Adaptation des populations virales aux résistances variétales et exploitation des ressources génétiques des plantes pour contrôler cette adaptation

Lucie Tamisier

Adaptation des populations virales aux résistances variétales et exploitation des ressources génétiques des plantes pour contrôler cette adaptation

Lucie Tamisier

Université d'Avignon et des Pays de Vaucluse
Ecole doctorale 536 « Agrosciences et Sciences »

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE D'AVIGNON ET DES PAYS DE VAUCLUSE

présentée et soutenue publiquement par

Lucie TAMISIER

le 7 décembre 2017

Adaptation des populations virales aux résistances variétales et exploitation des ressources génétiques des plantes pour contrôler cette adaptation

Thèse dirigée par Benoît MOURY et Alain PALLOIX

Jury :

Bruno Le Cam, Directeur de Recherche, INRA	Rapporteur
Olivier Le Gall, Directeur de Recherche, INRA	Rapporteur
Laurence Albar, Chargée de Recherche, IRD	Examinatrice
François Delmotte, Directeur de Recherche, INRA	Examinateur
Julie Lederer, Chargée de Recherche, HM Clause	Examinatrice
Benoît Moury, Directeur de Recherche, INRA	Directeur de thèse

Remerciements

Ce travail de thèse a été réalisé au sein des unités de recherche de Génétique et Amélioration des Fruits et Légumes (GAFL) et de Pathologie Végétale (PV) de l'INRA d'Avignon avec le soutien de la région PACA, du département BAP ainsi que du métaprogramme SMaCH. Cette thèse a également été soutenue par la société HM Clause à qui j'adresse mes remerciements.

Je tiens tout d'abord à remercier les membres de mon jury de thèse d'avoir accepté d'évaluer mon travail : merci à Bruno Le Cam et Olivier Le Gall pour leur travail de rapporteur. Merci également à Laurence Albar, François Delmotte et Julie Lederer d'avoir accepté d'être les examinateurs de ma thèse.

Je souhaite également remercier les membres de mon comité de thèse pour leurs remarques et leurs suggestions constructives concernant mes travaux de recherche. Merci à Charles-Eric Durel, Frédéric Fabre, Fabien Halkett, Julie Lederer et Christopher Sauvage.

Mes premières pensées en écrivant ces remerciements s'adressent naturellement à Alain. Je garde de lui le souvenir d'un directeur particulièrement bienveillant et sympathique, passionné par son métier et qui savait transmettre son savoir avec enthousiasme et talent. Il aura accompagné et inspiré de nombreux jeunes chercheurs, et je sais que c'est une chance pour moi d'en avoir fait partie.

Je tiens ensuite à adresser un immense merci à Benoît. Travailler avec quelqu'un d'aussi passionnant, enthousiaste et investi a été un réel plaisir et a grandement contribué à rendre mon expérience de thèse positive et formatrice. Merci d'avoir été un directeur toujours disponible, que ce soit pour les discussions scientifiques, la participation aux manips ou les corrections de la thèse. J'ai pu bénéficier grâce à toi d'un encadrement extrêmement enrichissant et je t'en remercie sincèrement.

Je remercie également l'ensemble des techniciens avec qui j'ai eu l'occasion de travailler durant ces trois années. Marion et Ghislaine, un immense merci de m'avoir toujours soutenue dans mes grosses manips en faisant preuve d'une organisation sans faille. Cette thèse n'aurait pas pu se faire sans votre aide et je sais que j'ai eu beaucoup de chance de travailler avec des

collègues aussi dévouées. Ghislaine profite de ta retraite bien méritée, et Marion je te souhaite le meilleur et j'espère que tu t'épanouiras dans tes nouveaux projets.

Merci également à Grégory pour toute l'aide que tu m'as apportée, notamment en début de thèse pendant les semaines d'extraction, ainsi qu'à Pauline et à Karine pour votre implication face aux montagnes de plaques ELISA. Je tiens également à remercier Nathalie, Joël et l'ensemble des équipes techniques qui s'occupent si bien des plantes. Merci aussi à Catherine et ses étiquettes qui m'ont été d'une aide précieuse pour organiser au mieux mes expériences, ainsi qu'à Patrick pour l'aide en ELISA et les conseils pour les photos.

Le génotypage des piments aura été un travail de longue haleine qui s'est bien déroulé grâce à plusieurs collègues qui m'ont permis de bénéficier de leur expertise. J'adresse donc de très grands remerciements à Renaud pour nous avoir si bien formées Marion et moi à l'extraction d'ADN, à Sylvain pour ses conseils avisés sur le GBS et le travail réalisé par son équipe sur les banques et enfin à Gautier pour m'avoir accueillie à Montpellier et pour l'aide qu'il m'a apportée sur le traitement des données.

Le séquençage des populations virales a également été le fruit d'un travail collaboratif. Merci à Catherine pour m'avoir accueillie à Toulouse et permis de participer au séquençage de mes échantillons. Un grand merci également à Frédéric et Elsa pour leur super modèle qui m'a fourni des données essentielles.

Je tiens aussi à remercier tous les membres des équipes Résistance Durable chez les Solanacées (RDS) et Virologie que j'ai côtoyés durant ma thèse. Un merci particulier à Karine et Emmanuel pour les discussions scientifiques (et les repas !) ainsi qu'à Judith pour nos échanges et son aide lors des dernières expériences. Merci aussi à nos supers secrétaires Claudine et Pascale pour tous les aspects administratifs.

J'ai également une pensée pour tous les collègues qui ont rendu mon séjour au château agréable. Je remercie notamment Joël qui partageait le bureau d'à côté et bon nombre de mes pauses café. Merci pour ton humour, ta bonne humeur permanente, tes petits tours (au sens propre) dans mon bureau, mais aussi pour tous tes conseils. Je remercie également Jean-Paul pour la disponibilité et l'efficacité dont il a fait preuve lorsque j'avais besoin d'aide en bioinformatique.

Un grand merci à tous les doctorants et collègues que j'ai pu côtoyer en dehors, pour tous les à-côtés, les randonnées, les concerts, les jeudis « Explo » ou encore les soirées jeux qui me laissent d'excellents souvenirs de ce séjour avignonnais. Merci à Elise, Gaëtan, Stéphanie, Zoé, Anna, Benoît, Maxime, Carole ainsi qu'à tout le groupe des doctorants de Saint Paul. Un merci tout particulier aux membres des « ratatouilles », Mariem et Isidore, pour nos sessions guitare du jeudi que j'ai tant appréciées. Enfin, je remercie évidemment mes proches pour leur soutien, et notamment les amis de longue date. Elodie, c'est toujours un plaisir de partir à l'aventure avec toi. Marion, merci pour tous ces bons moments et vivement notre prochaine escapade musicale. Arno, mille mercis pour ta joie de vivre, pour nos débats interminables et surtout pour le soutien sans faille durant ces trois ans (Rise up!). Les parents, vous êtes les meilleurs, merci pour tout.

Liste des abréviations

ADN: Deoxyribonucleic acid

ARN: Ribonucleic acid

AUDPC: Area under the disease progress curve

bp: Base pair

CMV: *Cucumber mosaic virus* (genus *Cucumovirus*)

DAS-ELISA: Double antibody sandwich enzyme-linked immunosorbent assay

DIECA: Diethyldithiocarbamate

DH: Doubled haploid

dpi: Days post inoculation

ddRADseq: Double digest restriction-site associated DNA sequencing

eIF(iso)4E: Eukaryotic initiation factor (iso)4E

eIF4E: Eukaryotic initiation factor 4E

FDR: False discovery rate

GBS: Genotyping-by-sequencing

GFP: Green fluorescent protein

GLM: Generalized linear models

GWAS: Genome-wide association studies

***h²*:** Heritability

LD: Linkage disequilibrium

MOI: Multiplicity of infection

MQM: Multiple QTL mapping

***N_e*:** Effective population size

PCR: Polymerase chain reaction

PVY: *Potato virus Y* (genus *Potyvirus*)

QTL: Quantitative trait locus

RB: Resistance breaking

RT-PCR: Reverse-transcription polymerase chain reaction

RdRp: RNA-dependent RNA polymerase

***s*:** Selection coefficient

SNP: Single-nucleotide polymorphism

TEV: *Tobacco etch virus* (genus *Potyvirus*)

ToMV: *Tomato mosaic virus* (genus *Tobamovirus*)

VA: Viral accumulation

VPg: Virus protein genome-linked

Sommaire

Remerciements 1

Liste des abréviations 5

Introduction générale 13

Introduction **17**

A. Playing on genetic drift to improve plant disease control strategies **19**

1 Using evolutionary principles to take control of plant pathogen adaptation 19

2 Why use genetic drift to increase the durability of crop resistance to disease? 23

3 Measures of genetic drift in pathogen populations at different stages of the plant infection process and at different scales 30

3.1 Empirical and modelling approaches to measure the effective population size of plant pathogen populations 30

3.2 Estimates of the effective population size of plant pathogens 33

3.2.1 Plant inoculation 34

3.2.2 Plant colonization 39

3.2.3 Plant-to-plant transmission 40

4 Agricultural practices to enhance the effect of genetic drift and decrease the evolutionary potential of pathogens 42

4.1 Decrease of the inoculum size and increase of the founder effect at within-host scale 42

4.2 Intra-host bottleneck and Muller's ratchet 44

4.3 Superinfection exclusion 46

4.4 Reduction of pathogen dispersal and survival 47

5 Future issues 48

B. Le pathosystème piment-PVY **61**

1 Le PVY 61

1.1 Description générale 61

1.2 Génome du PVY 61

1.3 Origine et classification 61

2 Le piment 63

2.1 Origine géographique et diversité 63

2.2 Génome du piment 64

2.3 Résistances aux potyvirus 66

2.3.1 Les résistances qualitatives 66

2.3.2 Les résistances quantitatives 68
2.3.3 La combinaison gène majeur/QTL 69
3 Un modèle biologique pour l'étude de la durabilité des résistances 71
3.1 Mécanismes d'action des QTL pour orienter l'évolution des populations virales. 71
3.2 Mesure des forces évolutives imposées par la plante aux populations virales 76
C. Objectifs de la thèse 79

Chapitre 1 : Quantitative trait loci in pepper control the effective population size of two RNA viruses at inoculation 81

1 Introduction 83
2 Results 84
2.1 Very few primary infection foci are initiated by two PVY variants simultaneously 84
2.2 The numbers of primary infection foci and local lesions are highly heritable traits 84
2.3 Detection of QTLs controlling the numbers of primary infection foci and local lesions for PVY and CMV 84
2.4 The number of infection foci induced by PVY correlates with the number of local lesions induced by ToMV 85
3 Discussion 85
3.1 Link between the number of primary infection foci or local lesions and the effective population size 85
3.2 Common and virus-specific QTLs control the effective population size of PVY and CMV at inoculation 87
3.3 Hypothesis on the mechanisms of action of the QTLs 88
3.4 Relationship between QTLs of effective population size and plant resistance 89
4 Methods 89
4.1 Plant and virus material 89
4.2 Evaluation of the N_e estimation method 90
4.3 Counting the primary infection foci and local lesion numbers induced by PVY, CMV and ToMV in a DH population 90
4.4 Statistical analyses 90
4.5 QTL analysis 90
5 Supplementary material 92

Chapitre 2 : Genome-wide association mapping of QTLs implied in Potato virus Y population sizes in the pepper germplasm **97**

1 Introduction 100

2 Results 102

2.1 A core-collection representative of the pepper germplasm 102

2.2 Distribution of SNPs in the pepper genome 104

2.3 Population structure and kinship relationships 104

2.4 Variation in the number of primary infection foci and the virus accumulation among the pepper core-collection 107

2.5 Genome-wide association mapping of pepper resistance to PVY 108

3 Discussion 112

3.1 Benefits and limits of the pepper core-collection to perform genome-wide association 112

3.2 Common genetic factors control the effective population size at inoculation and the virus accumulation 114

4 Materials and methods 117

4.1 Pepper core-collection sampling 117

4.2 Phenotyping of the core-collection 117

4.2.1 Number of PVY primary infection foci 117

4.2.2 Virus accumulation 117

4.3 SNP detection 118

4.4 Population structure and linkage disequilibrium estimations 119

4.5 Genome-wide association study 119

4.6 Statistical analyses 120

5 Supplementary material 121

Chapitre 3 : Impact of genetic drift, selection and within-host accumulation on virus adaptation to its host plants **135**

1 Introduction 138

2 Results 141

2.1 Estimates of N_e and s corresponding to a PVY composite population in 89 pepper DH lines 142

2.2 Correlation between putative explanatory variables and with the RB response variable 145

3 Discussion 150

3.1 Lack of strong relationships between N_e^{inoc}, N_e, s and VA 150

3.2 Which evolutionary forces contribute most to resistance breaking? 151
3.3 Applied consequences to improve the durability of major-effect resistance genes 155
4 Materials and methods 156
4.1 Previous data 156
4.2 Analysis of composite PVY populations infecting pepper DH lines 157
4.3 Inference of virus N_e and s 159
4.4 Statistical analyses of the links between variables related to the evolution of PVY populations 159
5 Supplementary material 160

Chapitre 4 : Taking control of virus adaptation by choosing host plant genotype 167

1 Introduction 170
2 Results 173
2.1 Divergent evolutionary trajectories among PVY lineages: extinction, *status quo* or parallel fixation of mutations 173
2.2 Significant increase in virus accumulation but little change in aggressiveness after experimental evolution 176
2.3 Validation of the impact of the two most frequent de novo mutations on virus accumulation 180
2.4 Contrasted effects of selection, genetic drift and initial accumulation on virus evolution 180
3 Discussion 183
3.1 Closely-related pepper lines impose divergent evolutionary trajectories to PVY 183
3.2 Choosing the plant traits most effective to avoid PVY adaptation 185
3.2.1 Relevance and precision of estimation of plant traits used to explain PVY evolutionary trajectories 185
3.2.2 Effects of plant traits on virus evolution 186
3.2.3 Agronomical perspectives 189
4 Materials and methods 190
4.1 Virus and plant material 190
4.2 Experimental evolution 191
4.3 Virus sequencing 192
4.4 Measures of aggressiveness and virus accumulation 192
4.5 Measures of virus competitiveness 193
4.6 Statistical analyses 193

5 Supplementary material 195
6 Complément d'information à l'article 201

Chapitre 5 : Discussion et perspectives **207**

1 Synthèse des principaux résultats 208
2 Comparaison des approches employées 212
2.1 Cartographie de QTL en population biparentale et GWAS 212
2.2 Evolution expérimentale par passages successifs et mesure de la durabilité du gène majeur *pvr2*[3] 213
2.2.1 La méthodologie employée 213
2.2.2 Les résultats obtenus 217
3 Par quels mécanismes le fonds génétique contraint-il l'évolution des pathogènes ?.. 218
3.1 Effet des goulets d'étranglement sur les mutations adaptatives 218
3.2 Peut-on confirmer l'action du cliquet de Muller ? 220
4 Conséquences pour la sélection variétale et l'application en champs 221
5 Conclusion générale 225

Références bibliographiques **227**

Annexe **247**

Introduction générale

L'amélioration de la protection des cultures contre les agents pathogènes est un des enjeux majeurs de la recherche agronomique. Actuellement, le niveau global des pertes de production agricole engendrées par les agents pathogènes demeure élevé, des pertes allant de 26 à 40 % de la production en champs ayant été estimées selon les espèces (Oerke 2006). De plus, les risques d'émergence d'agents pathogènes sont accrus par les changements climatiques ainsi que l'augmentation des échanges commerciaux au niveau international qui favorisent l'introduction d'espèces dans de nouvelles zones géographiques (Anderson et al. 2004).

Afin de limiter les dégâts occasionnés par les bio-agresseurs, l'utilisation de pesticides a longtemps été la solution privilégiée. Cependant, l'impact néfaste de ces intrants sur la santé et l'environnement rend aujourd'hui nécessaire la mise en place de méthodes de lutte alternatives telle que la lutte génétique. Cette stratégie couramment employée consiste à sélectionner des variétés génétiquement résistantes vis-à-vis de l'agent pathogène ciblé. La principale limite de la lutte génétique est la capacité des agents pathogènes à s'adapter aux variétés résistantes. Plusieurs gènes majeurs sont ainsi devenus inefficaces après quelques années de culture seulement (McDonald and Linde 2002; García-Arenal and McDonald 2003). De plus, la sélection d'une nouvelle variété résistante n'est pas toujours possible car les gènes majeurs sont rares dans les collections variétales. Aussi, les pratiques de sélection des plantes, qui ont toujours privilégié l'efficacité de la protection, doivent maintenant prendre en compte sa durabilité afin de ne pas épuiser les ressources génétiques.

D'autres sources de résistances existent dans les ressources génétiques. Il s'agit des gènes de résistance quantitative ou QTL (Quantitative Trait Loci), qui ne font que réduire les dégâts dus aux épidémies mais qui sont beaucoup plus fréquents dans les collections variétales. Chez plusieurs pathosystèmes, la combinaison d'un gène majeur et de QTL dans une même variété a permis d'augmenter significativement la durabilité de la résistance (Palloix et al. 2009; Brun et al. 2010; Fournet et al. 2013). Les QTL ont ainsi empêché ou réduit le contournement du gène majeur en orientant l'évolution des agents pathogènes. On suppose que les QTL affecteraient l'évolution des populations de pathogènes en agissant sur différentes forces évolutives, notamment la dérive génétique et la sélection. L'association d'un gène majeur et de QTL est donc une stratégie prometteuse. En effet, l'enjeu de la durabilité deviendrait accessible en déployant des gènes de résistance en fonction non seulement de leur capacité à réduire les populations pathogènes mais aussi à orienter et contrôler leur évolution dans l'agrosystème.

Le sujet de cette thèse s'intègre donc dans le contexte de la durabilité des résistances variétales aux agents pathogènes et fait également appel à des concepts de biologie évolutive. Afin de comprendre au mieux les résultats présentés dans la suite de ce manuscrit, l'introduction se structurera en trois parties. Elle débutera par un article d'opinion portant sur l'intérêt d'utiliser la dérive génétique pour augmenter la durabilité des résistances aux pathogènes. Ensuite, le modèle d'étude piment (*Capsicum annuum*) – PVY (*Potato virus Y*) utilisé au cours de cette thèse sera présenté. Les résultats récemment obtenus sur la durabilité des résistances chez ce pathosystème y seront plus particulièrement développés. Enfin, les objectifs de la thèse seront énoncés.

Introduction

A. Playing on genetic drift to improve plant disease control strategies

Tamisier, L. (1,2), Fabre, F. (3), Papaïx, J. (4), Berthier, K. (2), Moury, B. (2).

(1) GAFL, INRA, 84140 Montfavet, France

(2) Pathologie Végétale, INRA, 84140 Montfavet, France

(3) UMR 1065 Santé et Agroécologie du Vignoble, INRA, Villenave d'Ornon, France

(4) BioSP, INRA, 84000 Avignon, France

Keywords: genetic drift, evolutionary principles, plant disease management, effective population size

Abstract

The widespread use of monogenic resistances has proved to be an efficient method to fight against many plant diseases. However, the pathogen ability to evolve and breakdown these resistances is a major limit of this practice. Several authors have therefore highlighted the need to include evolutionary principles in the design of resistance management programs. They aim to limit pathogen evolutionary potential by playing on the evolutionary forces imposed by the pathogen environment. Among these forces, the role of selection to constrain pathogen evolution has been widely discussed. On the contrary, studies on the effects of genetic drift are scarce although it is one of the main evolutionary forces, even for plant microbe pathogens that possess huge census population sizes. In this opinion article, we argue that genetic drift should also be considered to achieve durable resistance. We discuss the benefits of genetic drift in the context of plant resistance durability, the different methods to assess its impact on pathogen population and we make proposals for its implementation in agricultural practices.

1 Using evolutionary principles to take control of plant pathogen adaptation

In plants, major-effect resistance genes have been widely used to fight against pathogens and protect crops. These genes provide a high level of resistance against pathogens, are environmentally safe and relatively easy to use. However, the main drawback of this genetic

control is the ability of the pathogen to overcome the resistance gene. Indeed, the deployment of a resistance gene in the field will unavoidably exert a strong selective pressure on the pathogen population. Many pathogen microbes have high mutation rates, short generation times, huge census population sizes and they usually show high evolution rates. If a resistance-breaking pathogen variant appears by mutation, recombination or migration into the population, the strong selective pressure imposed by the major resistance gene will lead to a rapid increase of the variant frequency and eventually to its fixation in the population. As a consequence, the new pathogen population may be able to invade the population of hosts that carry the resistance gene and the resistance may become completely inefficient. A high number of examples of such resistance breakdowns exist in the literature (Parlevliet 2002). Alternative gene types or genetic constructions are therefore needed to achieve more durable resistances. Several authors have proposed to use population genetics principles to improve resistance breeding strategies (McDonald and Linde 2002; Kinkel et al. 2011; Thrall et al. 2011; Zhan et al. 2014; Brown 2015; Zhan et al. 2015). Their idea is to view the plant as an environment in which the pathogen population will evolve. In this respect, the evolutionary forces*[1] that shape the pathogen population are therefore imposed by the plant itself. Indeed, the plant will select certain pathogen genotypes, will impose bottlenecks* to the pathogen population during the infection cycle, will impact the recombination rate between co-infecting pathogen variants by spatially structuring the population at the within-plant scale and will condition the fraction of possible mutations as well as their effects on pathogen fitness*. It is therefore of primary importance to understand how plant genotypes influence the evolution of pathogens during infection, because it could allow the breeding of plant cultivars preventing or limiting pathogen adaptation. Manipulating plant genetics according to selection* intensity (Brown 2015), mutation rate (Leach et al. 2001) and genetic drift* strength (Abel et al. 2015) to slow down pathogen adaptation has been discussed and experimented in several studies.

As mentioned above, plant resistance genes inherently exert selection effects on pathogen populations as far as there is some diversity in the capacity of infection of the resistance carrying plants among the pathogen population. Indeed, by definition, plant resistance reduces growth rate and potentially the within-host density of the pathogen. As soon as resistance-breaking (RB) variants, that escape the effect of the plant resistance gene, appear or are introduced into the pathogen population, differential selection effects are manifested in the resistance-carrying host plant. They can be defined as $s = r_{RB} - r_{WT}$, where r_{RB} and r_{WT} are the intrinsic growth rates

[1] Les termes marqués d'un astérisque sont définis dans un glossaire figurant page 50.

of the RB and wild-type (i.e. non-RB) pathogen variants. Multiple strategies have been proposed to use these differential selective effects to fight against crop pathogens. At the plant scale, the use of quantitative resistance* can slow down the evolution rate of the pathogens. This type of resistance confers only partial resistance and allows pathogens to accumulate in the plant and colonize it. Therefore, it decreases the competition among pathogen variants, making it a strategy potentially more durable than qualitative resistance conferred by a major gene, which, in contrast, can rapidly select and fix a RB allele in the pathogen population (Poland et al. 2009; Zhan et al. 2015). The slower action of selection on pathogens due to quantitative resistance has been observed for multiple pathosystems, such as *Potato virus Y* (PVY) with pepper (Quenouille et al. 2013) in the laboratory and in greenhouse as well as the barley pathogen *Rhynchosporium secalis* (Abang et al. 2006), and the wheat pathogens *Mycosphaerella graminicola* (Zhan et al. 2002) and *Phaeosphaeria nodorumin* (Sommerhalder et al. 2011) in the field. At the landscape scale, heterogeneity in the host population can impose divergent selective pressures on the pathogen population. The landscape can be represented as a heterogeneous environment with different fitness peaks. Pathogens will not be able to reach all adaptive peaks in this fitness landscape, which could increase the durability of the plant resistance genes used. Indeed, a pathogen variant able to infect multiple plant genotypes may have a lower infection efficiency than a pathogen specialized in one plant genotype of the mixture (Mundt 2002). This type of control is particularly efficient for diseases spread by spores, such as mildews and rusts. The efficiency of this strategy is more difficult to predict for pathogens like viruses, because the outcome also relies on the abundance and behavior of the vectors. Cultivar mixtures with different major resistance genes, different partial resistance genes or both can be used to impose this disruptive selection. This strategy has efficiently reduced the selection coefficient of *P. nodorumin* populations in the field. The genetic diversity of the pathogen population did not change during two years when evolving in a cultivar mixture of partially resistant hosts, whereas the pathogen evolution was faster in monoculture (Sommerhalder et al. 2011). Mixing resistant and susceptible cultivars has also been proposed to enhance resistance durability. One of the reasons is the dilution effect: non-RB pathogen variants will be maintained in the susceptible plant genotypes, whereas RB pathogen variants will probably be eliminated from these plants because of the fitness cost associated with the resistance breakdown. Using a mixture of susceptible and less-susceptible rice varieties to the blast disease in a large field scale, Zhu et al. (2000) have shown that the disease was reduced in both varieties and obtained 94% less severe rice blast than in monoculture. They explained it by the dilution effect of the inoculum of a strain thanks to the distance between the two varieties.

A significant decrease of *Phytophthora infestans* symptoms on potato was also obtained by mixing susceptible and resistant potato plants during two years (Garrett and Mundt 2000). Temporal patterning of the selective forces could also be applied using crop rotations. At the spatio-temporal scale, applying Red Queen* principles could help to limit the selection of pathogen variants (Zhan et al. 2015). The goal is to withdraw the crop carrying the resistance gene before the increase in frequency of the RB pathogen. The new crop will impose a negative selection on the variant, preventing its increase in the population.

Playing on the mutational trajectories accessible for the pathogen is another way to protect a major resistance gene from breakdown. Mutation can be used in two ways. First, the cost in fitness of a RB mutation can be too high for the mutation to be maintained in the population (Leach et al. 2001). In the absence of compensatory mutations, the RB strain will be removed from the population because of the action of negative selection. This is particularly true for pathogens with small genomes composed of multifunctional genes like viruses, where a small number of mutations can have strong negative impact on pathogen fitness (Carrasco et al. 2007). Indeed, plant durability is usually higher for plant resistance to viruses than to other plant pathogens because the genetic changes needed to overcome the resistance often imposed a very high fitness cost to the viruses (García-Arenal et al. 2003). The fitness cost associated to resistance-breakdown is the reason why, after a long period of deployment, some resistance genes are still effective against bacteria (Cruz et al. 2000) or viruses (Janzac et al. 2010). Second, the number of mutations needed to overcome the resistance is directly related to resistance durability (Harrison 2002). For viruses, if two or more mutations are needed to overcome the major gene, the resistance is usually considered as durable (Lecoq et al. 2004). The explanation for this outcome is that the intermediate variants, carrying one of the two mutations needed, are not necessarily fit enough to be maintained in the population and to acquire the second mutation. For example, in tobacco, the well-known *N* gene confers a highly durable resistance against *Tobacco mosaic virus* (TMV) because of the high number of mutations required to overcome the resistance (Padgett et al. 1997; Harrison 2002).

To date, the link between genetic drift and resistance durability has been less studied than for the other evolutionary forces, even is several authors have proposed to take advantage of this force to limit the emergence of adapted pathogens (Abel et al. 2015). We will see in the next section that genetic drift may, however, have a strong impact on resistance durability, and that this force should also be taken in consideration in plant disease control strategies.

2 Why use genetic drift to increase the durability of crop resistance to disease?

Genetic drift is the random variation of allele frequencies from one generation to the next. Like selection, it always induces a loss of genetic diversity in the population over time. However, since genetic drift is a stochastic process, it leads to the fixation or the loss of alleles independently of their fitness effects. The population genetics parameter used to quantify the impact of genetic drift on the genetic structure of the population is the effective population size* (N_e). It is defined as the size of an ideal Wright-Fisher* population that would experience the same amount of genetic drift as the population under study (Charlesworth 2009) (Box 1). When N_e is small, the effect of genetic drift is strong and can overwhelm the action of selection, while when N_e is high, the effect of genetic drift is low and selection is the predominant force. Wright (1931) first introduced the concept of effective population size. He also demonstrated that, when population varies over time, N_e can be approximated as the harmonic mean of the effective populations sizes over all generations. Since then, several methods have been developed to estimate N_e leading to the emergence of different concepts of N_e such as the variance effective size, the inbreeding effective size, the eigenvalue effective size or the coalescent effective size (Ewens 1982; Crow and Denniston 1988; Caballero 1994; Nordborg and Krone 2002; Charlesworth et al. 2003; Wang 2005). For instance, the variance effective size measures the change in allele frequencies from generation to generation while the inbreeding effective size measures the rate of increase in homozygosity. These two measures are the most used, and in this article, we will essentially refer to the variance effective size. Another feature of N_e is that it is usually much smaller than the census population size* (N). In the case of plant viruses for example, the size of a *Tobacco mosaic virus* (TMV) population infecting a tobacco leaf can reach between 10^{11} and 10^{12} viral particles (Gibbs et al. 2008), while N_e estimates range from one to a few hundred particles (Zwart and Elena 2015). Since N_e is directly related to the amount of genetic drift undergone by the population, it is more relevant to consider N_e than N when studying evolutionary processes like the adaptation of a pathogen to a resistant plant.

Box 1. The joint effect of selection and genetic drift: illuminating example from simulations of a Wright-Fisher model

The Wright-Fisher model occupies a central position in population genetics (Ewens 2004). It explicitly accounts for the effects of the main evolutionary forces – mutation, genetic drift and selection – on the dynamics of allele frequencies. The population considered in the Wright-Fisher model is an ideal population, which is a well-mixed, randomly-mating haploid population of finite size reproducing in discrete non-overlapping generations. The concept of ideal population is central for understanding the definition of effective population size, N_e. Indeed, N_e is the size of an ideal population (*i.e.* obeying previous assumptions) that would display the same degree of randomness in variant frequencies as the real population under study.

Simulating allele frequencies over time with a Wright-Fisher model is a good way to visualize how genetic drift and selection impact evolution. Let's consider a haploid population of constant size N with 2 alleles A and a and equal initial frequencies (p=q=0.5). In this situation, N is equal to the effective population size (N_e). These 2 alleles can represent two virus variants characterized by two alternative nucleotides at a given position in an RNA sequence. The allele A has a selection coefficient s, its relative fitness being $1+s$ (the fitness of allele a is 1).

We simulated the dynamic of the frequencies of A using a Wright-Fisher model over an array of increasing N and s (Figure 1). Selection is a deterministic force that increases the frequency of the fittest variants at the expense of the weakest ones. Simulations highlighted that the distinctive footprint of selection is contained in the mean trajectories of variant frequency (Figure 1, compare each row). Genetic drift, unlike selection, acts equally on all variants. It is the outcome of random sampling effects between generations in finite population size (Charlesworth 2009). The distinctive footprint of genetic drift is contained in the variance of the trajectories of variant frequency, higher variance being associated to lower N (Figure 1, compare each column). These distinctive footprints can be used to jointly estimate selection and genetic drift by observing allele frequencies at several time points in independent hosts.

Classical results are illustrated throughout these simulations. First, with a neutral allele (Figure 1, a to c; first raw), the mean allele frequency remains constant: $p_A(t)= p$. Half of A allele goes to fixation, half goes to extinction. This is a classical result: the proportion of populations expected to go to fixation for a given neutral allele is equal to its initial frequency (here 0.5).

The time to absorption (corresponding either to a loss or a fixation of *A*) increases with *N* as the effect of drift per generation becomes smaller. Allele *A* is always fixed or lost after 40 generations with *N*=10 (Figure 1a, d, g) whereas all the populations remained polymorph after 100 generations with the largest size considered (*N*=1000, Figure 1c). 5000 generations are then needed to go systematically to fixation or loss. In a haploid population, the mean time to fixation of a neutral allele depends on the population size and on its initial frequency. It is approximated by Tps(fix) = -(1/p)[2*N*(1-p)ln(1-p)] (Kimura and Ohta 1969). In figure 1 (a to c), the average times to fixation are 13.8 (*N*=10), 138 (*N*=100) and 1386 (*N*=1000) generations. Average times to fixation are quite fluctuating but are on the order of *N* (Rouzine et al., 2001). By contrast, average times to fixation are much less sensitive to the initial allele frequency (Figure 2a). When p is small, as for example in the case of a new variant generated by mutation, the average time to fixation of a neutral allele is 2*N* (Figure 2a, $p=10^{-4}$). Indeed, for small p, ln(1-p)≈-p and the above formula provides Tps(fix) = 2*N*.

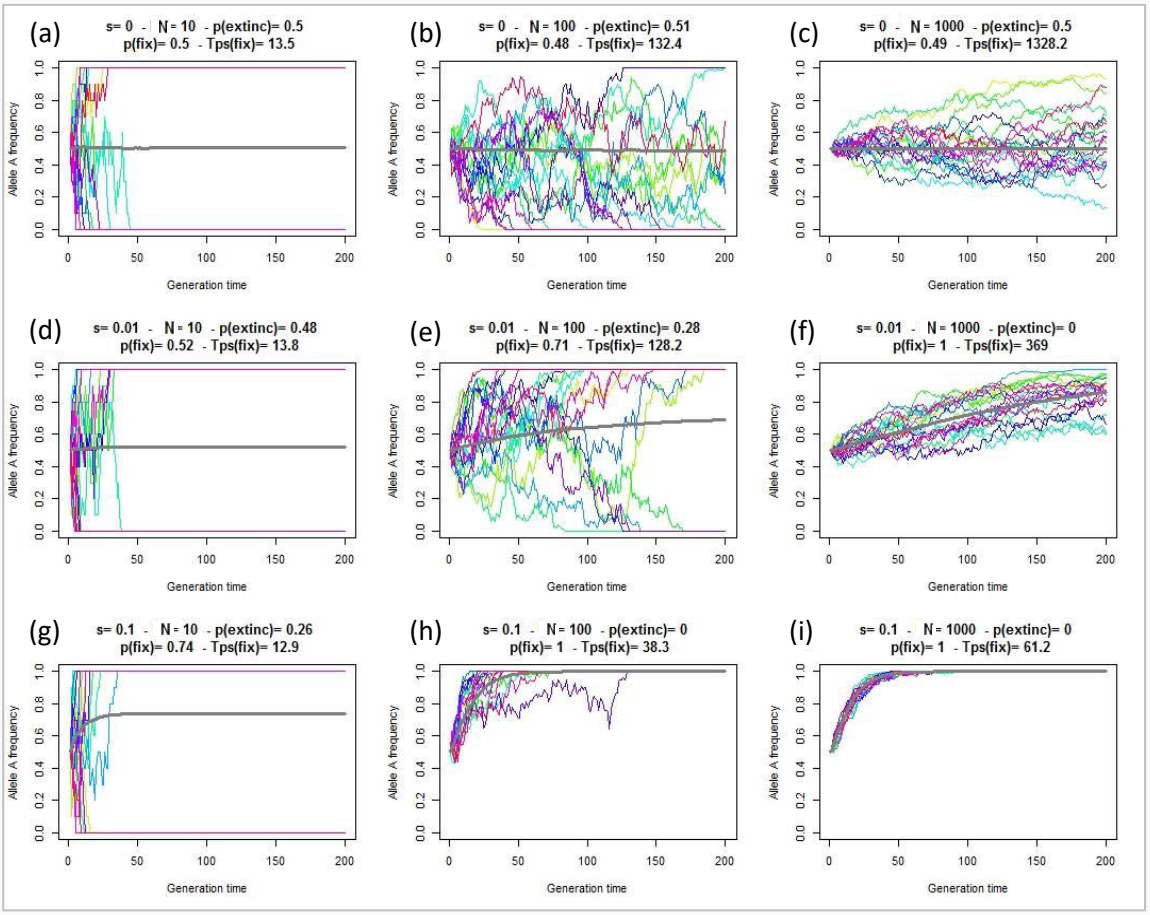


Figure 1: Simulations (20 replicates) of allele frequencies over 200 generations using bi-allelic Wright-Fisher model for 3 population sizes *N* (10, 100 and 1000) and 3 selection coefficients *s* (0, 0.01, 0.1). Initially the two alleles have equal frequencies. The mean trajectories correspond to the bold black lines. Probabilities of fixation (p(fix)) and extinction (p(extinct)) and the mean fixation time (Tps(fix)) are estimated after 5000 generations using 10^4 simulated trajectories.

Second, when selection and genetic drift act simultaneously, the dominant regime of evolution is controlled by the product Ns (Rouzine et al. 2001; Ewens 2004; Charlesworth 2009). If $Ns \ll 1$, then genetic drift predominates over selection and evolution is mostly stochastic (Figure 1, in particular panel (d) but also (e) and (f)). Allele dynamics behave close to neutrality. The fixation probability, close to initial allele frequencies, is almost independent of selection. So is the mean time to fixation. This is clearly illustrated in figure 2b (fixation probability) and in figure 2c (fixation time) for (s, N) pairs below the second diagonal (*i.e.* $Ns<1$). The typical footprints of selection and genetic drift are then hard to distinguish, making their joint inference a hard task. If $Ns \gg 1$, selection becomes effective and evolution is mostly deterministic (Figure 1, f, h, i). The beneficial allele considered almost always goes to fixation (Figure 2b for (s, N) pairs above the second diagonal (*i.e.* $Ns>1$)). The time to fixation is also considerably reduced compared to the neutral case (Figure 2c). The joint inference problem becomes much simpler in that case.

With both selection and drift, no analytical form of the mean time to fixation is available. Motoo Kimura derived an approximation of the probability of fixation of allele A as

$$\pi \simeq \frac{1-\exp(-2Nsp)}{1-\exp(-2Ns)}$$ (see Patwa and Wahl (2008) for a review).

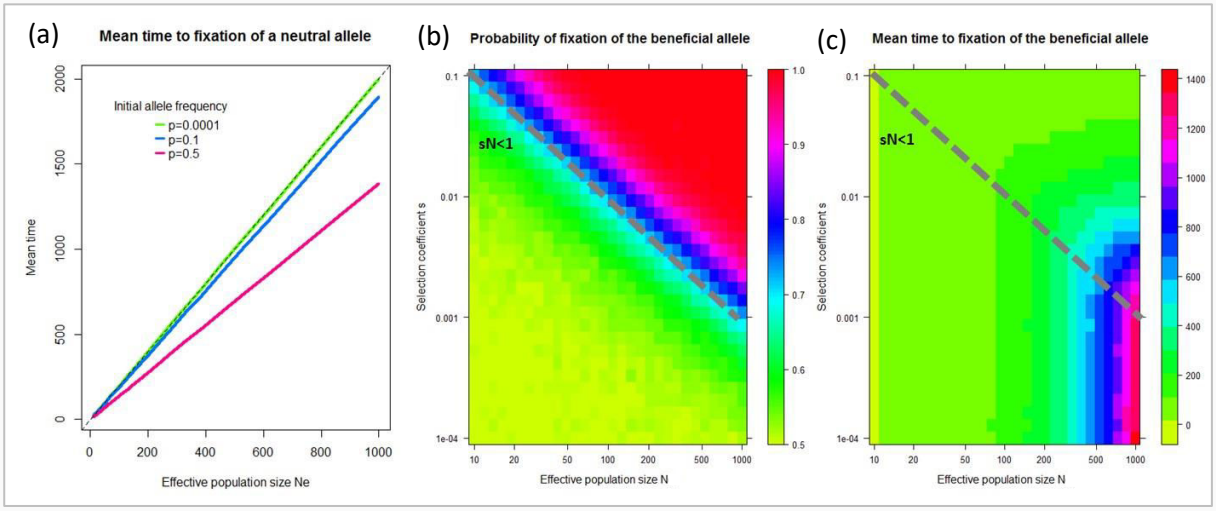


Figure 2. **(a):** Expected time to fixation for a neutral allele as a function of effective population size for 3 initial frequencies. Lines converge to the dash line for which Tps(fix)=2N as initial frequency decreases. **(b):** Probability of fixation of a beneficial allele with selection coefficient s (x-axis) in an ideal bi-allelic Wright-Fisher population of size N (y-axis) and equal initial frequencies. **(c):** Same as B for the mean time to fixation when fixation occurred. Probabilities of fixation and mean fixation times are estimated after 5000 generations using 10^4 simulated trajectories. The gray dotted line (B and C) separates the selection-drift space where $sN<1$ (below) and $sN>1$ (above).

The main goal of this article is to highlight the important role of genetic drift imposed by the host on the pathogen populations and how we could use it to improve plant resistance durability. As we have seen above, a strong genetic drift will reduce the genetic variation of plant pathogen populations. Pathogen population with a high N_e (i.e. low genetic drift) have a better ability to adapt and to potentially break down resistance genes or become resistant to antibiotics and fungicides. Consequently, we can hypothesize that a strong genetic drift will reduce the evolutionary potential* of the pathogen population and that it will avoid or slow down the adaptation of the population to the control method, leading to an increase in plant resistance durability. Both experimental studies and meta-analyses have provided evidences supporting these assumptions. Indeed, several experiments have demonstrated that the drastic reduction of population size through repetitive bottlenecks induced a strong genetic drift usually associated with a decrease of the pathogen's fitness. These 'mutation accumulation' experiments consist in imposing repeatedly bottlenecks to the population through time. As a result, all mutations that are non-lethal can be randomly fixed independently of their fitness effects and without being eliminated by selection (Barrick and Lenski 2013). When these experiments are performed with asexual populations and in the absence of recombination events, a process called Muller's ratchet* occurred (Box 2). It predicts that, since the majority of the mutations are deleterious, an irreversible accumulation of deleterious mutations will occur over time and will be responsible of a constant decrease in the mean fitness of the population. Eventually, the population may go to extinction because of a mutational meltdown (Lynch et al. 1993). The first experimental proof of this concept was provided by a famous experiment of Chao (1990). By passing the RNA bacteriophage Φ6 through a series of bottlenecks, he showed that the transferred clones had lost 78% of fitness compared to the parental clones. Since then, the Muller's ratchet has also been proved to occur with animal RNA viruses such as the *Vesicular stomatitis virus* (Duarte et al. 1992) and the Foot-and-mouth disease virus (FMDV) (Escarmís et al. 1996). The action of Muller's ratchet has also been demonstrated for the human immunodeficiency virus type 1 (HIV-1) (Yuste et al. 1999) and the herpes simplex virus type 1 (HSV-1) (Jaramillo et al. 2013). A loss in fitness through naturally occurring bottlenecks due to local lesions in plants has been shown for TEV (*Tobacco etch virus*) as well (de la Iglesia and Elena 2007). The same results were obtained for bacteria with *Escherichia coli* (Kibota and Lynch 1996) and for yeast with *Saccharomyces cerevisiae* (Zeyl and DeVisser 2001). Even if experiments directly testing the impact of bottlenecks on resistance durability are still needed, the multiple demonstrations that repetitive bottlenecks

caused a loss of evolutionary potential in the pathogen population are already a good indicator of the potential advantages of the use of genetic drift.

The problem of pathogens that evolve drug resistance in medicine field is a similar issue as plant resistance breakdown from an evolutionary viewpoint. Recently, Feder et al. (2016) have compared the HIV-1 sequences from patients treated with older drug therapies, where virus evolved rapidly and frequently acquired resistance, to HIV-1 sequences from patients treated with newer and more effective drugs, where the virus evolves slowly and rarely acquired resistance. They demonstrated that, when resistance mutation occurred, they were associated to an increase in genetic diversity with a less effective treatment and to a reduction in genetic diversity with an effective treatment. They propose that, in the case of old treatment administration, the same adaptive mutation can arise several times independently in different virus genetic backgrounds, leading to the higher genetic diversity observed, and can be predicted in a deterministic manner. In contrast, better drugs diminish the population size which decreases in turn the genetic diversity, the acquisition of drug resistance becoming partly an unlucky occurrence. These results suggest that a reduction in N_e and the increase in genetic drift associated could be one of the mechanisms explaining the better efficacy of the new treatments.

Meta-analyses on plant pathogens have also highlighted the fact that genetic drift could have a significant impact on resistance durability. Using data from 52 pathosystems, McDonald and Linde (2002) have developed a risk assessment model to estimate the risk of resistance breakdown according to the evolutionary potential of bacteria, fungi and nematodes. The evolutionary potential of the pathogens was determined by multiple criteria such as the mating system, the gene flow* intensity or the pathogen effective population size. In their model, the effective population size was a categorical variable with three levels: small, average and large. In this model, a small population size decreased the evolutionary potential of the pathogen. Even if it is an over simplification of the pathogens biology, they used it as a risk factor for crop damages and they concluded that disease management programs that maintain small pathogen population sizes are helpful to control resistance breakdown. García-Arenal and McDonald (2003) have performed the same analysis for plant viruses. They also showed that effective population size plays a role in the risk of resistance breakdown by viruses.

Box 2. Muller's ratchet

This process describes the irreversible accumulation of deleterious mutations in an asexual population (Muller 1964). Every individual is sorted according to the number of deleterious mutations it carries. The class with the lowest number of deleterious mutations carries 'n' mutations, the class with one more deleterious mutation carries 'n+1' mutations etc. The appearance of a mutation will move an individual from one class to another. Selection will eliminated the classes carrying most of the deleterious mutations. Genetic drift will eliminate, by chance, the classes carried the lowest amount of mutations. In this model, there is no recombination allowed, as well as no reverse or compensatory mutations. Therefore, once eliminated by drift, the class with the lowest amount of mutations could not be recreated. Since there is no turning back, the number of deleterious mutations increases over time in a ratchet-like manner. Eventually, this irreversible accumulation of deleterious mutations can cause the extinction of the population. The speed of the ratchet is increased when the population size is small, the effect of genetic drift being stronger in small populations, when the mutation rate is high and when the negative selection against deleterious mutations is low (Gordo and Charlesworth 2000). The mechanism of Muller's ratchet is not only useful to explain the extinction of asexual populations, but has also been used to explain the degeneration of Y chromosome, the evolution of sexual reproduction, the advantage of recombination and the limit to the genome size of asexual organisms (Charlesworth 1978; Engelstädter 2008).

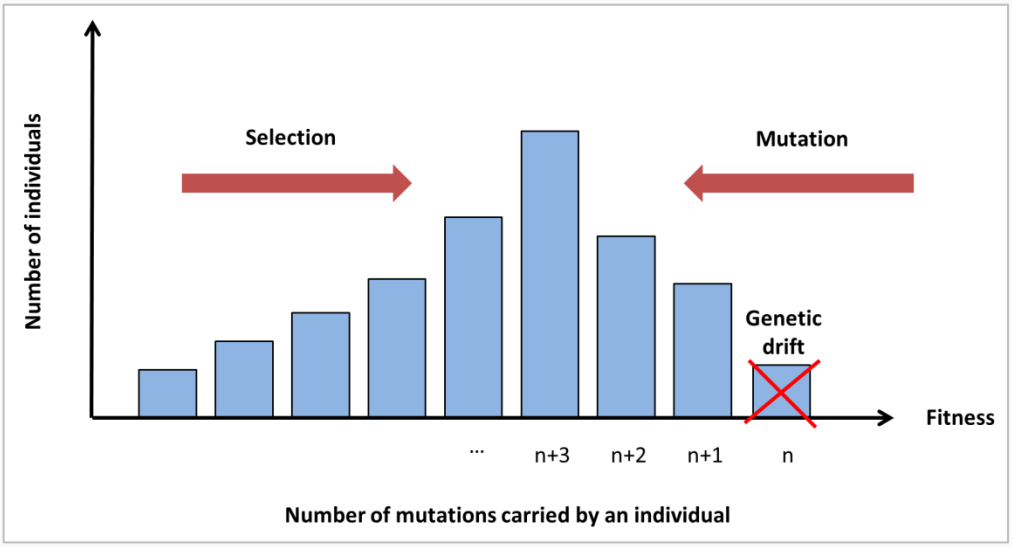


Figure 3: Illustration of the Muller's ratchet process. Each bar corresponds to a class of individuals carrying the same number of deleterious alleles. The effects of the evolutionary forces on the frequencies of these classes of individuals are drawn in red. Adapted from Lefevre et al. (2016).

All these results demonstrated that N_e is an important parameter to take into consideration when studying pathogen adaptation to plant resistances. However, to take advantage of N_e, we first need to accurately estimate it, with both empirical and mathematical approaches, and to understand its effects at different infection stages.

3 Measures of genetic drift in pathogen populations at different stages of the plant infection process and at different scales

3.1 Empirical and modelling approaches to measure the effective population size of plant pathogen populations

Two types of empirical approaches exist to measure N_e: direct and indirect methods. The direct estimation method follows directly the demographic changes occurring in the population. For example, the size of a bottleneck is measured by quantifying the number of individuals that survive after the reduction of the population size. One way to measure the bottleneck size at inoculation for viruses is to use a host producing necrotic local lesions in response to the infection. In this case, the number of local lesions is a direct and easy estimation of the number of founders (Sacristán et al. 2011; Zwart and Elena 2015). This approach is not restricted to plants producing necrotic lesions. Using viruses expressing fluorescent markers, it is possible to visualize and quantify the number of primary infection foci in the inoculated leaf. Zwart et al. (2011) have demonstrated that one infection focus was caused by only one viral particle for TEV marked with GFP and mCherry in *Nicotiana tabacum*. The number of foci on the infected leaf is therefore also a direct estimation of N_e at inoculation. We can make the hypothesis that this method to measure N_e could be extended to other pathogens. For example, the number of spots caused by bacteria on a leaf could also be used as a direct assessment of the number of founders initiating the infection. For pathogens that enter into the leaf through stomata like some bacterial and fungal pathogens, the number of stomata could be correlated to the number of founders. For pathogens that penetrate the leaf only when the stomata are opened, the stomatal pore size could also be correlated to founder effect* size. However, these assumptions still need to be tested. Nevertheless, the direct estimation method is not always achievable because pathogens use a complex migration pathway within the host with overlapping generations, making the direct counting of the number of individuals contributing to the next generation impossible (Abel et al. 2015).

The indirect estimation method uses the genetic changes occurring in the population over time to estimate N_e. This estimation can be made on natural populations when the genetic polymorphism is high enough, but can also be carried out experimentally. To do so, a pathogen population composed of two or more genotypes of known concentrations is inoculated into the host. The changes in genotype frequencies can then be followed at different time points thanks to specific markers. Since some models that estimate N_e assume neutrality, neutral markers are often used to differentiate between genotypes and these genotypes are usually equally fit. Multiple types of markers have been used. For example, to study the effect of bottlenecks on viral diversity, Li and Roossinck (2004) have created an artificial CMV population of 12 restriction enzyme-bearing mutants. The genotypes can also be tagged with fluorescent markers (Miyashita and Kishino 2010), non-coding sequences (Monsion et al. 2008; Lam and Monack 2014) or can differ in one or two specific non-synonymous mutations (Fabre et al. 2012). The frequencies of each genotype can then be assessed by hybridization, real-time quantitative PCR (qPCR) or DNA sequencing (Abel et al. 2015). In virology, the observation of fluorescently labeled variants have been performed in situ (Miyashita and Kishino 2010; Bergua et al. 2014) or in protoplasts by microscopy (González-Jara et al. 2009) or flow cytometry (Tromas et al. 2014). Deep sequencing analyses of the pathogen populations probably provide the best estimations of N_e. The high sequencing depth allows an accurate estimation of the genotypes frequencies within the population and enables the reconstruction of haplotypes (Kutnjak et al. 2017). However, the reconstruction is technically difficult to perform and requires several properties such as a sufficient amount of overlap between reads or a sufficient number of segregating polymorphic sites.

From these empirical data, a range of methods have been developed to estimate N_e. When a combination of two pathogen genotypes have been inoculated into plants, simple mathematical models that consider the probabilities for a plant or an organ of being infected by one or both variants have been used (Sacristán et al. 2003; Betancourt et al. 2008; González-Jara et al. 2009). When hosts showed only mixed-variant infections, two methods have been employed (Zwart and Elena 2015). First, the Wright's F_{ST} has been used to partition genetic variation within and between populations, allowing to estimate N_e between different time-points and/or plant organs (Monsion et al. 2008; Fabre et al. 2014; Tromas et al. 2014). The second method measures the changes in variant frequencies over time, since a small N_e induces a strong change in pathogen frequencies (Monsion et al. 2008; Fabre et al. 2014). However, these methods rely on the assumption that all variants are equally competitive, which allows to ignore

the effect of selection. Indeed, since both selection and genetic drift reduce the genetic diversity of the population, being able to distinguish the action of both forces on pathogen population is not an easy task, especially because the polymorphism studied is not necessarily neutral. For that reason, many methods require neutral markers to detect only the action of genetic drift (Nei and Tajima 1981; Waples 1989; Williamson and Slatkin 1999; Anderson et al. 2000; Wang 2001; Berthier et al. 2002). These methods cannot always be applied when studying pathogens with small genomes, because few sites are selectively neutral. A few methods jointly estimating selection and genetic drift are available (Table 1). Some of them make strong assumptions, like a high N_e ($N_e > 5000$) and a small s ($s < 0.01$) (Bollback et al. 2008; Malaspinas et al. 2012; Mathieson and McVean 2013; Steinrücken et al. 2014). When studying pathogen adaptation to plant resistances, these assumptions are not always relevant because a major resistance gene imposes a strong selective pressure to the pathogen population. A strong competition between pathogen variants and high s values are therefore expected. Besides, the pathogen population undergoes narrow bottlenecks within the host, which decrease N_e to very low values. Foll et al. (2015) have developed a method that can be applied to a wide range of N_e and s values. However, it also requires a large proportion of neutral markers ($\geq 90\%$). Terhorst et al.'s (2015) method can be used for non-neutral markers. It requires that allele frequencies are far from fixation or extinction, and that N_e is moderate ($N_e \approx 1000$). However, it cannot estimate both parameters simultaneously and assumes that N_e is known in order to estimated s, and vice versa. Lacerda and Seoighe's (2014) as well as Rousseau et al's (2017) methods can handle non-selectively neutral markers and can estimate jointly N_e and s. Lacerda and Seoighe's (2014) method can deal with very high selection coefficients ($s = 0.5$) but has only been tested for small effective population size ($N_e = 1000$). On the opposite, Rousseau et al's (2017) method is effective for much lower N_e values (i.e. a few tens individuals) but has not been tested for cases of very strong selection. This method has been developed for haploid and asexual organisms, and therefore is not suitable to all types of pathogen populations. Recently, Ali et al. (2016) have developed a method that can be applied to partially-clonal organisms like fungal pathogens. It allows the estimation of N_e during the periods of clonal reproduction and the proportion of the population that reproduces sexually.

Table 1: Overview of models estimating the effective population size (N_e) and selection coefficient (s). Adapted from Malaspinas (2016)

References	Model and approximation	Assumptions	Single/multiple locus	Estimated parameters[a]
Bollback et al. (2008)	WF model, diffusion approx.	High N_e, small s	single	$2N_es$
Malaspinas et al. (2012)	WF model, diffusion approx.	High N_e, small s	single	s
Mathieson and McVean (2013)	WF model, Gaussian approx.	Small s, $0 << x_i << 1$[b]	single	s
Foll et al. (2014)	WF model	≥ 90% neutral markers	multiple	N_e, s
Lacerda and Seoighe (2014)	WF model, Delta method	High N_e	single	N_e, s
Steinrücken et al. (2014)	WF model, diffusion approx.	High N_e, small s	single	N_e, s
Terhorst et al. (2015)	WF model, Gaussian approx.	$0 << x_i << 1$	multiple	either N_e or s
Ali et al. (2016)	WF model	Partially clonal organisms	multiple	N_c (i.e. N_e of a partially clonal organism)
Rousseau et al. (2017)	WF model	Both high and/or small N_e and s	multiple	N_e, s

WF: Wright-Fisher
x_i: frequency of the allele i within the population
[a]: Only the estimations for s and for N_e are provided
[b]: allele frequencies are assumed to remain far from extinction and fixation

3.2 Estimates of the effective population size of plant pathogens

All the N_e estimation methods have been applied to different pathosystems and have provided an overview of the effect of genetic drift during the infection. In this section, we will review the estimates for N_e of different pathogens and at different steps of the infection process, from within-host infection to transmission between hosts (Table 2). The examples came mainly from the virology field because most of the N_e estimates have been done for viruses, but results for other pathogens will be provided when available.

3.2.1 Plant inoculation

During the inoculation of a new plant, only a small fraction of the source pathogen population will be part of the inoculum and will be able to establish a new population. This strong bottleneck is called the founder effect and it contributes for a large part of the genetic drift. For viruses, several studies have estimated the number of viral particles initiating an infection. This number has proven to be very small, independently of the inoculation method. For aphid transmission, between 0.5 and 3.2 infectious particles of PVY per aphid was estimated to be inoculated on *Capsicum annuum* (Moury et al. 2007). Consistent values were obtained for *Cucumber mosaic virus* (CMV), with an average of 1 or 2 founders initiating the infection (Betancourt et al. 2008). Using mechanical inoculation, Zwart et al. (2011) estimated that between 1 and 50 TEV particles caused the infection of *Nicotiana tabacum*, according to the initial virion dose. Similar results were obtained by Sacristán et al. (2011) for contact transmission of TMV. They estimated that the number of founders initiating a single local lesion was one, and that the total number of founders initiating infection after a contact event transmission lies between 1 and 4. A strong genetic drift was also observed during vertical transmission of *Pea seedborne mosaic virus* (PSbMV), the number of founders contributing to the infection of a seedling from an infected mother plant being close to 1 (Fabre et al. 2014). Only one study has provided an estimation of the N_e at inoculation for plant bacteria. The number of founders of *Ralstonia solanacearum* during root infection of tomato plants (*Solanum lycopersicum*) was estimated to be approximatively 458 (Jiang et al., 2016). The same trend as plant viruses is found for animal viruses. For example, between 5 and 42 DENV (*Dengue virus*) founders are able to pass the midgut cells of the mosquito host (Lequime et al. 2016). Given those results, genetic drift turns out to be a major force acting at the initiation of an infection.

Table 2: Estimates for the effective population size (N_e) of different pathogens and at different steps of the infection cycle. The methods used to follow the demographic (direct method) or genetic changes (indirect method) in the pathogen population and the estimation methods for N_e are indicated.

Scale	Vector/ transmission mode	Pathogen		Host	Pathogen genotyping	N_e estimation method	N_e estimates	Reference
Transmission	Aphid (*Myzus persicae*)	Virus	*Potato virus Y* (PVY)	*Capsicum annuum*	Competition between viruses	Probabilistic model	0.5 - 3.2	Moury et al. 2007
	Aphid (*Aphis gossypii*)	Virus	*Cucumber mosaic virus* (CMV)	*Solanum lycopersicum*	Genotype-specific oligonucleotide probes	Probabilistic model	1 - 2	Betancourt et al. 2008
	Leaf contact	Virus	*Tobacco mosaic virus* (TMV)	*Nicotiana tabacum*	Genotype-specific oligonucleotide probes	Probabilistic model	1.3 -3.3	Sacristán et al. 2011
	Mechanical inoculation (leaves)	Virus	*Tobacco etch virus* (TEV)	*Nicotiana tabacum*	Fluorescent labeling	Quantification of the mean number of primary infection foci	1.20 - 47.86	Zwart et al. 2011
	Mechanical inoculation (leaves)	Virus	*Tobacco etch virus* (TEV)	*Capsicum annuum*	Fluorescent labeling	Quantification of the mean number of primary infection foci	1.1 - 5.4	Zwart et al. 2011
	Soil inoculation (roots)	Bacteria	*Ralstonia solanacearum*	*Solanum lycopersicum*	Antibiotic resistance marker	Probabilistic model	458	Jiang et al. 2016
	Vertical transmission	Virus	*Pea seedborne mosaic virus* (PSbMV)	*Pisum sativum*	Sequencing of polymorphic loci	Probabilistic model	1	Fabre et al. 2014

Table 2: *Continued*

Scale	Organ	Pathogen		Host	Pathogen genotyping	MOI estimation method	MOI estimates	Reference
Cell	Bark tissue	Virus	*Citrus tristeza virus* (CTV)	*Citrus macrophylla*	Fluorescent labeling	Probabilistic model	1.066	Bergua et al. 2014
	Indvidual leaves	Virus	*Turnip mosaic virus* (TuMV)	*Brassica rapa*	Fluorescent labeling	Probabilistic model	1 - 30	Gutiérrez et al. 2015
	Indvidual leaves	Virus	*Tobacco Mosaic Virus* (TMV)	*Nicotiana tabacum*	Fluorescent labeling	Probabilistic model	0.85 - 1.22	González-Jara et al. 2009 González-Jara et al. 2013
	Indvidual leaves	Virus	*Cauliflower mosaic virus* (CaMV)	*Brassica rapa*	Genotype-specific DNA inserts	Probabilistic model	2 - 13	Gutiérrez et al. 2010
	Indvidual leaves	Virus	*Tobacco etch virus* (TEV)	*Nicotiana tabacum*	Fluorescent labeling	Probabilistic model	1.001 - 1.431	Tromas et al. 2014
	Inoculated leave	Virus	*Soil-borne wheat mosaic virus* (SBWMV)	*Chenopodium quinoa*	Fluorescent labeling	Probabilistic model	5.02 - 5.97	Miyashita & Kishino 2010

Table 2: *Continued*

Scale	Organ	Pathogen		Host	Pathogen genotyping	N_e estimation method	N_e estimates	Reference
Within-host infection	Indvidual leaves	Virus	*Cauliflower mosaic virus* (CaMV)	*Brassica rapa*	Genotype-specific DNA inserts	Change in variance of neutral markers	8.8 - 131	Gutiérrez et al. 2012
	Indvidual leaves	Virus	*Tobacco etch virus* (TEV)	*Nicotiana tabacum*	Fluorescent labeling	FST statistics	5.83 - 107.00	Tromas et al. 2014
	Indvidual leaves	Virus	*Pea seedborne mosaic virus* (PSbMV)	*Pisum sativum*	Sequencing of polymorphic loci	Change in variance of neutral markers and Fst statistics	56 - 216	Fabre et al. 2014
	Pool of leaves	Virus	*Potato virus Y* (PVY)	*Capsicum annuum*	Variant frequencies over time using NGS	FST statistics	1 - 4	Fabre et al. 2012
	Pool of leaves	Virus	*Potato virus Y* (PVY)	*Capsicum annuum*	Sequencing of polymorphic loci	Rousseau et al (2017)'s method estimating conjointly the intensities of selection and genetic drift	13 - 1469	Rousseau et al. 2017
	Tiller	Virus	*Wheat streak mosaic virus* (WSMV)	*Triticum aestivum*	RFLP markers	Probabilistic model	4	Hall et al. 2001 French & Stenger 2003
	Indvidual leaves	Virus	*Tobacco Mosaic Virus* (TMV)	*Nicotiana tabacum*	Genotype-specific oligonucleotide probes	Probabilistic model	3.1 - 5.6	Sacristán et al. 2003
	Indvidual leaves	Virus	*Cauliflower mosaic virus* (CaMV)	*Brassica rapa*	Genotype-specific DNA inserts	Change in variance of neutral markers and Fst statistics	298 - 484	Monsion et al. 2008

Table 2: *Continued*

Scale	Spatio-temporal environment	Pathogen		Host	Pathogen genotyping	Genetic drift estimation method	N_e estimates	Reference
Between-host transmission	Multiple fields, 6-year period	Fungus	*Mycosphaerella graminicola*	*Triticum aestivum*	RFLP markers	Ewens (1972)'s method	3,400 - 700,000	Zhan et al. 2001
	9 fields in 9 countries, different years	Fungus	*Rhynchosporium commune*	*Hordeum spp.*	Microsatellite markers	Qst/Fst statistics	Not measured	Stefansson et al. 2014
	Multiple sampled locations (~150km long), 1-year period	Nematode	*Heterodera schachtii*	*Beta vulgaris*	Microsatellite markers	Pseudolikelihood method (Wang 2001) Moment-based method (Jorde & Ryman, 2007)	85	Jan et al. 2016

3.2.2 Plant colonization

After the inoculation step, the remaining parasite genotypes will start to replicate and to colonize the inoculated organ. Plant pathogens use different colonization routes: bacteria and fungi usually multiply in the intercellular space, viruses are exclusively intracellular and some fungi use both pathways. During host colonization, all pathogen populations will encounter several barriers that could affect N_e, regardless of their colonization route.

For intra-cellular parasites, plasmodesmata represent one of these barriers. Plasmodesmata are intercellular channels that connect adjacent plant cells and allow the diffusion of small plant metabolites. These are dynamic channels that can dilate to spread macromolecules like proteins and RNAs. Pathogens such as viruses and some hemibiotrophic fungi exploit this property to perform cell-to-cell movement through plasmodesmata and colonize the inoculated leaf until they reach the phloem (Lee and Lu 2011). In virology, the MOI (multiplicity of infection) is a common measure that estimates the number of viral particles that contribute to individual cell infection, that is, the bottleneck size during the cell-to-cell movement through plasmodesmata. This step has also been proved to impose a narrow bottleneck to the population and low values ranging from one to a few tens have been reported to date. González-Jara et al. (2009, 2013) were the first to provide an estimation of the MOI of a virus. They estimated a very low MOI for TMV in *N. tabacum*, varying from 0.85 to 1.22 during host colonization. Small values of MOI were also reported for *Soil-borne wheat mosaic virus* (SBWMV), with a size of bottleneck during cell-to-cell movement between 5 and 6 in the inoculated leaf (Miyashita and Kishino 2010). For *Tobacco etch virus* (TEV), this value was not greater than 1.5 (Tromas et al. 2014), as well as for *Citrus tristeza virus* (CTV) which has shown an MOI around 1 (Bergua et al. 2014). Several studies have demonstrated that the MOI is not a constant number but fluctuates during the host colonization according to the plant organ. For *Cauliflower mosaic virus* (CaMV), the MOI was estimated around 2 at the beginning of the infection, then was shown to increase at 13 and to decrease to the initial level in the latest infection stages (Gutiérrez et al. 2010). Moreover, CaMV concentration in vascular tissue is positively correlated with the MOI (Gutiérrez et al. 2012). Recently, Gutiérrez et al. (2015) demonstrated that the MOI of *Turnip mosaic virus* (TuMV) was very low in companion cells primarily infected from the vasculature but increases around a value of 30 during cell-to-cell movement in the mesophyll.

Estimations of N_e within the plant have also been made to a larger scale than the cellular level. During host colonization, the N_e of pathogens will be influenced by the action of the host

immune system, the ability of pathogens to reach different plant organs or the pathogen capacity to establish in different plant tissues (i.e. intercellular or intracellular colonization, invasion of phloem sieve tubes or xylem vessels, etc.). Hall et al. (2001) were the first to estimate the number of founder during virus spread in wheat, with an estimated value of 4. Low N_e values during systemic infection have also been found for TMV (Sacristán et al. 2003) and PVY (Fabre et al. 2012). In contrast, higher N_e were reported during leaf colonization for CaMV (Monsion et al. 2008) and PSbMV (Fabre et al. 2014). Since N_e during colonization can depend on N_e at inoculation (Zwart et al. 2011), different inoculum doses could be behind these differences. The correlation between the viral dose in the phloem and the bottleneck size in the systemic leaf found by Gutiérrez et al. (2012) is also in agreement with this hypothesis.

To our knowledge, estimations of N_e for other pathogens than viruses within the plant host are not available yet. However, bottlenecks probably also occur, modifying N_e and the effect of genetic drift during host colonization. For example, during infection by pathogens such as fungal or nematode parasites, plants showing partial resistance to the pathogens may modify their physical structure after the inoculation and produce barriers that will block the pathogens. Plants may limit the spread of the pathogens into cells by producing polymers like callose, lignin and suberin, which reinforces cell walls and probably modifies N_e (Zhan et al. 2014).

3.2.3 Plant-to-plant transmission

At the host population scale, multiple variables will affect the N_e of the pathogen population. First, the pathogen evolutionary potential will influence the overall effective population size. Short generation time increase population size and probably increase N_e. Pathogens showing long-distance dispersal tend to have a greater N_e, since it reduces the probability to mate with a conspecific and to decrease the population genetic diversity. High gene flow links subpopulations of pathogens and increases N_e.

Zhan et al. (2001) have studied the wheat pathogen *Mycosphaerella graminicola* during 6 years in the field. They have found a very high genetic diversity among the years, probably because the ascopore founder population was sufficiently large to prevent genetic drift effect. Therefore, they have estimated a large effective population size ranging from 3,400 to 700,000 individuals. This result is probably due to the pathogen features, which show sexual recombination, long-distance dispersal of ascospores and strong gene flow. More recently, a study on another fungus, the barley pathogen *Rhynchosporium commune*, has demonstrated that

many pathogen traits are under stabilizing selection and that high within-population genetic variation is maintained for these traits (Stefansson et al. 2014). In contrast, genetic drift little affects the pathogen traits. This high evolutionary potential is probably due to selection imposed by a wide range of agricultural conditions combined with gene flow. At the opposite, the evaluation of the N_e of beet cyst nematode *Heterodera schachtii* within a year provided a low effective population size estimated at 85 individuals (Jan et al. 2016). This result was explained by the variation in population size through time, systematic inbreeding, restricted dispersal, development on short-lived hosts and unbalanced sex-ratios.

Second, multiple environmental parameters will probably influence the pathogen N_e. For example, the presence/absence of plant reservoirs, the pathogen vector densities, the host density, the plant life cycle (annual or living longer than one year) or the period of host absence most likely play a role in N_e. However, few data directly linking these parameters to pathogen N_e values are available. Bergstrom et al. (1999) have investigated the role of host population size on the effect of horizontal transmission bottlenecks. In their model, the N_e of the virus was equal to the product of the N_e within the host and the number of hosts. Therefore, they argued that the genetic drift intensity decreased as the host population size increased. However, their model make strong assumption about the number of infected hosts, which stays constant over time. A more realistic model allowing variation of the number of infected plants during the epidemic could lead to different conclusions on the link between host population size and N_e of the pathogen. Regarding empirical studies, Rieux et al. (2013) have studied the impact of agricultural landscape heterogeneity on the genetic structure of the plant pathogenic fungus *Mycosphaerella fijiensis* infecting banana plantations. They expected genetic drift to occur in areas treated with fungicides compared with untreated areas, because the application of chemicals will induces strong reduction in population size, resulting in a decrease in pathogen genetic diversity. Indeed, they found a drastic reduction in *M. fijiensis* population size due to fungal control methods. However, no decrease in genetic diversity was observed between the two areas, probably because the population size remains large enough to avoid genetic drift effects. Gene flow from the outside could also have restored the genetic diversity of the population evolving in the treated area.

4 Agricultural practices to enhance the effect of genetic drift and decrease the evolutionary potential of pathogens

In this section, we will review several mechanisms that could enhance the effet of genetic drift and discuss how we could implement them in agricultural practices.

4.1 Decrease of the inoculum size and increase of the founder effect at within-host scale

During the inoculation step, decreasing the inoculum size (i.e. the N_e at inoculation) thanks to the founder effect will reduce the genetic diversity of the pathogen population, thereby reducing the evolutionary potential of the pathogens and increasing the effect of genetic drift. Therefore, if a RB variant is already present in the population, a strong founder effect could reduce its transmission to the next host plant or even cause its loss by chance (Figure 4). Indeed, several authors have found a link between the number of founders at inoculation and the level of intra-host infection. For TEV, Zwart et al. (2011) have shown that the N_e at inoculation was positively correlated to the N_e at the level of systemic infection, and the model of Lafforgue et al. (2012) has proved that the number of founders determines the number of infected cells at the systemic level. Modeling the time for viruses to reach systemic infection, Rodrigo et al. (2014) found that small N_e at inoculation will impose a delay before the occurrence of systemic infection. Decreasing N_e as soon as the inoculation step could therefore have important consequences to limit the spread of the pathogens during the other steps of the infection process.

A first way to increase the founder effect is to use genetic control using plant factors controlling the number of founders at inoculation. In a previous section, we have seen that for viruses we can measure the founder effect by simply quantifying the number of local lesions caused by a virus on a leaf and that this method could potentially be applicable on a wide range of pathogens. For example, Caranta et al. (1997) have measured the number of local lesions caused by a necrotic strain of *Cucumber mosaic virus* (CMV) on a pepper doubled-haploid population. They have found a highly-significant variance of the trait (from 1 to more than 100 lesions per leaf) and have estimated a heritability of 0.94, meaning that the plant genotype explained 94% of the trait variance. After performing QTL detection, they have identified three pepper QTLs controlling the number of necrotic local lesions, that is, the number of founders at inoculation. Combining a major resistance gene with these QTLs could be an efficient way

to fight against pathogens. Indeed, by reducing N_e at inoculation, these QTLs could potentially protect the major gene and make the resistance more durable. This hypothesis has already been proposed to explain the durability of a major resistance gene against PVY when introgressed into a partially-resistant genetic background in pepper (Quenouille et al. 2013). The same strategy could be applied for other traits, like the number of stomata on a leaf. In mulberry (*Morus* spp.), it has been showed that a reduction in stomatal density increases the resistance to powdery mildew (Chattopadhyay et al. 2011). Morevover, the closure of stomata in response to *Pseudomonas syringae* infection has been demonstrated in *Arabidopsis* (Melotto et al. 2006). These mechanisms could also be selected by breeders and used to reduce N_e at inoculation.

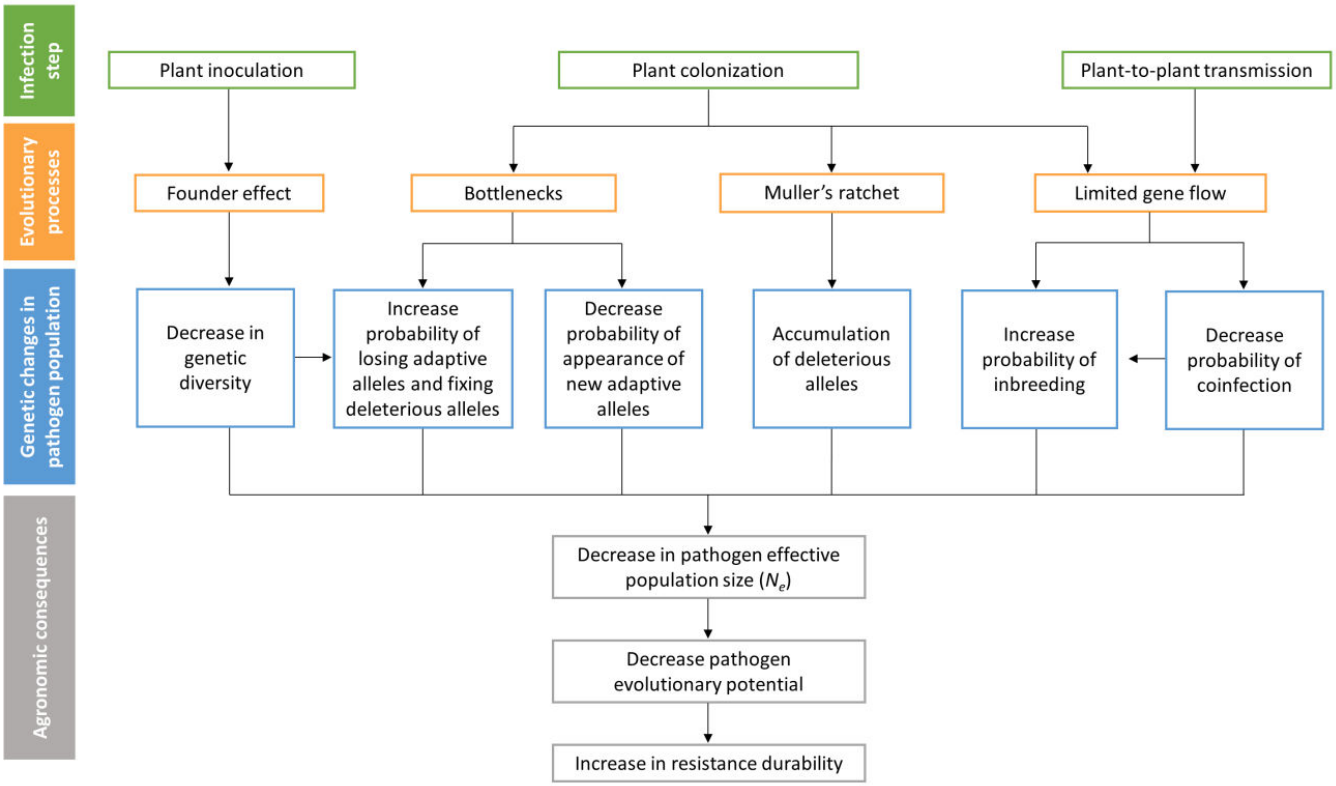


Figure 4: The diagram describes, for different steps of the infection cycle (green), the evolutionary processes (orange) leading to a decrease in population size (N_e) and their consequences on the genetic structure of the pathogen population (blue).

A second way to increase the founder effect is the crop management in agricultural landscapes (Table 3). Every strategy that will induce a succession of crops in a specific land or any crop-free period will cause a founder effect. Such changes will impose a severe reduction in pathogen population size contrary to monoculture where the pathogen population size

remains constant (McDonald and Linde 2002). Consequently, a small part of the pathogen population will remain and, potentially, RB alleles will be lost. Strategies such as crop rotations or fallow cropland may help to increase the founder effect. Field hygiene, like stubble burning, and the use of clean seeds will also reduce the inoculum available in the pathogen population and increase the founder effect (Zhan et al. 2015). Finally, we have seen that cultivar mixture between susceptible and resistant crops will impose a dilution effect on the pathogen population. This effect reduces the inoculum size and could help to decrease N_e at inoculation.

Table 3: Impact of crop management strategies on disease epidemic frequency and pathogen evolution regarding N_e, s and pathogen dissemination. Adapted from Zhan et al. (2014).

Management approach	Epidemic frequency	Evolution of pathogens		
		Effective population size (N_e)	Selection (s)	Dissemination
Agrochemical	Reduced	Reduced	Directional	Not affected
Clean seeds	Reduced	Reduced	Not affected	Reduced
Hygiene	Reduced	Reduced	Not affected	Reduced
Monoculture	Increased	Increased	Increased	Not affected
Quarantine	Reduced	Reduced	Not affected	Reduced
R Mixture	Reduced	Reduced	Disruptive	Not affected
R Rotation	Reduced	Reduced	Disruptive	Not affected
Resistance	Reduced	Reduced	Directional	Not affected
Species diversification	Reduced	Reduced	Disruptive	Reduced

4.2 Intra-host bottleneck and Muller's ratchet

At the intra-host scale, we have seen that the pathogens will encounter several barriers that will impose bottlenecks and lead to a decrease in N_e and an increase in genetic drift. In asexual population, the repetitive bottlenecks can lead to the extinction of the population because of the Muller's ratchet process (Box 2). In sexual population, it will promote inbreeding and decrease

the genetic diversity and the pathogen evolutionary potential (Figure 4). To play on this effects, we first need to understand how plants influence the bottleneck size.

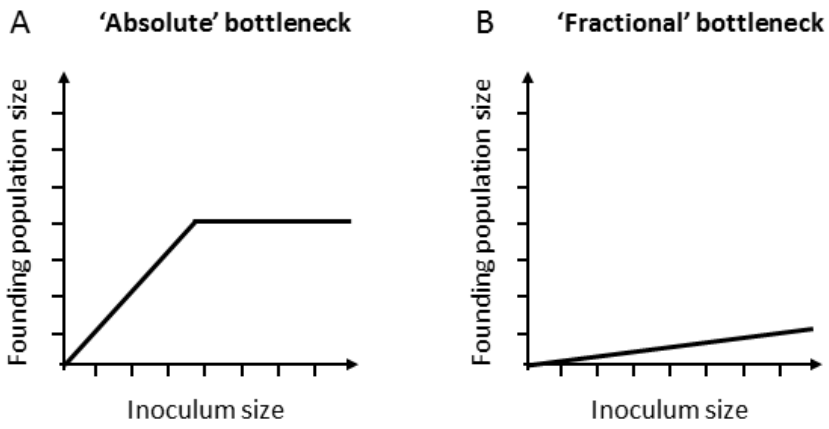


Figure 5: Hypothetical bottlenecks leading to different relationship between the inoculum size and founding population size. **A**: During a fractional bottleneck, a proportion of the inoculum does not survive. **B**: An absolute bottleneck allows the passage of organisms until its capacity is exhausted. Adapted from Abel et al. 2015.

According to the mechanisms set up by the plant to stop infection, Abel et al. (2015) have described different kinds of bottlenecks that could occur within the host. First, they have proposed that any host defence that counteract infection could be seen as a bottleneck. Indeed, secondary metabolites like phenol or alcaloids that acts as toxin against pathogens, as well as specialized chemicals and proteins produced by resistant plants during infection will reduce pathogen population and play to a certain extent on N_e. The host defences are more likely to produce a "fractional" bottleneck (Figure 5A). In this scenario, only a proportion of the population survives to the bottleneck, this proportion increasing with the inoculum size. Another kind of bottleneck probably occurring in the plant is "absolute" bottleneck (Figure 5B). If the number of entry points within an organ is limited, the number of founders that enter in this organ will be constant (i.e. reaching an "absolute" number) and independent of the size of the population before the bottleneck. Gutiérrez et al. (2012) have studied which type of bottleneck was met in turnip plant when CaMV goes from vascular system to leaves. They have not found an "absolute" bottleneck but have demonstrated that N_e within the leave was rather dependent of the virus concentration in plant sap, according to the "fractional" bottleneck definition. However, bottlenecks inducing an upper limit of the population size entering in a plant organ could still occur for other pathogens or in other parts of the plant. For example, the number of primary infection foci induced by a virus falls into this category of bottlenecks. At

first, the number of foci on the leaf increases with the viral dose (Zwart et al. 2011), but then it will reach a limit due to the physical space on the leaf that cannot produces much foci. The multiplicity of infection (MOI), that is the number of viral particles able to enter in a cell, also belong to this category of bottlenecks. Plants with low MOI could reduce coinfection, increase genetic drift effect and potentially lead to Muller's ratchet in infecting pathogen populations. MOI also modifies virus variant interaction. For example, in specific conditions, low MOI can decrease complementation between viruses. Using vesicular stomatitis virus (VSV), Novella et al. (2004) have showed that in competition assays, a high MOI increases the fitness of a deleterious mutant compared to the wild type. They developed a model that confirmed that a high MOI increases complementation and allows the viruses to share gene products. In this situation, select plants with low MOI could decrease the pathogen population genetic diversity by eliminating the defective mutant.

A better understanding of the different types of bottleneck imposed by the plant, from the cellular level to the organ level, and an accurate estimation of the size of these bottlenecks during plant colonization are currently needed and could help to select plants imposing very narrow bottlenecks to the pathogen population. It could also be useful to investigate more deeply the effects of the pathogen concentration and composition on the plant bottlenecks.

4.3 Superinfection exclusion

Superinfection takes place when a conspecific or closely-related pathogen is inoculated into a host individual or in a cell after infection was established or initiated by a first pathogen. The second pathogen can be excluded, leading to a complete lack of infection. This phenomenon is called cross-protection when considering plant-virus interactions. It will diminish N_e at the host scale by limiting the number of variants entering in a host/cell (Syller 2016).

Researchers have proposed to take advantage of this mechanism by inoculating plants with a lower aggressive strain, which will not induce damages to the plant but will provide exclusion from superinfection by more virulent strains (i.e. like a vaccine in human medicine). Attempts to use this mechanism have been done for example with *Zucchini yellow mosaic virus* (ZYMV) in squash (Lecoq et al. 1991), *Papaya ringspot virus* (PRSV) in papaya (Yeh and Gonsalves 1984) or *Citrus tristeza virus* (CTV) in citrus trees (Folimonova 2013). For CTV, cross-protection has been successful in some citrus areas. However, the mechanism is only efficient

for isolates belonging to the same CTV strain. Since infections are often caused by mixture of different virus strains, researches are still needed to develop a broad-spectrum cross-protection against this virus. Cross-protection between two virus strains can also be asymmetrical, the efficiency of superinfection exclusion depending on the first infecting strain (Fabre et al. 2010). In that case, genetic drift resulting from cross-protection and selection resulting from its asymmetry will jointly accelerate the decrease of genetic diversity in the virus population.

4.4 Reduction of pathogen dispersal and survival

High dispersal and survival abilities increase the genetic diversity of the pathogen population. First, it enhances the probability of coinfection, increasing the number of pathogen variants in the population (Figure 4). Second, coinfection increases the probability of reproduction and recombination between individuals. It promotes sexual reproduction or parasexual recombination for some fungi, bacterial conjugation as well as virus recombination and reassortment. In these conditions, the genetic diversity of the pathogen population increases and it could lead to the appearance of new RB variants and/or to a chance in pathogen host range (McDonald and Linde 2002). Pathogens with high dispersal abilities also have increased N_e, which limits the impact of genetic drift and the random loss of alleles in the population.

Genetic control could also be useful to limit pathogen dispersal. Regarding vector-borne pathogens, we have already seen that vectors impose a very strong bottleneck to the pathogen population (Moury et al. 2007; Betancourt et al. 2008). However, if the number of vectors feeding in a plant is high, a large part of the source population will be transmitted and the effect of the bottleneck imposed by a single vector individual will be low. Several studies have already proposed to control the disease by reducing the vector population density (Escriu et al. 2003). In this way, N_e will stay low and the action of genetic drift during plant-to-plant transmission will still be effective to maintain a low evolutionary potential in the pathogen population. The use of plant resistance genes targeting the vector population could be a way to achieve this objective. For example, the *Vat* gene in melon confers a resistance both to *A. gossypii* and the viruses it transmits by initiating a hypersensitive reaction in mesophyll and epidermal cells (Boissot et al. 2016). Note that this gene is atypical because it confers both resistance to plant colonization by the aphid and to the transmission of viruses by the vector. For pathogens that spread by spores, reducing the numbers of spores produced by the pathogen population within the plant will also reduce dispersal. Richardson et al. (2006) have created multiple barley lines by introgressing different QTL combinations in a susceptible genetic background. They have

measured the disease resistance of the lines, and have notably found that QTLs which decreased the pustule density on the leaf were associated with a higher resistance. Such QTLs could help to limit gene flow between hosts during fungal infection.

At the landscape scale, physical barriers can help to minimize the spread of propagules. In a cultivar mixture, resistant plants act as physical barriers and reduce the gene flow between susceptible plants. Disease management approaches including the regular elimination of the founder population, for example the elimination of necrotic tissues where sexual ascospores are produced (Rieux et al. 2013), may be helpful to diminish dispersal. Crop rotations may once again help to decrease N_e and the survival of the population by modifying host availability over time. However, modelling studies have highlighted that, in certain situations, pathogens can adapt to the periodic absence of the host, which produces a higher pathogen transmission rate (Van Den Berg et al. 2011). Finally, reducing the density of an alternate host needed by the pathogen to complete is life cycle may be efficient to limit epidemic development and pathogen survival.

5 Future issues

Could genetic drift favour pathogen adaptation to resistant plants?

Several authors have highlighted the fact that, in specific conditions, genetic drift could be advantageous for the pathogens and speed up their adaptation to resistance factors. Miyashita and Kishino (2010) have shown that within-host bottlenecks can make the action of selection more efficient for advantageous alleles acting in *trans*. For example, a viral mutation improving the activity of an inhibitor of apoptosis will be advantageous for all virus variants if the number of variants in a cell is high, while it will only confer benefice to the variant carrying the mutation when the number of variants in a cell is low due to bottleneck (Zwart and Elena 2015). The strong bottleneck will therefore accelerate the selection of this specific adapted variant. Zwart and Elena (2015) also proposed that bottlenecks could be advantageous in a multi-peaks fitness landscape. If the fitness of a single mutant is lower than the fitness of the wild-type virus and the double mutant, the single mutant will be eliminated by selection. Genetic drift could help to isolate this single mutant and could allow it to cross a fitness valley and reach a higher fitness peak. Further studies on the potential adaptive role of bottlenecks are crucial to better

understand their impact if we want to take advantage of them to increase resistance durability and avoid such drawbacks.

Can we detect plant genetic factors controlling N_e and to which extent could they be used to fight against various pathogens?

In the same way as we have characterized resistance genes or QTL imposing different levels of selection on pathogen populations, we could detect plant genetic factors imposing different levels of genetic drift. Moreover, since genetic drift is a stochastic force, we can make the hypothesis that these genetic factors will show a broad-spectrum of action, and could be used to fight against several pathogen variants or even species.

Can pathogens adapt to genetic factors controlling genetic drift?

Since genetic drift is a stochastic process acting on every pathogen variant independently of its fitness, it seems counterintuitive to propose that pathogen population could adapt to this mechanism. However, like the erosion of a quantitative resistance, we could imagine that some pathogen variants could adapt to the mechanism inducing genetic drift and make it less effective. For example, some pathogens could show a higher probability to survive after a bottleneck. During the inoculation of a leaf with a virus, highly stable virions could bypass the resistance mechanisms induced by the plant and show a higher probability to initiate an infection than less stable virions. Moreover, a better ability to survive outside the host plant could help to reduce the effect of genetic drift during a crop-free period. Regarding the superinfection exclusion mechanism, the selection of mutants that partly overcome it has already been reported for viruses (Webster et al. 2013). These examples strengthen the need to develop models jointly estimating the effect of selection and genetic drift in order to better characterize the part of the mechanism inducing stochastic changes and the part inducing determinist changes.

How does the host physiology influence bottlenecks?

Tromas et al. (2014) have demonstrated that the bottleneck size differed between leaves in tobacco and that these differences could be explained by the host physiology. How the phloem is wired in the host plant, the sink-source transitions and the overall host anatomy could largely contribute to the bottleneck size. A better understanding of the impact of the host physiology on the bottleneck size is therefore crucial if we want to later use this trait.

Do very low N_e values observed for viruses within the plant stand for others pathogens?

At the intra-plant scale, several values of N_e and MOI are now available for viruses. However, the data are very scarce for other pathogens. Such estimations would be useful to evaluate the interest of using genetic drift to increase durability against these pathogens.

Glossary

Bottleneck: Severe reduction in population size which can lead to the random loss or fixation of alleles.

Census population size: The total number of individuals in the population, usually symbolized as N.

Effective population size: The size of an ideal Wright-Fisher population that would experience the same rate of genetic drift as the observed population. Usually symbolised as N_e, it determines the strength of genetic drift in the population. Genetic drift is stronger when N_e is small and lower when N_e is large.

Evolutionary forces: The forces leading to changes in the genetic composition of a population. The five main evolutionary forces are selection, genetic drift, mutation, recombination and migration.

Evolutionary potential: The ability of a population to adapt to its environment. This potential depends of the effects of evolutionary forces such as the mutation rate, recombination, migration, the intensity of selection or the effective population size.

Founder effect: Isolation of a small number of individuals that will establish a new population. The new population carries only a small part of the genetic diversity of the initial one.

Fitness: Survival and reproductive success of an individual in its environment. For microorganisms, population growth rate or accumulation level are often used as fitness proxies at the within-host scale.

Genetic drift: Random change in allele frequencies in the population over time, independently of their fitness effect. The impact of genetic drift is stronger in small populations.

Gene Flow: Exchange of alleles or individuals among separated populations.

Muller's ratchet: A process leading to the irreversible accumulation of deleterious mutations in an asexual population. The individuals are categorized under different classes according to the number of deleterious mutations they carry. If the class with the lowest number of mutations (i.e. the fittest class) is lost by genetic drift, then, in the absence of recombination and reverse mutations, the lost class cannot be recreated. Consequently, the number of deleterious mutations in the population increases in a ratchet-like manner. Eventually, this process can lead to the extinction of the population.

Qualitative resistance: This resistance exhibits a discrete variation in the host plant population, showing only completely sensitive and completely resistant plant genotypes. It confers an almost-total resistance against the pathogen and it is usually under the control of one major gene.

Quantitative resistance: This resistance exhibits a continuous phenotypic variation in the host plant population, showing a continuum between completely sensitive and completely resistant plant genotypes. It can confer a partial resistance to the plant because it usually allows the multiplication and the colonization of the pathogen. This type of resistance is usually controlled by multiple loci (QTLs for quantitative trait loci) but can also be under the control of a single gene.

Red Queen hypothesis: A scenario of coevolution where oscillations in pathogen and host genotype frequencies are under negative frequency-dependent selection. The most frequent host genotype will be massively infected by pathogens adapted to this genotype. In consequence, the less frequent host genotypes are favoured and increase in frequency, while the most common ones begins to decrease in frequency.

Selection: Determinist change in allele frequencies in the population which increases the frequencies of advantageous alleles at the expense of deleterious ones.

Wright-Fisher model: A popular stochastic model in population genetics. It describes the evolution of allele frequencies in a population under genetic drift only. The assumptions of the model are a panmictic population, a constant number of individuals from generation to generation, non-overlapping generations and the absence of selection, mutation and migration.

References

Abang, M. M., Baum, M., Ceccarelli, S., Grando, S., Linde, C. C., Yahyaoui, A., Zhan, J. and McDonald, B. A. (2006). Differential selection on *Rhynchosporium secalis* during parasitic and saprophytic phases in the barley scald disease cycle. *Phytopathology* **96**, 1214–1222.

Abel, S., Abel zur Wiesch, P., Davis, B. M. and Waldor, M. K. (2015). Analysis of Bottlenecks in Experimental Models of Infection. *PLoS Pathog.* **11**, e1004823.

Ali, S., Soubeyrand, S., Gladieux, P., Giraud, T., Leconte, M., Gautier, A., Mboup, M., Chen, W., Vallavieille-Pope, C. and Enjalbert, J. (2016). CLONCASE: estimation of sex frequency and effective population size by clonemate resampling in partially clonal organisms. *Mol. Ecol. Resour.* **16**, 845–861.

Anderson, E. C., Williamson, E. G. and Thompson, E. A. (2000). Monte Carlo evaluation of the likelihood for N_e from temporally spaced samples. *Genetics* **156**, 2109–2118.

Barrick, J. E. and Lenski, R. E. (2013). Genome dynamics during experimental evolution. *Nat. Rev. Genet.* **14**, 827.

Bergstrom, C. T., McElhany, P. and Real, L. A. (1999). Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. *Proc. Natl. Acad. Sci. USA.* **96**, 5095–5100.

Bergua, M., Zwart, M. P., El-Mohtar, C., Shilts, T., Elena, S. F. and Folimonova, S. Y. (2014). A viral protein mediates superinfection exclusion at the whole-organism level but is not required for exclusion at the cellular level. *J. Virol.* **88**, 11327–11338.

Berthier, P., Beaumont, M. A., Cornuet, J.-M. and Luikart, G. (2002). Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics* **160**, 741–751.

Betancourt, M., Fereres, A., Fraile, A. and Garcia-Arenal, F. (2008). Estimation of the Effective Number of Founders That Initiate an Infection after Aphid Transmission of a Multipartite Plant Virus. *J. Virol.* **82**, 12416–12421.

Boissot, N., Schoeny, A. and Vanlerberghe-Masutti, F. (2016). *Vat*, an amazing gene conferring resistance to aphids and viruses they carry: from molecular structure to field effects. *Front. Plant Sci.* **7**, 1-18.

Bollback, J. P., York, T. L. and Nielsen, R. (2008). Estimation of $2N_eS$ from temporal allele frequency data. *Genetics* **179**, 497–502.

Brown, J. K. M. (2015). Durable Resistance of Crops to Disease: A Darwinian Perspective. *Annu. Rev. Phytopathol.* **53**, 513–539.

Caballero, A. (1994). Developments in the prediction of effective population size. *Heredity* **73**, 657–679.

Caranta, C., Palloix, A., Lefebvre, V. and Daubeze, A. M. (1997). QTLs for a component of partial resistance to *Cucumber mosaic virus* in pepper: restriction of virus installation in host-cells. *Theor. Appl. Genet.* **94**, 431–438.

Carrasco, P., de la Iglesia, F. and Elena, S. F. (2007). Distribution of fitness and virulence effects caused by single-nucleotide substitutions in *Tobacco etch virus*. *J. Virol.* **81**, 12979–12984.

Chao, L. (1990). Fitness of RNA virus decreased by Muller's rachet. *Nature* **348**, 454–455.

Charlesworth, B. (1978). Model for evolution of Y chromosomes and dosage compensation. *Proc. Natl. Acad. Sci. USA.* **75**, 5618–5622.

Charlesworth, B. (2009). Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205.

Charlesworth, B., Charlesworth, D. and Barton, N. H. (2003). The effects of genetic and geographic structure on neutral variation. *Annu. Rev. Ecol. Evol. Syst.* **34**, 99–125.

Chattopadhyay, S., Ali, K. A., Doss, S. G., Das, N. K., Aggarwal, R. K., Bandopadhyay, T. K., Sarkar, A. and Bajpai, A. (2011). Association of leaf micro-morphological characters with powdery mildew resistance in field-grown mulberry (*Morus* spp.) germplasm. *AoB Plants* **2011**, plr002.

Crow, J. F. and Denniston, C. (1988). Inbreeding and variance effective population numbers. *Evolution* **42**, 482–495.

Cruz, C. M. V., Bai, J., Oña, I., Leung, H., Nelson, R. J., Mew, T.-W. and Leach, J. E. (2000). Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. *Proc. Natl. Acad. Sci. USA.* **97**, 13500–13505.

de la Iglesia, F. and Elena, S. F. (2007). Fitness Declines in *Tobacco Etch Virus* upon Serial Bottleneck Transfers. *J. Virol.* **81**, 4941–4947.

Duarte, E., Clarke, D., Moya, A., Domingo, E. and Holland, J. (1992). Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proc. Natl. Acad. Sci. USA.* **89**, 6015–6019.

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361.

Engelstädter, J. (2008). Muller's ratchet and the degeneration of Y chromosomes: A simulation study. *Genetics* **180**, 957–967.

Escarmís, C., Dávila, M., Charpentier, N., Bracho, A., Moya, A. and Domingo, E. (1996). Genetic lesions associated with Muller's ratchet in an RNA virus. *J. Mol. Biol.* **264**, 255–267.

Escriu, F., Fraile, A. and García-Arenal, F. (2003). The evolution of virulence in a plant virus. *Evolution* **57**, 755–765.

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3**, 87–112.

Ewens, W. (1982). On the concept of the effective population size. *Theor. Popul. Biol.* **21**, 373–378.

Ewens, W. J. (2004). Mathematical population genetics. I. Theoretical introduction. Interdisciplinary applied mathematics, 27.

Fabre, F., Chadœuf, J., Costa, C., Lecoq, H. and Desbiez, C. (2010). Asymmetrical over-infection as a process of plant virus emergence. *J. Theor. Biol.* **265**, 377–388.

Fabre, F., Montarry, J., Coville, J., Senoussi, R., Simon, V. and Moury, B. (2012). Modelling the Evolutionary Dynamics of Viruses within Their Hosts: A Case Study Using High-Throughput Sequencing. *PLoS Pathog.* **8**, e1002654.

Fabre, F., Moury, B., Johansen, E. I., Simon, V., Jacquemond, M. and Senoussi, R. (2014). Narrow bottlenecks affect *Pea seedborne mosaic virus* populations during vertical seed transmission but not during leaf colonization. *PLoS Pathog.* **10**, e1003833.

Feder, A. F., Rhee, S.-Y., Holmes, S. P., Shafer, R. W., Petrov, D. A. and Pennings, P. S. (2016). More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. *Elife* **5**, e10670.

Folimonova, S. Y. (2013). Developing an understanding of cross-protection by *Citrus tristeza virus*. *Front. Microbiol.* **4**, 76.

Foll, M., Shim, H. and Jensen, J. D. (2015). WFABC: a Wright–Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Mol. Ecol. Resour.* **15**, 87–98.

García-Arenal, F. and McDonald, B. A. (2003). An analysis of the durability of resistance to plant viruses. *Phytopathology* **93**, 941–952.

García-Arenal, F., Fraile, A. and Malpica, J. M. (2003). Variation and evolution of plant virus populations. *Int. Microbiol.* **6**, 225–232.

Gibbs, A., Gibbs, M., Ohshima, K. and Garcia-Arenal, F. (2008a). More about plant virus evolution; past, present and future. In *Origin and Evolution of Viruses,* 229–50.

González-Jara, P., Fraile, A., Canto, T. and Garcia-Arenal, F. (2009). The multiplicity of infection of a plant virus varies during colonization of its eukaryotic host. *J. Virol.* **83**, 7487–7494.

González-Jara, P., Fraile, A., Canto, T. and García-Arenal, F. (2013). The multiplicity of infection of a plant virus varies during colonization of its eukaryotic host. *J. Virol.* **87**, 2374–2374.

Gordo, I. and Charlesworth, B. (2000). On the speed of Muller's ratchet. *Genetics* **156**, 2137–2140.

Gutiérrez, S., Yvon, M., Thébaud, G., Monsion, B., Michalakis, Y. and Blanc, S. (2010). Dynamics of the multiplicity of cellular infection in a plant virus. *PLoS Pathog.* **6**, e1001113.

Gutiérrez, S., Yvon, M., Pirolles, E., Garzo, E., Fereres, A., Michalakis, Y. and Blanc, S. (2012). Circulating virus load determines the size of bottlenecks in viral populations progressing within a host. *PLoS Pathog.* **8**, e1003009.

Gutiérrez, S., Pirolles, E., Yvon, M., Baecker, V., Michalakis, Y. and Blanc, S. (2015). The Multiplicity of Cellular Infection Changes Depending on the Route of Cell Infection in a Plant Virus. *J. Virol.* **89**, 9665–9675.

Hall, J. S., French, R., Hein, G. L., Morris, T. J. and Stenger, D. C. (2001). Three distinct mechanisms facilitate genetic isolation of sympatric wheat streak mosaic virus lineages. *Virology* **282**, 230–236.

Harrison, B. D. (2002). Virus variation in relation to resistance-breaking in plants. *Euphytica* **124**, 181–192.

Jan, P., Gracianne, C., Fournet, S., Olivier, E., Arnaud, J., Porte, C., Bardou-Valette, S., Denis, M. and Petit, E. J. (2016). Temporal sampling helps unravel the genetic structure of naturally occurring populations of a phytoparasitic nematode. 1. Insights from the estimation of effective population sizes. *Evol. Appl.* **9**, 489–501.

Janzac, B., Montarry, J., Palloix, A., Navaud, O. and Moury, B. (2010). A point mutation in the polymerase of *Potato virus Y* confers virulence toward the *Pvr4* resistance of pepper and a high competitiveness cost in susceptible cultivar. *Mol. Plant. Microbe Interact.* **23**, 823–830.

Jaramillo, N., Domingo, E., Muñoz-Egea, M. C., Tabares, E. and Gadea, I. (2013). Evidence of Muller's ratchet in *Herpes simplex virus* type 1. *J. Gen. Virol.* **94**, 366–375.

Jiang, G., Peyraud, R., Remigi, P., Guidot, A., Ding, W., Genin, S. and Peeters, N. (2016). Modeling and experimental determination of infection bottleneck and within-host dynamics of a soil-borne bacterial plant pathogen. *bioRxiv*.

Jorde, P. E. and Ryman, N. (2007). Unbiased estimator for genetic drift and effective population size. *Genetics* **177**, 927–935.

Kimura, M. and Ohta, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763.

Kinkel, L. L., Bakker, M. G. and Schlatter, D. C. (2011). A coevolutionary framework for managing disease-suppressive soils. *Annu. Rev. Phytopathol.* **49**, 47–67.

Kutnjak, D., Elena, S. F. and Ravnikar, M. (2017). Time-sampled population sequencing reveals the interplay of selection and genetic drift in experimental evolution of *Potato virus Y*. *J. Virol.* JVI-00690.

Lacerda, M. and Seoighe, C. (2014). Population genetics inference for longitudinally-sampled mutants under strong selection. *Genetics*. **198**, 1237-1250.

Lafforgue, G., Tromas, N., Elena, S. F. and Zwart, M. P. (2012). Dynamics of the establishment of systemic potyvirus infection: independent yet cumulative action of primary infection sites. *J. Virol.* **86**, 12912–12922.

Lam, L. H. and Monack, D. M. (2014). Intraspecies competition for niches in the distal gut dictate transmission during persistent *Salmonell*a infection. *PLoS Pathog.* **10**, e1004527.

Leach, J. E., Vera Cruz, C. M., Bai, J. and Leung, H. (2001). Pathogen fitness penalty as a predictor of durability of disease resistance genes. *Annu. Rev. Phytopathol.* **39**, 187–224.

Lecoq, H., Lemaire, J. and Wipf-Scheibel, C. (1991). Control of *Zucchini yellow mosaic virus* in squash by cross protection. *Plant Dis.* **75**, 208–211.

Lecoq, H., Moury, B., Desbiez, C., Palloix, A. and Pitrat, M. (2004). Durable virus resistance in plants through conventional approaches: a challenge. *Virus Res.* **100**, 31–39.

Lee, J.-Y. and Lu, H. (2011). Plasmodesmata: the battleground against intruders. *Trends Plant Sci.* **16**, 201–210.

Lefevre, T., Raymond, M. and Thomas, F. (2016). *Biologie évolutive*. De Boeck Superieur.

Lequime, S., Fontaine, A., Gouilh, M. A., Moltini-Conclois, I. and Lambrechts, L. (2016). Genetic drift, purifying selection and vector genotype shape dengue virus intra-host genetic diversity in mosquitoes. *PLoS Genet.* **12**, e1006111.

Li, H. and Roossinck, M. J. (2004). Genetic Bottlenecks Reduce Population Variation in an Experimental RNA Virus Population. *J. Virol.* **78**, 10582–10587.

Lynch, M., Bürger, R., Butcher, D. and Gabriel, W. (1993). The mutational meltdown in asexual populations. *J. Hered.* **84**, 339–344.

Malaspinas, A.-S., Malaspinas, O., Evans, S. N. and Slatkin, M. (2012). Estimating allele age and selection coefficient from time-serial data. *Genetics* **192**, 599–607.

Mathieson, I. and McVean, G. (2013). Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics* **193**, 973–984.

McDonald, B. A. and Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* **40**, 349–379.

Melotto, M., Underwood, W., Koczan, J., Nomura, K. and He, S. Y. (2006). Plant stomata function in innate immunity against bacterial invasion. *Cell* **126**, 969–980.

Miyashita, S. and Kishino, H. (2010). Estimation of the size of genetic bottlenecks in cell-to-cell movement of *Soil-borne wheat mosaic virus* and the possible role of the bottlenecks in speeding up selection of variations in trans-acting genes or elements. *J. Virol.* **84**, 1828–1837.

Monsion, B., Froissart, R., Michalakis, Y. and Blanc, S. (2008). Large bottleneck size in *Cauliflower mosaic virus* populations during host plant colonization. *PLoS Pathog.* **4**, e1000174.

Moury, B., Fabre, F. and Senoussi, R. (2007). Estimation of the number of virus particles transmitted by an insect vector. *Proc. Natl. Acad. Sci. USA*. **104**, 17891–17896.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutat. Res. Mol. Mech. Mutagen.* **1**, 2–9.

Mundt, C. (2002). Use of multiline cultivars and cultivar mixtures for disease management. *Annu. Rev. Phytopathol.* **40**, 381–410.

Nei, M. and Tajima, F. (1981). Genetic drift and estimation of effective population size. *Genetics* **98**, 625–640.

Nordborg, M. and Krone, S. M. (2002). Separation of time scales and convergence to the coalescent in structured populations. In *Modern Developments in Population Genetics. The Legacy of Gustave Malécot,* 194–232. Oxford, UK: Oxford Univ. Press.

Novella, I. S., Reissig, D. D. and Wilke, C. O. (2004). Density-dependent selection in *Vesicular stomatitis virus*. *J. Virol.* **78**, 5799–5804.

Padgett, H. S., Watanabe, Y. and Beachy, R. N. (1997). Identification of the TMV replicase sequence that activates the N gene-mediated hypersensitive response. *Mol. Plant. Microbe Interact.* **10**, 709–715.

Parlevliet, J. E. (2002). Durability of resistance against fungal, bacterial and viral pathogens; present situation. *Euphytica* **124**, 147–156.

Patwa, Z. and Wahl, L. M. (2008). The fixation probability of beneficial mutations. *J. R. Soc. Interface* **5**, 1279–1289.

Poland, J. A., Balint-Kurti, P. J., Wisser, R. J., Pratt, R. C. and Nelson, R. J. (2009). Shades of gray: the world of quantitative disease resistance. *Trends Plant Sci.* **14**, 21–29.

Quenouille, J., Montarry, J., Palloix, A. and Moury, B. (2013). Farther, slower, stronger: how the plant genetic background protects a major resistance gene from breakdown: Mechanisms of polygenic resistance durability. *Mol. Plant Pathol.* **14**, 109–118.

Richardson, K., Vales, M., Kling, J., Mundt, C. and Hayes, P. (2006). Pyramiding and dissecting disease resistance QTL to barley stripe rust. *Theor. Appl. Genet.* **113**, 485–495.

Rieux, A., De Bellaire, L. D. L., Zapater, M.-F., Ravigné, V. and Carlier, J. (2013). Recent range expansion and agricultural landscape heterogeneity have only minimal effect on the spatial genetic structure of the plant pathogenic fungus *Mycosphaerella fijiensis*. *Heredity* **110**, 29–38.

Rodrigo, G., Zwart, M. P. and Elena, S. F. (2014). Onset of virus systemic infection in plants is determined by speed of cell-to-cell movement and number of primary infection foci. *J. R. Soc. Interface* **11**, 20140555.

Rousseau, E., Fabre, F., Senoussi, R., Mailleret, L., Palloix, A., Simon, V., Valière, S., Moury, B. and Grognard, F. (In press). Estimating virus effective population size and selection without neutral markers. *PLoS Pathog.*

Rouzine, I. M., Rodrigo, A. and Coffin, J. M. (2001). Transition between Stochastic Evolution and Deterministic Evolution in the Presence of Selection: General Theory and Application to Virology. *Microbiol. Mol. Biol. Rev.* **65**, 151–185.

Sacristán, S., Malpica, J. M., Fraile, A. and Garcia-Arenal, F. (2003). Estimation of Population Bottlenecks during Systemic Movement of *Tobacco mosaic virus* in Tobacco Plants. *J. Virol.* **77**, 9906–9911.

Sacristán, S., Diaz, M., Fraile, A. and Garcia-Arenal, F. (2011). Contact Transmission of *Tobacco mosaic virus*: a Quantitative Analysis of Parameters Relevant for Virus Evolution. *J. Virol.* **85**, 4974–4981.

Sommerhalder, R. J., McDonald, B. A., Mascher, F. and Zhan, J. (2011). Effect of hosts on competition among clones and evidence of differential selection between pathogenic and saprophytic phases in experimental populations of the wheat pathogen *Phaeosphaeria nodorum*. *BMC Evol. Biol.* **11**, 188.

Stefansson, T. S., McDonald, B. A. and Willi, Y. (2014). The influence of genetic drift and selection on quantitative traits in a plant pathogenic fungus. *PLoS One* **9**, e112523.

Steinrücken, M., Bhaskar, A. and Song, Y. S. (2014). A novel spectral method for inferring general diploid selection from time series genetic data. *Ann. Appl. Stat.* **8**, 2203.

Syller, J. (2016). Antagonistic within-host interactions between plant viruses: molecular basis and impact on viral and host fitness. *Mol. Plant Pathol.* **17**, 769–782.

Terhorst, J., Schlötterer, C. and Song, Y. S. (2015). Multi-locus analysis of genomic time series data from experimental evolution. *PLoS Genet.* **11**, e1005069.

Thrall, P. H., Oakeshott, J. G., Fitt, G., Southerton, S., Burdon, J. J., Sheppard, A., Russell, R. J., Zalucki, M., Heino, M. and Ford Denison, R. (2011). Evolution in agriculture: the application of evolutionary approaches to the management of biotic interactions in agro-ecosystems. *Evol. Appl.* **4**, 200–215.

Tromas, N., Zwart, M. P., Lafforgue, G. and Elena, S. F. (2014). Within-Host Spatiotemporal Dynamics of Plant Virus Infection at the Cellular Level. *PLoS Genet.* **10**, e1004186.

Van Den Berg, F., Bacaër, N., Metz, J., Lannou, C. and Van Den Bosch, F. (2011). Periodic host absence can select for higher or lower parasite transmission rates. *Evol. Ecol.* **25**, 121–137.

Wang, J. (2001). A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genet. Res.* **78**, 243–257.

Wang, J. (2005). Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. **360**, 1395–1409.

Waples, R. S. (1989). A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**, 379–391.

Webster, B., Ott, M. and Greene, W. C. (2013). Evasion of superinfection exclusion and elimination of primary viral RNA by an adapted strain of *hepatitis C virus*. *J. Virol.* **87**, 13354–13369.

Williamson, E. G. and Slatkin, M. (1999). Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics* **152**, 755–761.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.

Yeh, S. and Gonsalves, D. (1984). Evaluation of induced mutants of *Papaya ringspot virus* for control by cross protection. *Phytopathology* **74**, 1086–1091.

Yuste, E., Sánchez-Palomino, S., Casado, C., Domingo, E. and López-Galíndez, C. (1999). Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. *J. Virol.* **73**, 2745–2751.

Zhan, J., Mundt, C. C. and McDonald, B. A. (2001). Using restriction fragment length polymorphisms to assess temporal variation and estimate the number of ascospores that initiate epidemics in field populations of *Mycosphaerella graminicola*. *Phytopathology* **91**, 1011–1017.

Zhan, J., Mundt, C., Hoffer, M. and McDonald, B. (2002). Local adaptation and effect of host genotype on the rate of pathogen evolution: an experimental test in a plant pathosystem. *J. Evol. Biol.* **15**, 634–647.

Zhan, J., Thrall, P. H. and Burdon, J. J. (2014). Achieving sustainable plant disease management through evolutionary principles. *Trends Plant Sci.* **19**, 570–575.

Zhan, J., Thrall, P. H., Papaïx, J., Xie, L. and Burdon, J. J. (2015). Playing on a Pathogen's Weakness: Using Evolution to Guide Sustainable Plant Disease Control Strategies. *Annu. Rev. Phytopathol.* **53**, 19–43.

Zhu, Y., Chen, H., Fan, J. and Wang, Y. (2000). Genetic diversity and disease control in rice. *Nature* **406**, 718.

Zwart, M. P. and Elena, S. F. (2015). Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annu. Rev. Virol.* **2**, 161–179.

Zwart, M. P., Daròs, J.-A. and Elena, S. F. (2011). One Is Enough: In Vivo Effective Population Size Is Dose-Dependent for a Plant RNA Virus. *PLoS Pathog.* **7**, e1002122.

B. Le pathosystème piment-PVY

1 Le PVY

1.1 Description générale

Le *Potato virus Y* (PVY) appartient à la famille des *Potyviridae* et au genre des *Potyvirus*. Il est capable d'infecter une large gamme d'hôtes qui comprend neuf familles de plantes parmi lesquelles les Solanacées sont particulièrement représentées. Au sein de cette famille, le PVY occasionne principalement des dégâts sur pomme de terre, tabac, tomate et piment. La transmission du virus peut se faire par propagation végétative, par pucerons et plus rarement par contact. Plus de 40 espèces de pucerons sont capables de transmettre le PVY selon le mode non persistant. Cela signifie que le virus ne peut ni circuler, ni se multiplier dans son hôte, et que la durée de rétention dans le stylet du puceron est courte. Il est également possible de réaliser des inoculations mécaniques du virus en laboratoire (Kerlan 2006; Quenouille et al. 2013b).

1.2 Génome du PVY

Le génome du PVY est composé d'un ARN simple brin qui mesure environ 9,7 kb et qui est de polarité positive, ce qui signifie qu'il peut être directement traduit pour synthétiser les protéines virales. La molécule d'ARN est liée à la protéine virale VPg (viral protein genome-linked) en extrémité 5' et est polyadénylée en extrémité 3'. Le génome contient un cadre ouvert de lecture qui code pour une polyprotéine d'environ 3062 acides aminés. Cette polyprotéine est par la suite clivée en dix protéines fonctionnelles par trois protéases virales (P1, HCPro et NIa) (Figure 1). Une onzième protéine, nommée P3N-PIPO, peut être synthétisée à partir d'un autre cadre de lecture (décalage de phase +2) dans la région codant pour la protéine P3 (Chung et al. 2008). La majorité des protéines sont multifonctionnelles (voir Urcuqui-Inchima et al. (2001) et Quenouille et al. (2013b) pour revue).

1.3 Origine et classification

Le PVY appartient à un groupe comprenant 19 espèces de potyvirus infectant les Solanacées et qui sont majoritairement retrouvées en Amérique. L'Amérique du sud serait le centre d'origine et de diversification de ce groupe, ce qui coïncide avec les données obtenues sur leurs plantes hôtes qui sont principalement originaires d'Amérique centrale et du sud. La

diversification du PVY est difficile à dater de par le manque de structure temporelle dans les séquences des PVY. La radiation évolutive des *Potyvirus* s'étant probablement produite il y a 6600 ans (Gibbs et al. 2008b), la diversification de PVY est très certainement un événement récent (Quenouille et al. 2013b).

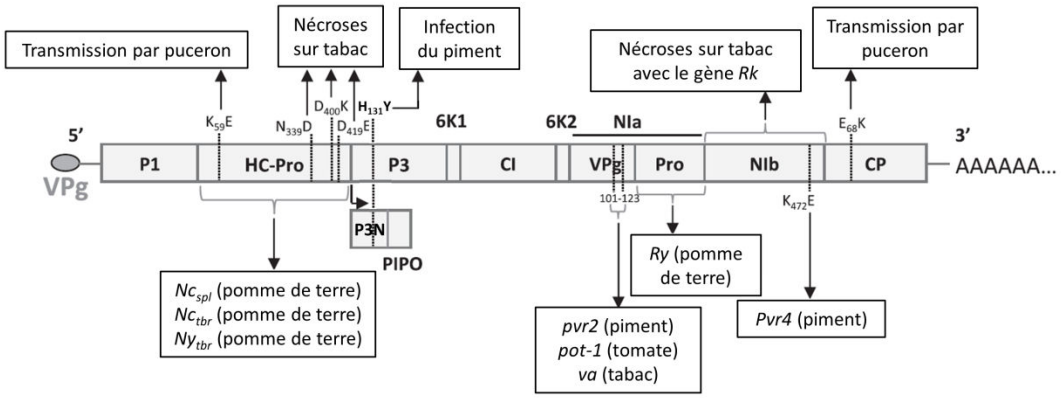


Figure 1 : Organisation génomique du *Potato virus Y* (PVY). Les protéines virales et/ou les mutations impliquées dans le contournement de gènes majeurs sont indiquées en bas du schéma. Modifié d'après Quenouille et al. (2013).

Les premiers critères utilisés pour classer les souches de PVY portaient sur les capacités d'infection et les symptômes induits par les virus sur pomme de terre (*Solanum tuberosum*) et sur tabac (*Nicotiana tabacum*). Depuis, le développement des outils de biologie moléculaire a permis de déterminer que la classification phylogénétique du PVY s'organise en 6 clades majeurs : PVY-O, PVY-N, PVY-C1, PVY-C2, PVY-Chili et PVY-Brésil (Janzac et al. 2015). Les quatre premiers clades ont été observés partout à travers le monde tandis que les deux derniers ont uniquement été décrits au Chili et au Brésil, respectivement. Cette classification est étroitement corrélée à la gamme d'hôtes des différentes souches. Par exemple, lorsque l'on compare les capacités d'infection du virus sur piment (*Capsicum* spp.) et sur pomme de terre, les souches des clades O, N et C2 infectent quasi exclusivement la pomme de terre tandis que les souches des clades C1 et Chile infectent quasi exclusivement le piment. Les souches du Brésil ne semblent quant à elles infecter que le tabac. Une étude récente suggère que l'adaptation du PVY au piment serait un caractère secondaire apparu plusieurs fois indépendamment au cours de son évolution. La capacité à infecter le piment aurait été acquise grâce à une mutation non synonyme dans la région codant pour les protéines P3 et P3N-PIPO (Vassilakos et al. 2016).

Plusieurs souches recombinantes entre les clades phylogénétiques ont également été décrites. Par exemple, les recombinants entre les clades O et N se répartissent en deux groupes principaux : PVY^{NTN} et PVY^{N}-W. Entre les années 1990 et 2000, la fréquence de ces deux souches recombinantes a considérablement augmenté en champs, leurs valeurs sélectives étant supérieures à celles des souches PVY-O et PVY-N non recombinantes (Rolland et al. 2008).

2 Le piment

2.1 Origine géographique et diversité

Le piment (*Capsicum* L.) appartient à la famille des Solanacées et est originaire d'Amérique centrale et d'Amérique du sud. Il s'agit d'une dicotylédone annuelle préférentiellement autogame. On estime que la culture du piment a débuté environ 7000 ans avant l'arrivée des Européens sur le continent américain (Djian-Caporalino et al. 2006; Perry et al. 2007). A la fin du XV^{e} siècle, le piment a été importé en Europe suite au premier voyage de Christophe Colomb. Il a ensuite été largement propagé en Afrique et en Asie et, près de deux siècles plus tard, il était cultivé partout à travers le monde (Sage-Palloix et al. 2007). Le piment est une plante d'intérêt commercial dont les usages sont multiples. Majoritairement utilisé pour l'alimentation, à la fois sous forme de fruit et de condiment, il est également employé dans la fabrication de colorants, dans l'industrie pharmaceutique ou encore comme plante ornementale. En 2011, la production mondiale de piment s'élevait à 34,6 millions de tonnes de fruits frais récoltés sur 3,9 millions d'hectares (www.fao.org).

Le genre *Capsicum* comprend au moins 35 espèces parmi lesquelles 5 espèces ont été domestiquées : *Capsicum annuum*, *Capsicum baccatum*, *Capsicum chinense*, *Capsicum frutescens* et *Capsicum pubescens* (Moscone et al. 2006). La domestication et la sélection engendrées par des conditions de cultures variées à travers le monde ont conduit à une forte diversification secondaire des types de fruits. Il en résulte une grande diversité de forme, de taille et de couleur des fruits entre ces espèces. *C. annuum* est l'espèce la plus cultivée à l'échelle mondiale et qui possède le plus grand nombre de variétés différentes, de l'ordre de plusieurs milliers. *C. chinense* est mieux adaptée aux climats chauds et humides et est majoritairement cultivée en Afrique centrale et de l'ouest. Bien que moins largement cultivée, *C. frutescens* est retrouvée sur la plupart des continents. Enfin, les espèces *C. baccatum* et *C. pubescens* sont principalement cultivées en Amérique latine (Djian-Caporalino et al. 2006).

2.2 Génome du piment

Les membres du genre *Capsicum* sont des espèces diploïdes qui possèdent 12 paires de chromosomes (2n = 24). En 2014, deux équipes de recherche ont publié indépendamment la séquence complète du génome du piment (Kim et al. 2014; Qin et al. 2014). L'équipe de Kim et al. (2014) a réalisé le séquençage *de novo* des génomes de *C. annuum* cv. Criollo de Morelos 334 (CM334), une variété très utilisée en recherche et en création variétale, et d'une espèce sauvage de *C. chinense*. Ils ont également effectué le reséquençage de deux variétés cultivées, c'est-à-dire l'alignement des séquences de ces variétés contre le génome de référence qu'ils ont produit afin de comparer les génomes de plusieurs espèces de piment. L'équipe de Qin et al. (2014) a quant à elle réalisé le séquençage *de novo* du génome de *C. annuum* cv. Zunla-1, une variété cultivée, et de son progéniteur sauvage chiltepin, ainsi que le reséquençage de 20 accessions comprenant des *C. annuum* et des *C. chinense*.

Les tailles des génomes estimées sont de 3,48 Gb pour CM334 et de 3,26 Gb pour Zunla-1, soit un génome trois à quatre fois plus grand que celui d'autres Solanacées comme la tomate *Solanum lycopersicum* ou la pomme de terre *Solanum tuberosum* (Figure 2a). En revanche, le nombre de gènes prédits est très proche, avec respectivement 34903 et 35336 gènes prédits par les deux études sur le piment, contre 34711 pour la tomate et 39031 pour la pomme de terre (Potato Genome Sequencing Consortium 2011; Tomato Genome Consortium 2012). Les deux études attribuent l'expansion génomique du piment à la présence d'un nombre important de séquences répétées dans l'ensemble des chromosomes, représentant entre 76,4 et 81,0 % du génome complet, soit environ 2,34 à 2,70 Gb. Ces valeurs sont supérieures à celles de la tomate et de la pomme de terre, chez qui la proportion de séquences répétées avoisine les 60 % du génome. Ainsi, lorsqu'on compare les chromosomes du piment et de la tomate, son plus proche parent chez les Solanacées, on constate la présence de nombreuses régions de synténie mais également une forte accumulation de séquences répétées dans le génome du piment, à la fois dans l'euchromatine et l'hétérochromatine (Figure 2b). La majorité de ces séquences répétées sont des rétrotransposons à LTR (long terminal repeats), parmi lesquels 87,7 % sont des éléments de type *Gypsy* et 7,4 % des éléments de type *Copia* (Figure 2c). Des séquences appartenant à des pararétrovirus de la famille des Caulimoviridae ont également été retrouvées dans des proportions 9,2 fois supérieures à ce qui est décrit chez la tomate.

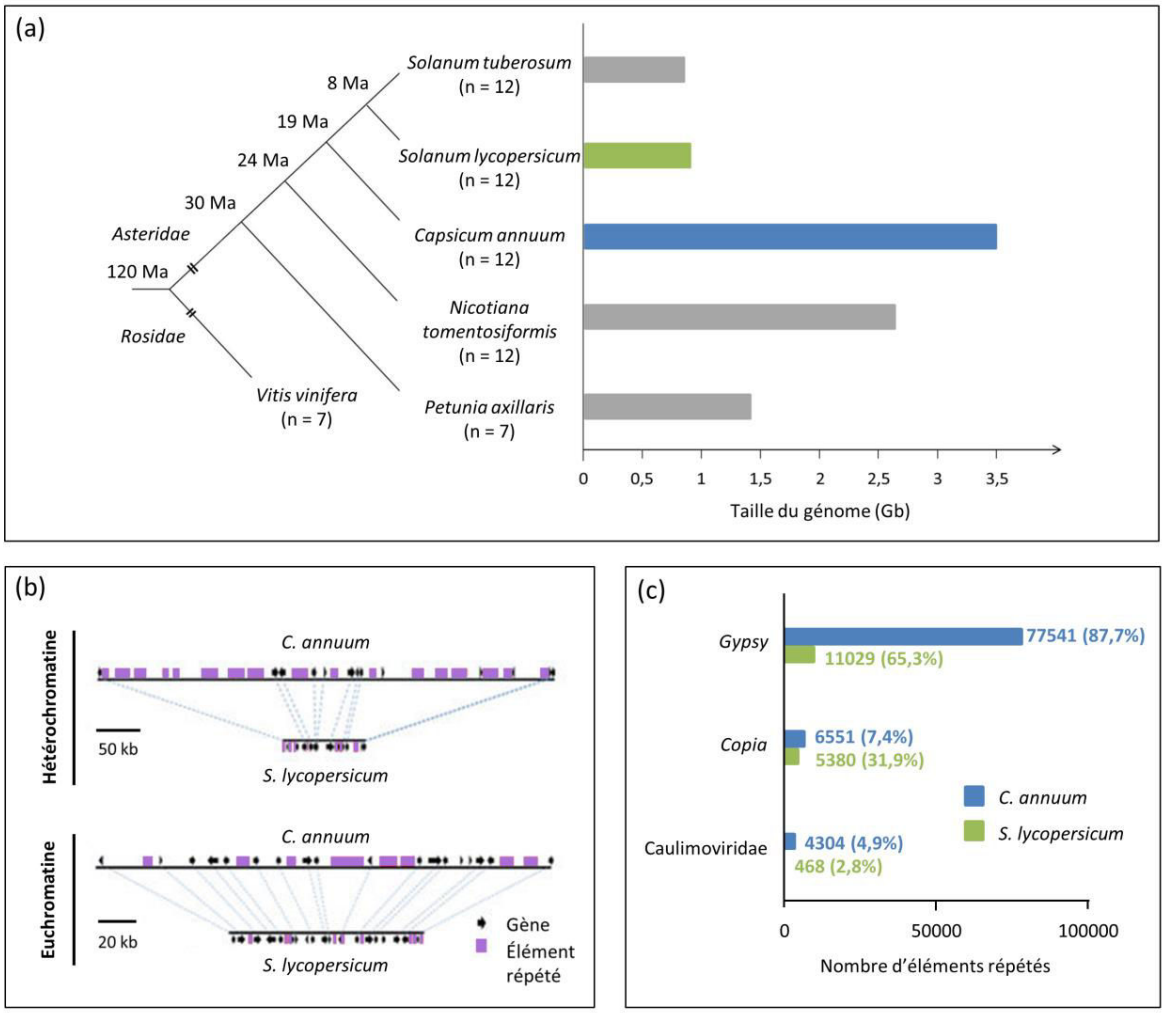


Figure 2 : Description de l'expansion génomique survenue chez *Capsicum* spp. (a) Arbre phylogénétique des Solanacées (d'après Bombarely et al. (2016)) et taille estimée du génome chez cinq représentants appartenant à quatre genres différents (Potato Genome Sequencing Consortium 2011; Tomato Genome Consortium 2012; Sierro et al. 2013; Kim et al. 2014; Bombarely et al. 2016). Le piment est illustré en bleu et la tomate, son plus proche parent chez les Solanacées, est en vert. **(b)** Comparaison du chromosome 2 du piment (barre du haut) et de la tomate (barre du bas), à la fois pour l'hétérochromatine et l'euchromatine. La position des gènes orthologues est indiquée par les lignes reliant les deux génomes. Modifié d'après Kim et al. (2014). **(c)** Comparaison de la proportion d'éléments répétés de type *Gypsy*, *Copia* et de séquences de Caulimoviridae entre les génomes du piment (bleu) et de la tomate (vert). Modifié d'après Kim et al. (2014).

Les régions génomiques répétées ne concernent cependant pas que les séquences non codantes. Plusieurs gènes impliqués dans la résistance aux maladies et les fonctions cellulaires ont subi des événements de duplication que l'on ne retrouve pas chez les autres Solanacées. Par exemple, 684 gènes ont été identifiés comme codant pour des protéines de type NBS-LRR (nucleotide-binding site-leucine-rich repeat), souvent impliquées dans la résistance aux pathogènes. A titre de comparaison, on ne dénombre que 267 gènes de ce type chez la tomate. Le cas du gène *Bs2* illustre bien cette expansion des NBS-LRR. Ce gène confère la résistance

à *Xanthomonas* spp. et possède trois orthologues chez la tomate et un orthologue chez la pomme de terre. Chez le piment, 89 copies regroupées sur le chromosome 9 ont été identifiées (Kim et al. 2014).

2.3 Résistances aux potyvirus

La vaste dispersion géographique des *Capsicum* spp. à travers le monde a confronté ce genre à de nombreux pathogènes. Concernant les virus, plus de 20 virus appartenant à 15 groupes taxonomiques différents sont connus pour infecter les cultures de piment (Djian-Caporalino et al. 2006). Les membres du genre potyvirus sont les plus représentés, avec 10 espèces décrites (Moury et al. 2005; Janzac et al. 2008). La coévolution entre piment et potyvirus a engendré la sélection de gènes de résistance chez le piment. On estime que 40% des accessions de piment sont résistantes aux souches communes de PVY, de nombreux facteurs de résistance différant dans leur spectre d'action, leur niveau de résistance et leur durabilité ayant été décrits (Charron et al. 2008). Comme énoncé dans la première partie de l'introduction, ces résistances peuvent être divisées en deux catégories en fonction de la ségrégation du caractère lié à la résistance dans la population étudiée : les résistances qualitatives et les résistances quantitatives. On parle de résistance qualitative lorsque le caractère ségrège de manière discrète, les génotypes de plantes étant soit sensibles, soit résistants au pathogène, et de résistance quantitative lorsque le caractère ségrège de manière continue, des plantes avec des niveaux de résistance intermédiaires étant alors observées. Les résistances qualitatives sont généralement sous le contrôle d'un gène majeur, et les résistances quantitatives sous le contrôle de plusieurs gènes.

2.3.1 Les résistances qualitatives

Plusieurs gènes majeurs conférant une résistance qualitative aux potyvirus ont été décrits chez le piment. Tout d'abord, le gène de résistance dominant *Pvr4*, localisé sur le chromosome 10, confère une résistance à tous les pathotypes de PVY connus ainsi qu'à l'*Ecuadorian rocoto virus* (ERV), au *Pepper mottle virus* (PepMoV), au *Pepper yellow mosaic virus* (PepYMV) et au *Pepper severe mosaic virus* (PepSMV) (Dogimont et al. 1996; Janzac et al. 2008, 2009a). Il a été cloné récemment et code pour une protéine de type NBS-LRR (Kim et al. 2017). De nombreuses résistances dominantes conférées par des protéines à NBS-LRR résultent de l'interaction entre le facteur de résistance de la plante et le facteur d'avirulence du virus déclenche une réaction de type « hypersensibilité » (HR) ou de type « résistance extrême » (ER). Dans le cas du PVY, la région codant pour l'ARN polymérase ARN dépendante (NIb)

constitue le facteur d'avirulence vis-à-vis de *Pvr4*. Ce gène de résistance s'est révélé être extrêmement durable malgré des dizaines d'années de déploiement en champs. Il a été démontré qu'une seule mutation dans la NIb du virus permet de contourner le gène majeur, mais que le coût d'adaptation à la résistance sur la valeur sélective est trop important, expliquant ainsi la forte durabilité de cette résistance (Janzac et al. 2010). Un second gène de résistance dominant, *Pvr7*, avait été identifié sur le chromosome 10 comme étant étroitement lié au gène *Pvr4* et conférant une résistance au PepMoV (Grube et al. 2000). Récemment, la cartographie fine de cette région génomique a révélé que *Pvr4* et *Pvr7* seraient en fait un seul et même gène (Venkatesh et al. 2017).

Des résistances récessives aux potyvirus ont également été décrites chez le piment. La résistance récessive provient de la perte d'interaction entre le facteur de l'hôte nécessaire à l'infection virale et le facteur de virulence du virus. Le gène de résistance récessif *pvr2*, localisé sur le chromosome 4, est impliqué dans la résistance au PVY et au *Tobacco etch virus* (TEV). Chez *C. annuum*, on dénombre 34 allèles de résistance à ce locus, allant de $pvr2^{1}$ à $pvr2^{34}$, ainsi qu'un allèle de sensibilité nommé $pvr2^{+}$. La protéine pour laquelle codent ces allèles appartient à la famille des eIF4E (eukaryotic translation initiation factor 4E). Il s'agit d'un facteur d'initiation de la traduction eucaryote qui interagit avec la coiffe des ARN messagers de la plante pour initier leur traduction (Ruffel et al. 2002, 2004). Chez un génotype de plante porteur de l'allèle $pvr2^{+}$, le virus détourne ce mécanisme en recrutant le facteur eIF4E avec sa VPg pour initier la traduction et/ou la réplication de l'ARN viral. Lorsque les plantes sont porteuses de deux copies d'un allèle de résistance au locus *pvr2*, l'interaction physique entre les protéines est rompue et l'infection n'a pas lieu. Néanmoins, plusieurs mutations dans la partie centrale de la VPg permettant le rétablissement de l'interaction, et donc le contournement de la résistance, ont été mises en évidence. La durabilité de *pvr2* est variable selon les allèles au locus, les allèles $pvr2^{1}$ et $pvr2^{3}$ présentant une durabilité plus faible que l'allèle $pvr2^{2}$. La durabilité dépend à la fois du nombre de mutations dans la VPg nécessaires à l'acquisition de la virulence ainsi que du coût sur la valeur sélective du virus associé à la mutation de contournement (Ayme et al. 2006, 2007). De nombreuses mutations dans la VPg permettent au virus de contourner $pvr2^{1}$ et $pvr2^{3}$, tandis que les mutations permettant le contournement de $pvr2^{2}$ sont plus rares. De plus, chez certaines souches de PVY, une seule substitution d'acide aminé est suffisante pour contourner $pvr2^{1}$ et $pvr2^{3}$, alors que deux substitutions sont nécessaires pour contourner $pvr2^{2}$.

D'autres types de résistances récessives ont été décrits, comme le gène *pvr3* qui confère au piment la résistance au PepMoV (Murphy et al. 1998). Enfin, le gène *pvr6* localisé sur le

chromosome 3 code pour un eIF(iso)4E, qui est une isoforme de l'eIF4E codé par *pvr2*. Lorsqu'il est associé à certains allèles de résistance au locus *pvr2*, la plante est résistante au *Pepper veinal mottle virus* (PVMV) et au *Chili veinal mottle virus* (ChiVMV) (Moury et al. 2005; Rubio et al. 2009).

2.3.2 Les résistances quantitatives

Des résistances partielles au PVY ont également été identifiées chez certaines accessions de piment (Caranta and Palloix 1996). Ces résistances quantitatives ont pu être étudiées plus en détails grâce à la création d'une population de lignées haploïdes doublées (HD) de *C. annuum*. Ces lignées sont issues de l'hybride F_1 obtenu par le croisement de Perennial, le parent résistant porteur de l'allèle $pvr2^3$, et de Yolo Wonder, le parent sensible porteur de l'allèle $pvr2^+$. Les lignées ségrègent donc pour les allèles des deux parents au locus *pvr2*, une partie de la population étant porteuse de $pvr2^+$ et l'autre partie du gène majeur $pvr2^3$. Les premiers travaux effectués sur cette descendance ont identifié onze régions génomiques (QTL) impliquées dans la résistance au PVY (Caranta et al. 1997a). Depuis, des QTL contrôlant la fréquence de contournement du gène majeur $pvr2^3$, le niveau d'accumulation du PVY et la sévérité des symptômes ont été cartographiés (Quenouille et al. 2014). Bien qu'ils confèrent une résistance partielle à la plante, la résistance conférée par les QTL seuls n'est pas pour autant plus durable en l'absence de gène majeur. Une efficacité moindre au cours du temps, généralement qualifiée d'érosion, ou un contournement de ces résistances ont été observés chez plusieurs espèces (Chain et al. 2007; Le Guen et al. 2007; Dowkiw et al. 2010; Caffier et al. 2014). Concernant le piment, Montarry et al. (2012) ont montré qu'une adaptation du PVY aux lignées HD porteuses uniquement de résistances quantitatives (sans $pvr2^3$) était possible. Ils ont effectué de l'évolution expérimentale à partir d'un clone infectieux de PVY sur différents génotypes de piment. Le clone infectieux était initialement adapté au génotype sensible Yolo Wonder et des passages répétés sur ce génotype n'ont pas montré de gain d'accumulation virale chez les populations finales, laissant penser que le clone de départ ne peut plus augmenter sa valeur sélective dans Yolo Wonder. Un de leur traitement a consisté à effectuer des passages en série de ce clone sur la lignée HD223 porteuse de 3 QTL conférant une résistance partielle au virus. A la fin de l'expérience, les lignages viraux s'accumulaient 12 fois plus dans HD223 que la population d'origine (Figure 3). De plus, leur accumulation virale dans ce génotype était beaucoup plus élevée que celle de la population de départ dans le parent sensible Yolo Wonder. Les lignages ayant évolué sur HD223 ont donc non seulement contourné la résistance mais ont également acquis un niveau d'adaptation supérieur à celui qu'ils avaient précédemment atteint

dans le génotype sensible. Il s'agit donc d'un cas de contournement qui rend la résistance quantitative totalement inefficace, et pas seulement d'un phénomène d'érosion. Par ailleurs, un coût d'adaptation sur la valeur sélective du virus a également été démontré. Ainsi, les lignages finaux ayant évolué sur HD223 montraient une accumulation virale significativement plus faible dans le génotype sensible Yolo Wonder que la population d'origine.

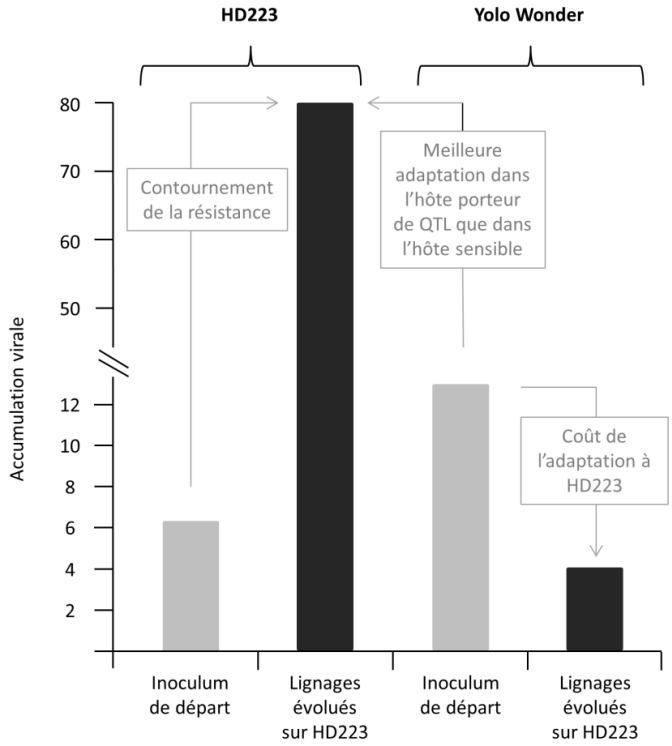


Figure 3 : Contournement d'une résistance quantitative chez le piment par le PVY. Accumulations virales obtenues dans la lignée HD223 (porteuse de QTL de résistance) et dans le génotype sensible Yolo Wonder (sans QTL), pour l'inoculum de départ (gris) ainsi que pour les populations virales ayant évoluées dans HD223 (noir). Modifié d'après Montarry et al. (2012).

2.3.3 La combinaison gène majeur/QTL

La combinaison d'un gène majeur avec des QTL de résistance représente une alternative prometteuse pour augmenter la durabilité du gène majeur. Plusieurs études ont d'ores et déjà prouvé l'efficacité de cette stratégie chez différents pathosystèmes. Brun et al. (2010) ont mené une expérience de 5 ans en champs au cours de laquelle ils ont inoculé une population de *Leptosphaeria maculans* à plusieurs génotypes de colza (*Brassica napus*). L'expérience comprenait notamment un génotype porteur uniquement du gène majeur *Rlm6* et un génotype porteur de la combinaison du gène *Rlm6* et des QTL. Au bout de trois ans, le gène majeur a été contourné chez le génotype porteur de *Rlm6* associé à un fonds génétique sensible. En revanche,

après cinq ans de culture, le gène majeur n'était toujours pas contourné dans le génotype porteur de la combinaison gène majeur/QTL, démontrant l'impact bénéfique du fonds génétique résistant sur la durabilité du gène majeur. Une seconde étude a été réalisée chez le pathosystème nématode/pomme de terre. Fournet et al. (2013) ont étudié la durabilité de *GpaV*, un QTL de résistance à effet fort responsable de la masculinisation des populations de nématodes. Ils ont réalisé des passages en série de populations de *Globodera pallida* sur quatre génotypes de pomme de terre porteurs du QTL à effet fort mais différant dans leurs fonds génétiques. Après huit passages effectués en champs puis en serre, ils ont montré que l'effet du QTL était aboli chez un seul génotype de pomme de terre et que les autres génotypes étaient probablement porteurs de facteurs de résistances additionnels ayant protégé le QTL du contournement.

Concernant le piment, Palloix et al. (2009) ont évalué la durabilité du gène majeur $pvr2^3$ au PVY. Ils ont inoculé un clone infectieux avirulent de PVY à deux génotypes de piment : Perennial, porteur du gène majeur $pvr2^3$ dans un fonds génétique résistant, et la lignée HD285, porteuse du gène majeur $pvr2^3$ dans un fonds génétique sensible. Cinq semaines après inoculation, le contournement du gène majeur a été observé chez 23% des plantes inoculées appartenant à la lignée HD285. Ces plantes présentaient des symptômes, étaient positives au test ELISA et des mutations conférant la capacité à contourner $pvr2^3$ ont été détectées par séquençage dans la VPg des populations virales. Sur les 471 plantes de type Perennial inoculées, aucune n'a montré de symptôme et aucun virus n'a été détecté en ELISA. Le fonds génétique a donc une nouvelle fois protégé le gène majeur du contournement. Quenouille et al. (2014) ont étudié plus en détails l'impact du fonds génétique sur la capacité de PVY à contourner le gène majeur $pvr2^3$. Pour ce faire, ils ont utilisé la chimère CI, un recombinant entre deux clones de PVY. La chimère possède le génome du clone SON41p et la région qui code pour la protéine CI (cylindrical inclusion) du clone LYE84.2. Sa capacité à contourner $pvr2^3$ est multipliée par 6 comparée au clone SON41p (Montarry et al. 2011). Cette faculté de contournement accrue s'explique non pas par l'acquisition de mutations supplémentaires dans la région codant pour la CI, mais par une sélection plus rapide des variants virulents mutés dans la VPg en présence de la CI de LYE84.2. Quenouille et al. (2014) ont inoculé la chimère à 153 lignées HD issues du croisement entre Perennial et Yolo Wonder. Toutes les lignées étaient porteuses de $pvr2^3$ mais différaient dans leurs fonds génétiques. Le taux de contournement du gène majeur s'est avéré être un caractère extrêmement variable au sein des lignées HD, allant de 0 à 90 % de plantes contournées un mois après inoculation. L'ensemble de ces expériences démontre que le fonds génétique de la plante a donc une influence majeure dans l'évolution des populations virales.

3 Un modèle biologique pour l'étude de la durabilité des résistances

Les nombreuses connaissances acquises sur la diversité et la durabilité des gènes majeurs chez le piment, la possibilité d'obtenir des variants viraux de PVY contournant le gène majeur en laboratoire ou encore la possibilité d'intégrer les mutations causales dans des clones infectieux pour faire de la génétique inverse font du pathosystème piment-PVY un modèle particulièrement adapté à l'étude de la durabilité des résistances. Nous avons vu que la combinaison d'un gène majeur avec des QTL représente une stratégie prometteuse pour réaliser un contrôle génétique durable des populations virales. De ce fait, les dernières études réalisées sur ce pathosystème se sont focalisées sur la compréhension des mécanismes mis en place par le fond génétique de la plante pour influencer l'évolution des virus.

3.1 Mécanismes d'action des QTL pour orienter l'évolution des populations virales

Quenouille et al. (2013a) ont démontré qu'au moins trois mécanismes permettent d'expliquer l'impact des QTL sur l'évolution des populations de PVY. Ils ont tout d'abord montré que le fond génétique confère une résistance additionnelle à celle du gène majeur en réduisant davantage l'accumulation virale dans la plante. Pour ce faire, ils ont mesuré le taux de contournement de *pvr2*^{3} dans 14 lignées HD inoculées avec la chimère CI de PVY, comme expliqué dans la partie précédente. Lors d'une seconde expérience, ces lignées HD ont été inoculées avec un variant de la chimère CI porteur de la mutation 119N[2] qui permet au virus de contourner *pvr2*^{3}. Ils ont alors mesuré le niveau d'accumulation virale du variant porteur de la mutation 119N au sein des 14 lignées, ce qui leur a permis de déterminer l'impact des QTL sur le niveau de résistance au virus en éliminant l'effet du gène majeur. Le taux de contournement s'est révélé être fortement corrélé au niveau d'accumulation virale induit par le fonds génétique de la plante (Figure 4), démontrant qu'une accumulation virale réduite grâce aux QTL permet de diminuer la fréquence de contournement du gène majeur.

[2] La nomenclature utilisée pour désigner les mutations non synonymes correspond à la position de l'acide aminé dans le génome du virus suivie du nom abrégé du nouvel acide aminé.

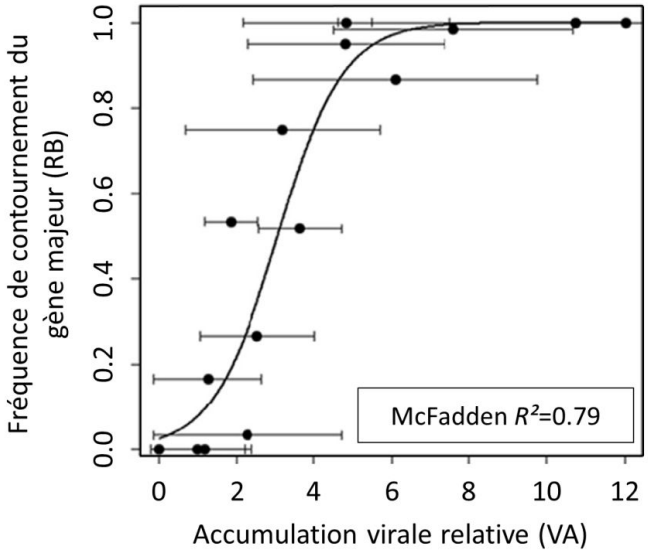


Figure 4 : Lien entre la fréquence de contournement du gène majeur *pvr2*^{3} et l'accumulation de PVY au sein de 14 lignées HD de piment et du génotype résistant Perennial. La fréquence de contournement (RB pour « resistance breaking ») a été mesurée avec la chimère CI de PVY sur 60 plantes par génotype et l'accumulation virale (VA pour « viral accumulation ») a été mesurée par ELISA quantitatif sur 10 plantes par génotype inoculées avec un variant de la chimère CI contournant *pvr2*^{3}. Les barres associées aux points indiquent les écart-types de chaque valeur et la courbe représente le modèle logistique ajusté aux données. Modifié d'après Quenouille et al. (2013a).

Une détection de QTL a par la suite été réalisée pour identifier les facteurs génétiques à l'origine de ce mécanisme. Des mesures du taux de contournement du gène de résistance (RB pour « resistance breaking »), d'accumulation virale (VA pour « viral accumulation ») ainsi que de la sévérité des symptômes (AUDPC pour « area under the disease progress curve ») ont été réalisées chez 153 lignées HD porteuses de *pvr2*^{3} (Quenouille et al. 2014). L'analyse a permis d'identifier trois QTL liés au taux de contournement, deux QTL liés à l'accumulation virale, trois QTL liés à l'AUDPC ainsi que des effets d'épistasie entre certains QTL (Figure 5). De plus, plusieurs QTL d'accumulation virale et d'AUDPC colocalisent avec des QTL liés au taux de contournement. Les effets alléliques sont cohérents entre ces QTL, puisque les allèles qui diminuent le taux de contournement diminuent également l'accumulation virale et l'AUDPC. Ces colocalisations entre QTL sont donc compatibles avec des effets de pléiotropie.

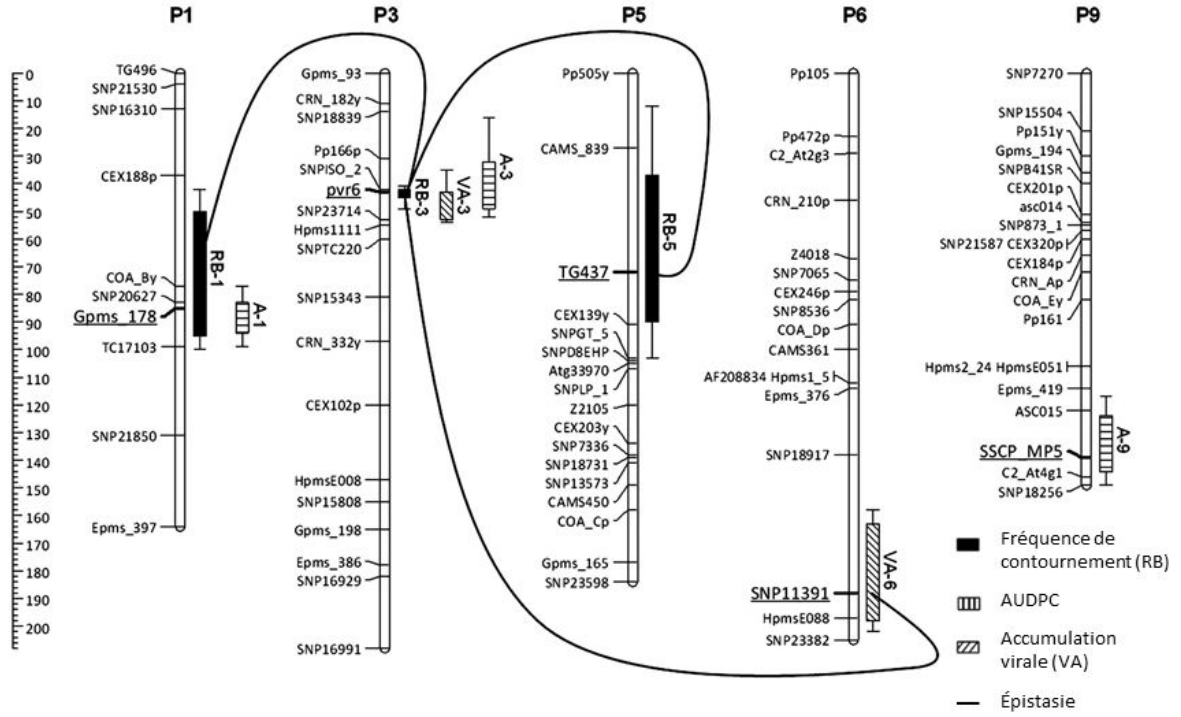


Figure 5 : Cartographie des QTL contrôlant la fréquence de contournement du gène majeur *pvr2*^{3} (RB), l'accumulation virale (VA) et la sévérité des symptômes (AUDPC) dans une population haploïde doublée de piment inoculée avec le PVY. Pour chaque QTL, la taille des rectangles représente les bornes de l'intervalle de confiance à 95% calculées en abaissant le score obtenu par le logiciel MCQTL de 1 par rapport au score maximum, et la taille des barres verticales représente l'intervalle de confiance à 95% calculé en abaissant le score de 2 par rapport au score maximum. Les traits noirs indiquent les effets d'épistasie entre les loci. Modifié d'après Quenouille et al. (2014).

Le second mécanisme proposé par Quenouille et al. (2013a) pour expliquer l'impact des QTL sur l'orientation de l'évolution des virus concerne la complexité du chemin mutationnel nécessaire au contournement du gène majeur. Pour étudier ce mécanisme, deux génotypes de piment ont été inoculés avec SON41p, un variant de PVY ne contournant pas *pvr2*^{3}. Le premier génotype était Perennial, porteur de *pvr2*^{3} dans un fonds génétique résistant, et le second la lignée HD285, porteuse de *pvr2*^{3} dans un fonds génétique sensible. Etant donné que la résistance conférée par Perennial ne permet pas de maintenir le variant non contournant dans la plante, les inoculations ont été réalisées par greffage pour les deux génotypes. Yolo Wonder a été utilisé comme porte-greffe sensible et a permis de fournir une source d'inoculum constante à Perennial et à HD285. Six semaines après inoculation, 72 % des plantes de la lignée HD285 étaient positives en ELISA. Le séquençage de la VPg des virus a révélé la présence de quatre variants viraux dans ces plantes, tous porteurs d'une mutation non synonyme (G, K, N ou C). Il avait déjà été démontré que chacune de ces mutations est suffisante pour contourner le gène

majeur *pvr2*^{3} (Ayme et al. 2006). A l'opposé, aucune plante n'a été détectée positive en ELISA chez Perennial six semaines après inoculation. Néanmoins, trois mois après inoculation, 2 plantes de génotype Perennial sur les 91 inoculées ont présenté des symptômes. Après séquençage, deux variants viraux porteurs chacun de deux mutations non synonymes ont été identifiés (GK et KN). La combinaison de deux mutations dans la VPg serait donc nécessaire au virus pour s'adapter à une plante porteuse d'un gène majeur et de QTL. Afin de tester cette hypothèse, cinq clones infectieux de PVY porteurs des mutations seules (G, K et N) et des mutations combinées (GK et KN) ont été créés. La capacité d'infection des cinq clones sur les deux génotypes a ensuite été mesurée. Pour HD285, les cinq variants étaient capables d'infecter 100 % des plantes inoculées. Pour Perennial, des capacités d'infection très différentes selon les mutants ont été décrites : les mutants K, GK et KN infectaient entre 41 et 46 % des plantes tandis que les mutants G et N n'infectaient respectivement que 9 et 3 % des plantes. L'absence du mutant K chez Perennial, alors que sa capacité d'infection est similaire à celle des deux doubles mutants GK et KN, s'explique par la nature des substitutions nucléotidiques de chaque variant. La substitution correspondant à la mutation K est une transversion, alors que celles correspondant aux mutations G et N sont des transitions. Or, les transitions sont en moyenne 5 à 8 fois plus fréquentes que les transversions chez le PVY (Ayme et al. 2006). Les virus inoculés dans Perennial ont donc d'abord dû acquérir une mutation fréquente comme G ou N, ce qui leur a permis de s'accumuler un peu plus dans la plante. Dans un second temps, ils ont acquis la mutation K qui leur a permis de s'adapter complétement à la combinaison du gène majeur et des QTL. Le chemin mutationnel pour s'adapter à la plante est donc plus complexe que lorsque le virus est uniquement confronté à un gène majeur et n'a besoin que d'une seule mutation pour contourner la résistance.

Enfin, le troisième mécanisme repose sur la fixation moins fréquente des variants qui contournent le gène majeur lorsqu'ils sont en présence du gène majeur et des QTL. Quatre expériences mettant en compétition les doubles mutants et les simples mutants de PVY correspondants ont été réalisées à l'aide de clones infectieux porteurs des différentes mutations (KN *vs* N, KN *vs* K, GK *vs* G et GK *vs* K). Pour chaque compétition, les mutants ont été inoculés en proportion 1:1 dans différents génotypes de plantes, notamment Perennial, porteur de *pvr2*^{3} dans un fonds génétique résistant, et la lignée HD285, porteuse de *pvr2*^{3} dans un fonds génétique sensible. Un mois après inoculation, la fréquence des plantes infectées uniquement par le simple mutant, uniquement par le double mutant ou par les deux types de mutants a été calculée. Seule la compétition entre les mutants KN et N n'a pas mis en évidence de différence

de compétitivité entre les mutants, quel que soit le génotype de plante. Pour les trois autres compétitions, le double mutant est en moyenne plus souvent fixé que le simple mutant dans la lignée HD285, ce qui signifie que le double mutant est plus compétitif que le simple mutant en présence du gène majeur seul. En revanche, la fixation du double mutant est significativement diminuée en présence du gène majeur et des QTL (Figure 6). Bien que les doubles mutants soient les variants viraux les plus adaptés au gène majeur, la présence des QTL permet donc de diminuer leur fixation.

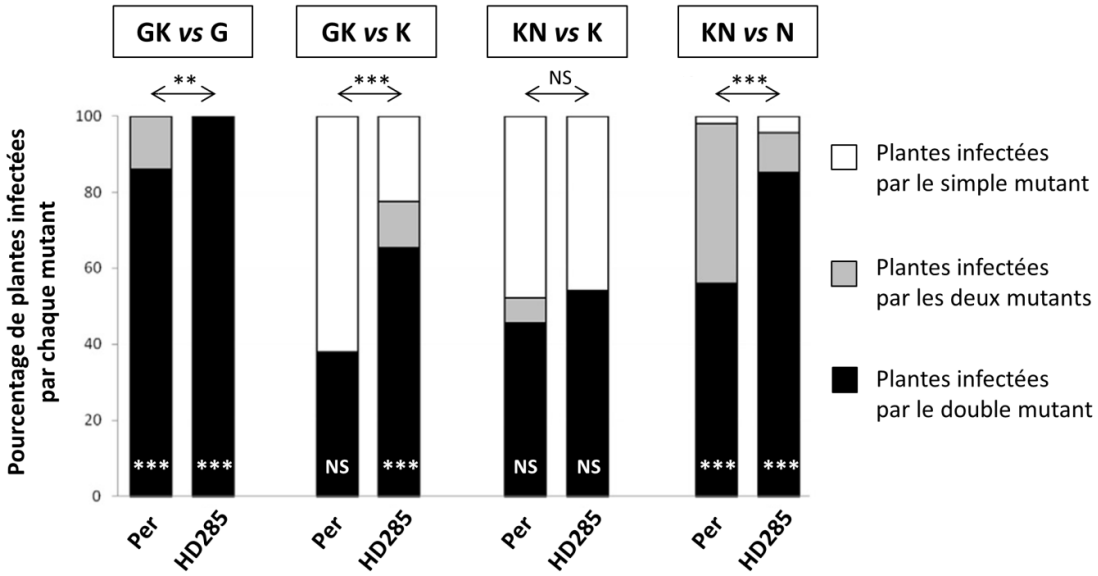


Figure 6 : Compétitivité des doubles mutants dans la VPg du PVY face aux simples mutants correspondants dans le génotype Perennial (Per), porteur du gène majeur *pvr2*^{3} dans un fonds génétique résistant, et dans la lignée HD285, porteuse du gène majeur *pvr2*^{3} dans un fonds génétique sensible. Les symboles blancs ** et *** dans les histogrammes représentent des différences significatives entre la fréquence du simple mutant (blanc) et du double mutant (noir) au sein de chaque génotype de plante comparé au ratio 1:1 de l'inoculum de départ, à un seuil d'erreur de type I de 1% et 0,1% respectivement (test exact de Fisher). Les symboles noirs ** et *** au-dessus des histogrammes correspondent à des différences significatives entre génotypes de plante pour les distributions de plantes infectées par le simple mutant (blanc), le double mutant (noir) ou les deux à la fois (gris), à un seuil d'erreur de type I de 1% et 0,1% respectivement (test exact de Fisher). NS : pas de différence significative (p > 5%). Modifié d'après Quenouille et al. (2013a).

Pour finir, les mécanismes décrits ci-dessus peuvent également être liés aux forces évolutives imposées par les QTL aux populations virales. Concernant le premier mécanisme, la diminution d'accumulation virale induite par les QTL pourrait diminuer la taille efficace de population et augmenter les effets de dérive génétique au sein de la plante. Des mutations conférant la capacité de contourner le gène majeur pourraient être perdues par effet du hasard, et le taux de contournement du gène majeur serait ainsi diminué. Pour le second mécanisme, la mutation permettrait de protéger le gène majeur. Elle jouerait un rôle à la fois de par le nombre

d'événements mutationnels plus faible dans les populations à faible accumulation virale, mais également de par la nature des mutations nécessaires à l'apparition des doubles mutants, les transversions étant plus rares que les transitions. Enfin, pour le troisième mécanisme, la plus faible fixation des variants viraux les plus adaptés au gène majeur en présence de QTL pourrait être expliquée par des effets de dérive génétique et de sélection. Les plantes porteuses uniquement du gène majeur pourraient induire une plus forte pression de sélection sur les populations virales, ce qui conduirait à la fixation plus fréquente des doubles mutants. A l'opposé, les plantes combinant un gène majeur et des QTL induiraient des pressions de sélection plus faibles et/ou une plus forte dérive génétique, diminuant ainsi la fixation des doubles mutants.

3.2 Mesure des forces évolutives imposées par la plante aux populations virales

Afin d'étudier plus en détail les hypothèses énoncées précédemment sur l'implication de chacune des forces évolutives dans la durabilité du gène majeur, il est nécessaire de mesurer précisément leur impact sur les populations virales.

Pour la mutation, il est très difficile d'avoir accès au taux intrinsèque de mutation d'un virus, qui correspond au nombre de substitutions par nucléotide et par événement de réplication. La principale difficulté rencontrée consiste à parvenir à distinguer le nombre total de mutations produites à chaque génération (= le taux de mutation) et le nombre de mutations fixées après l'action des forces évolutives à chaque génération (= le taux de substitution). Le taux de mutation est un paramètre d'intérêt en évolution puisqu'il renseigne sur la fréquence à laquelle de la diversité génétique est générée dans la population, et donc sur le potentiel évolutif du virus. Cependant, il est difficilement accessible car la plupart des mutations qui surviennent spontanément au cours de la multiplication du virus ont un effet délétère pour celui-ci et disparaissent rapidement de la population virale par contre-sélection. C'est donc le taux de substitution qui est directement observable. Cette difficulté à calculer le taux de mutation est en partie due au manque de marqueurs neutres dans les génomes viraux. Au cours de nos expériences, on suppose donc que le taux de mutation intrinsèque est similaire entre les populations de PVY. La probabilité d'apparition d'une mutation dépend également du nombre de réplications effectuées par le virus, un virus effectuant un grand nombre de réplications générant un plus grand nombre de mutations. Lors de nos expériences, nous faisons l'hypothèse que ce nombre de réplications est proportionnel à l'accumulation virale de la population dans

la plante, que l'on peut aisément calculer par un ELISA quantitatif. L'accumulation virale dans la plante est donc utilisée comme un proxy de l'intensité de la mutation.

Concernant la sélection et la dérive génétique, Rousseau et al. (2017) ont développé un modèle permettant de mesurer conjointement l'effet de la sélection et de la dérive génétique sur les populations virales, à différents intervalles de temps au cours de l'infection. Pour ce faire, ils ont inoculé 15 lignées HD de piment et 8 plantes par lignée HD avec une population composée de cinq variants de PVY (N, G, K, GK et KN) en proportion équimolaire. Ils ont ensuite prélevé la feuille inoculée 6 jours après inoculation, puis effectué des prélèvements sur des feuilles apicales 10, 14, 20, 27 et 34 jours après inoculation (Figure 7a). La méthode de prélèvement étant destructive, 8 nouvelles plantes par lignée HD étaient échantillonnées à chaque date. Le séquençage des populations virales au sein de chaque échantillon a ensuite été effectué pour comparer l'évolution des fréquences des variants viraux entre les plantes d'une même lignée (séquençage Illumina MiSeq générant en moyenne 3000 séquences de PVY par plante). La sélection étant une force déterministe, elle augmentera systématiquement la fréquence des variants viraux les plus adaptés au sein de la population au détriment des variants les moins adaptés. Chez une lignée HD induisant une forte pression de sélection, les compositions finales des populations virales devraient être similaires entre les plantes appartenant à cette lignée, les variants les plus adaptés étant toujours sélectionnés et les variants les moins adaptés étant toujours éliminés. C'est ce qu'on observe chez la lignée A, pour laquelle la composition de la population virale est très similaire entre les 8 plantes au cours du temps, toutes les plantes ayant systématiquement contre-sélectionné les variants N et G (bleu foncé et bleu clair) et majoritairement sélectionné le variant K (vert) (Figure 7b). A l'opposé, la dérive génétique est une force stochastique qui induira une variation aléatoire de la fréquence des variants viraux au cours du temps, indépendamment de leur valeur sélective. C'est ce qu'on observe chez la lignée B, pour laquelle la variance inter-plantes est très forte, les 8 plantes de cette lignée ne fixant jamais les mêmes variants viraux à chaque date. Ces deux lignées sont des exemples extrêmes, des lignées induisant des effets de sélection et de dérive génétique plus modérés existant également. Sur la base de ces données de fréquences virales, le modèle de Rousseau et al. (2017) estime deux types de paramètre. Tout d'abord le taux d'accroissement (r_i) de chacun des variants renseigne sur l'intensité de la sélection. Sur la base de ces taux d'accroissement, un coefficient de sélection (s) peut être obtenu soit en calculant la différence entre le variant avec le r_i le plus élevé et le variant avec le r_i le plus faible, soit en calculant l'écart-type entre les 5 valeurs de r_i correspondant aux 5 variants de PVY. Le second paramètre

calculé est la taille efficace de population (N_e) dont la valeur renseigne sur l'intensité de la dérive génétique. Les données acquises sur les 15 lignées HD montrent que la sélection et la dérive génétique imposées par les lignées aux populations virales sont des caractères fortement variables entre les lignées. De plus, ces deux caractères se sont révélés être très héritables (h^2 = 0.63 pour N_e et h^2 = 0.94 pour s). Le paramètre h^2 est l'héritabilité au sens large et correspond à la part de variance phénotypique d'origine génétique.

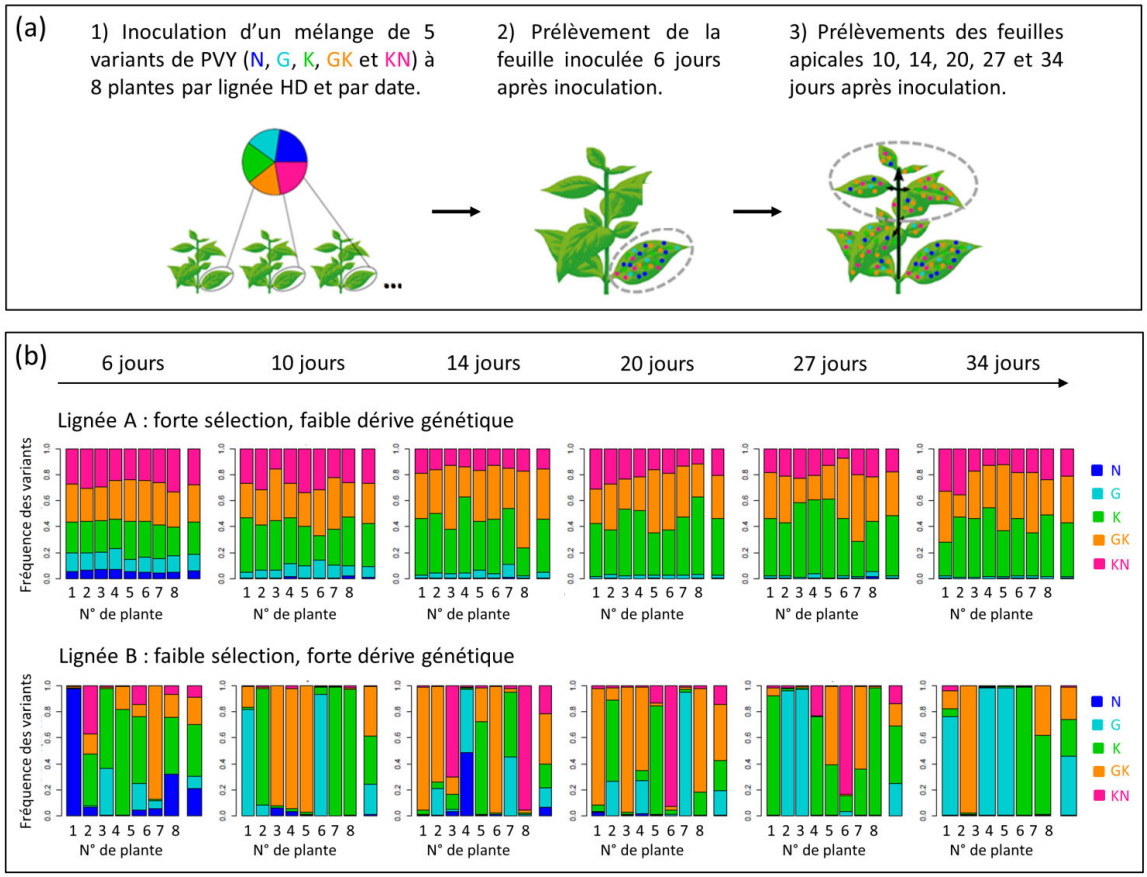


Figure 7 : Mesure de l'intensité de la sélection et de la dérive génétique dans des lignées HD. (a) Protocole expérimental employé pour suivre l'évolution des fréquences de 5 variants de PVY chez 15 lignées HD et à 6 dates différentes après inoculation. **(b)** Fréquences des 5 variants de PVY obtenues suite au séquençage des échantillons prélevés à 6 dates après inoculation et chez deux lignées HD contrastées en termes de sélection et de dérive génétique. Les 8 barres de chaque histogramme correspondent aux 8 plantes de la lignée étudiée, et la 9ème barre à la moyenne des 8 plantes.

Puisque des outils nous permettant de mesurer précisément les forces évolutives induites par la plante sur les populations virales sont à notre disposition et que ces caractères semblent variables et héritables, il est dorénavant possible de rechercher les déterminants génétiques de la plante à l'origine de ces forces évolutives. Il est également possible de sélectionner des lignées sur la base de ces critères pour étudier s'il est possible de contraindre l'évolution des virus en jouant sur ces effets de sélection et de dérive génétique

C. Objectifs de la thèse

Cette thèse comprend deux objectifs majeurs. Le premier est de caractériser les facteurs de résistance partielle disponibles dans les ressources génétiques du piment pour leur aptitude à induire des effets de dérive génétique sur les populations virales. Le second est d'étudier l'impact des forces évolutives induites par la plante sur l'adaptation des virus aux résistances variétales et de comprendre comment ces forces évolutives peuvent être employées pour contraindre au mieux l'évolution des virus. Pour répondre à ces objectifs, quatre approches expérimentales ont été mises en place. Ces approches font l'objet d'articles acceptés ou en préparation qui structurent ce manuscrit en quatre chapitres de résultats et un chapitre de discussion :

Le premier chapitre est consacré à l'identification de QTL de la plante contrôlant la taille efficace des populations virales (N_e), et donc l'intensité de la dérive génétique, lors de l'étape d'inoculation. Cette étude a été réalisée par une approche de détection de QTL dans une population de piment haploïde doublée. Afin de comparer la généricité des résultats chez plusieurs espèces virales, l'analyse a été réalisée à la fois pour le PVY et le CMV.

Le second chapitre porte sur la détection de régions génomiques contrôlant N_e à l'inoculation ainsi que l'accumulation virale dans la plante entière. Pour ce faire, une approche de génétique d'association a été réalisée sur une core-collection de piments représentative de la diversité génétique de l'espèce *Capsicum annuum*.

Le troisième chapitre est dédié à l'étude de l'effet de la dérive génétique, de la sélection et de l'accumulation virale induits par la plante sur la durabilité d'un gène majeur. En utilisant des données obtenues sur 89 lignées haploïdes doublées de piment, cette étude a cherché à identifier quelles forces évolutives expliquent au mieux la fréquence de contournement du gène majeur *pvr2*^{3}.

Le quatrième chapitre présente une étude sur l'adaptation des populations de PVY à 6 lignées haploïdes doublées de piment. Ces lignées sont toutes porteuses du gène majeur *pvr2*^{3} mais induisent des effets de dérive génétique et de sélection contrastés sur les populations virales et permettent différents niveaux d'accumulation virale. De l'évolution expérimentale par passages répétés de populations de PVY a été réalisée durant sept mois sur ces lignées afin d'évaluer les possibilités d'adaptation des virus par mutation ainsi que les éventuelles contraintes d'adaptation induites par les forces évolutives imposées par la plante. En retour, les

trajectoires évolutives observées lors de cette expérience ont été analysées par rapport à la dérive génétique, la sélection et l'accumulation virale chez les différentes lignées afin d'identifier lesquelles de ces variables jouaient les rôles les plus déterminants.

Le cinquième chapitre discute ces résultats et les perspectives de ce travail. Enfin, l'article « Estimating virus effective population size and selection without neutral markers » de Rousseau et al. (2017) est disponible en annexe de ce manuscrit. Cet article présente le modèle utilisé dans cette thèse pour caractériser les lignées haploïdes doublées de piment en fonction des effets de dérive génétique et de sélection qu'elles imposent aux populations virales (chapitres 3 et 4). Cet article vient d'être accepté pour publication dans PLOS Pathogens mais n'est pas encore disponible en ligne et a donc été ajouté en annexe.

Chapitre 1

Quantitative trait loci in pepper control the effective population size of two RNA viruses at inoculation

Résumé de l'article :

Objectifs : Le premier objectif était d'identifier les facteurs génétiques contrôlant la taille efficace de population (N_e), et donc l'intensité de la dérive génétique induite par la plante, lors de l'étape d'inoculation des virus. Ces facteurs ont été recherchés à la fois pour le PVY et le CMV. Le second objectif était d'améliorer notre compréhension des mécanismes employés par la plante pour contrôler N_e. Enfin, le troisième objectif était de comparer les facteurs détectés avec les facteurs de résistance précédemment identifiés chez le piment.

Stratégie : Une population haploïde doublée de piment issue d'un croisement biparental a été inoculée sur les cotylédons avec un clone infectieux de PVY exprimant la protéine fluorescente GFP et sur les feuilles avec une souche de CMV nécrotique. Le nombre de foyers fluorescents et le nombre de lésions locales nécrotiques ont été quantifiés et utilisés comme des proxys de N_e. Une détection de QTL a été réalisée et les résultats ont été comparés à ceux précédemment obtenus chez ces lignées.

Résultats : Les N_e lors de l'étape d'inoculation sont des traits fortement héritables au sein des lignées de piment, à la fois pour le PVY (h^2=0,93) et le CMV (h^2=0,98). Deux QTL communs aux deux virus et présentant un fort effet d'épistasie ont été identifiés sur les chromosomes 7 et 12. Des QTL spécifiques de chacun des virus ont également été identifiés sur le chromosome 6. La régulation des N_e à l'inoculation requiert donc la contribution de mécanismes à la fois généraux et spécifiques de chaque virus. Enfin, le QTL spécifique du PVY co-localise avec un QTL qui réduit l'accumulation virale ainsi que la capacité de contournement du gène majeur *pvr2*[3].

JOURNAL OF GENERAL VIROLOGY

RESEARCH ARTICLE

Tamisier *et al.*, *Journal of General Virology* 2017;98:1923–1931

DOI 10.1099/jgv.0.000835

MICROBIOLOGY SOCIETY

Quantitative trait loci in pepper control the effective population size of two RNA viruses at inoculation

Lucie Tamisier,[1,2,*]† Elsa Rousseau,[2,3,4]† Sebastien Barraillé,[2] Ghislaine Nemouchi,[1] Marion Szadkowski,[1] Ludovic Mailleret,[3,4] Frederic Grognard,[3] Frederic Fabre,[5] Benoit Moury[2,*] and Alain Palloix[1]

Abstract

Infection of plants by viruses is a complex process involving several steps: inoculation into plant cells, replication in inoculated cells and plant colonization. The success of the different steps depends, in part, on the viral effective population size (N_e), defined as the number of individuals passing their genes to the next generation. During infection, the virus population will undergo bottlenecks, leading to drastic reductions in N_e and, potentially, to the loss of the fittest variants. Therefore, it is crucial to better understand how plants affect N_e. We aimed to (i) identify the plant genetic factors controlling N_e during inoculation, (ii) understand the mechanisms used by the plant to control N_e and (iii) compare these genetic factors with the genes controlling plant resistance to viruses. N_e was measured in a doubled-haploid population of *Capsicum annuum*. Plants were inoculated with either a *Potato virus Y* (PVY) construct expressing the green fluorescent protein or a necrotic variant of *Cucumber mosaic virus* (CMV). N_e was assessed by counting the number of primary infection foci on cotyledons for PVY or the number of necrotic local lesions on leaves for CMV. The number of foci and lesions was correlated (r=0.57) and showed a high heritability (h^2=0.93 for PVY and h^2=0.98 for CMV). The N_e of the two viruses was controlled by both common quantitative trait loci (QTLs) and virus-specific QTLs, indicating the contribution of general and specific mechanisms. The PVY-specific QTL colocalizes with a QTL that reduces PVY accumulation and the capacity to break down a major-effect resistance gene.

INTRODUCTION

During the plant infection process, RNA viruses are generally able to evolve quickly and adapt to their host thanks to their high mutation rate and short generation time [1]. As a result, breakdown of plant resistance by the emergence of new virus variants may occur and cause important losses for agricultural production and quality [2, 3]. A better understanding of the evolutionary processes that shape viral populations and the extent to which we can control them is therefore required for the sustainable management of crop disease [4].

In plants, two well-known evolutionary forces act on the frequencies of the different variants comprising virus populations: natural selection and genetic drift. Natural selection is a deterministic force that increases the frequency of the fittest variants at each generation. In contrast, genetic drift is a stochastic force that randomly changes the frequencies of the virus variants from generation to generation [5]. The two forces act jointly on viral populations and can have opposite effects on their adaptation. Indeed, if genetic drift is strong, deleterious mutations may be randomly fixed or advantageous ones may be lost. The strength of genetic drift depends on a key parameter of virus evolution: the effective population size (N_e). N_e can be defined as the number of individuals that pass their genes to the next generation [6], and the strength of genetic drift is inversely proportional to this number. N_e is generally much smaller than the census population size N [7]. Through the infection process, the viral population will endure several bottlenecks that will strongly reduce N_e and consequently increase genetic drift [8, 9]. These bottlenecks can occur during all the infection steps, like vector transmission [10, 11], virus inoculation into plant cells [12], replication in infected cells [13] and cell-to-cell or long-distance movements

Received 31 January 2017; Accepted 12 May 2017

Author affiliations: [1]INRA, UR1052 GAFL, Unité de Génétique et Amélioration des Fruits et Légumes, Domaine St Maurice - 67 Allée des Chênes, CS 60094, F-84143 Montfavet Cedex, France; [2]INRA, UR407 PV, Unité de Pathologie Végétale, Domaine St Maurice - 67 Allée des Chênes, CS 60094, F-84143 Montfavet Cedex, France; [3]INRIA, Biocore Team, F-06902 Sophia Antipolis, France; [4]INRA, Université Nice Sophia Antipolis, CNRS, UMR 1355-7254 Institut Sophia Agrobiotech, Sophia Antipolis, France; [5]INRA, UMR 1065 Santé et Agroécologie du Vignoble, BP 81, 33883 Villenave d'Ornon cedex, France.

***Correspondence:** Lucie Tamisier, lucie.tamisier@inra.fr; Benoit Moury, benoit.moury@inra.fr

Keywords: effective population size; bottlenecks; quantitative trait loci; *Capsicum annuum*; virus infection; *Potato virus Y*; *Cucumber mosaic virus*.

Abbreviations: CMV, *Cucumber mosaic virus*; DH, doubled-haploid; h^2, broad-sense heritability; LOD, logarithm of odds; N_e, effective population size; PVY, *Potato virus Y*; QTL, quantitative trait locus; ToMV, *Tomato mosaic virus*; VA, virus accumulation.

†These authors contributed equally to this work.

One supplementary table and three supplementary figures are available with the online Supplementary Material.

in the infected plant [14]. Although estimation of bottleneck size and its effects on the genetic diversity of the viral population is well documented [15, 16], the plant genetic determinants controlling bottleneck size have been little studied to date. Previous studies have found quantitative trait loci (QTLs) involved in the reduction of the viral accumulation, which corresponds to the census population size N [17–19], but no study has searched for genomic regions causing direct changes in the N_e of viruses. However, since N_e could be a more relevant parameter than N for virus evolution, understanding how plant genetic factors may affect it could contribute to the development of cultivars slowing down pathogen evolution and increasing resistance durability [20].

In this study, we focused on N_e during the inoculation of pepper (*Capsicum annuum*) plants with two RNA viruses, *Potato virus Y* (PVY; genus *Potyvirus*, family *Potyviridae*) and *Cucumber mosaic virus* (CMV; genus *Cucumovirus*, family *Bromoviridae*). We aimed to (i) identify the plant QTLs that control N_e at the inoculation step, (ii) understand the mechanisms used by the plant to control N_e and (iii) compare these genetic factors with others factors controlling the virus accumulation and plant resistance durability.

RESULTS

Very few primary infection foci are initiated by two PVY variants simultaneously

To measure N_e, we inoculated *C. annuum* plants with either a PVY construct expressing the green fluorescent protein (GFP), the PVY-GFP, or a necrotic variant of CMV, the CMV-N strain of Fulton. We then quantified the number of primary infection foci under a specific light wavelength for PVY-GFP, or the number of necrotic local lesions for CMV-N observed on the inoculated organs. We hypothesized that one primary infection focus or one local lesion was caused by only one viral particle or one 'infectious unit' composed of several particles containing different genome components in the case of a multipartite virus like CMV. Therefore, counting the number of foci or lesions on the inoculated organs would be a direct assessment of N_e at inoculation.

To validate the N_e estimation method, we conducted a control experiment using co-infections by two PVY constructs expressing different fluorescences (GFP and mCherry), similar to Zwart *et al.* [21]. Among a population of 152 *C. annuum* doubled-haploid (DH) lines issued from the F_1 hybrid between the PVY-resistant parent Perennial and the susceptible parent Yolo Wonder, we selected 13 DH lines that showed contrasting numbers of foci when inoculated by PVY-GFP alone (Fig. S1, available with the online Supplementary Material). A 1 : 1 mixture of the PVY variants carrying either the GFP or the mCherry fluorescent markers was inoculated on the first true leaf of the plants. We kept only the leaves showing at least 1 infection focus of each colour, and ended up with a total of 49 leaves and 1 to 8 leaves per DH line (Table S1).

For 75.5 % of the inoculated leaves, no infection foci with dual fluorescences were observed. Overall, the mean frequency of foci showing both red and green fluorescences was 0.64 % among the DH lines, with a maximum frequency per plant of 5.3 %. The mean number of viral particles initiating a focus was assessed using the model proposed by Sacristán *et al.* [7] [equations (1) and (2)]. On average, 1.01 to 1.02 viral particles initiated an infection focus, with a maximum number per DH line of 1.06 (Table S1), which validated our hypothesis.

The numbers of primary infection foci and local lesions are highly heritable traits

The DH population comprising 152 lines of *C. annuum* was inoculated with either the PVY-GFP variant on the cotyledons or the CMV-N strain of Fulton on one leaf, which induced fluorescent infection foci or necrotic local lesions, respectively (Figs 1 and S2). The number of primary infection foci on the 2 cotyledons or local lesions on 1 leaf were then quantified on 10 plants per DH line. For PVY, the mean number of primary infection foci ranged from 1.15 to 44.05 among the different DH lines, with an overall mean number of 15.56±9.47 (mean ± standard deviation) (Fig. 2a). For CMV, the overall mean number of local lesions varied from 0.20 to 95.00, with a mean number of 24.59±21.75 (Fig. 2b). The two variables were significantly correlated among the DH lines (Pearson r=0.57, P=8.97×10^{-15}). They also both showed a high heritability, with h^2=0.93 for the foci induced by PVY and h^2= 0.98 for the lesions induced by CMV. Furthermore, the numbers of foci and lesions of the F_1 hybrid (Yolo Wonder×Perennial) were intermediate between those of the two parental lines. Since they were closer to Perennial (the parent showing the least number of foci or lesions) than to Yolo Wonder (the susceptible parent) in both cases (Fig. 2), we concluded that the numbers of foci and lesions have either a dominant or codominant inheritance.

Detection of QTLs controlling the numbers of primary infection foci and local lesions for PVY and CMV

Three QTLs were detected for each virus (Table 1 and Fig. 3). They were named PVY-6, PVY-7, PVY-12 and CMV-6, CMV-7, CMV-12 according to the virus used for the

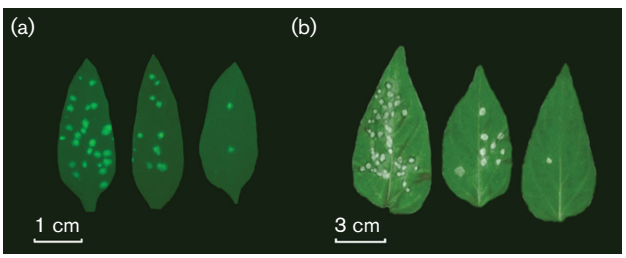


Fig. 1. Primary infection foci or necrotic local lesions on pepper cotyledons or leaves inoculated with PVY-GFP (a) or CMV-N (b).

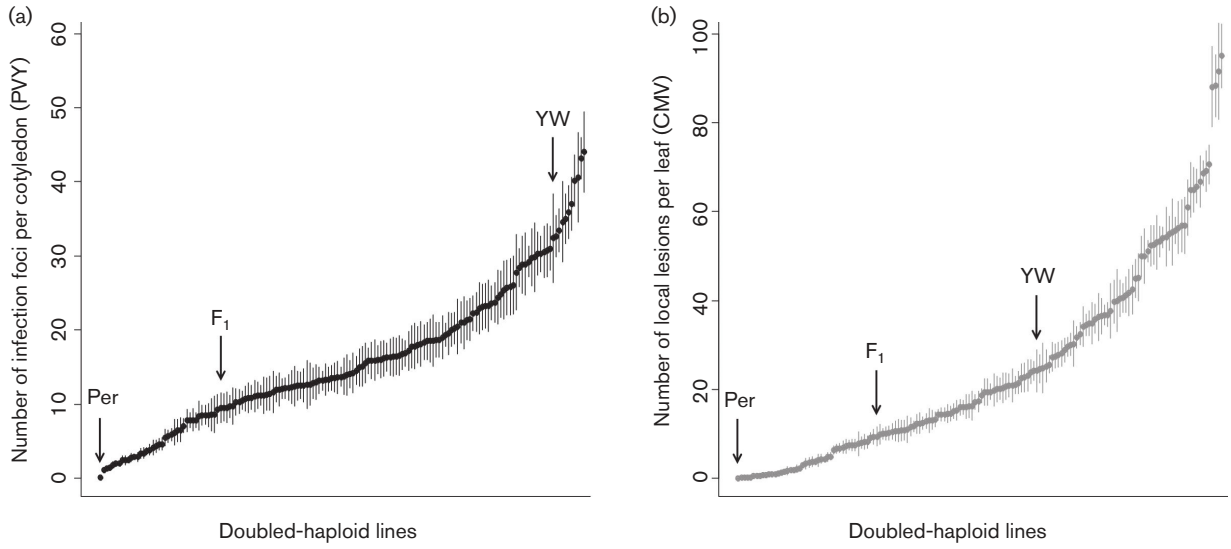


Fig. 2. The number of primary infection foci induced by PVY-GFP (a) and the number of local lesions induced by CMV-N (b) among the DH lines. Error bars represent the standard error of the mean. The positions of the parental lines, Perennial (Per) and Yolo Wonder (YW), and the F_1 hybrid are indicated.

inoculation and the chromosome number. The QTLs PVY-6, PVY-7 and PVY-12 explained respectively 6.28, 34.73 and 26.22 % of the variation of the primary infection foci numbers for PVY. Similarly, QTLs CMV-6, CMV-7 and CMV-12 explained respectively 11.18, 31.53 and 21.67 % of the variation of the local lesion numbers for CMV. For both viruses, analysis revealed a significant epistatic interaction between the QTLs on chromosomes 7 and 12. This epistatic interaction explained 11.46 and 9.38 % of the primary infection foci numbers and the local lesion numbers for PVY and CMV, respectively. More precisely, for the DH lines with the Perennial allele at both QTLs, the number of foci/lesions was on average lower than expected if there was no epistasis (i.e. it is a case of synergistic epistasis; Fig. 4). Regarding all the QTLs independently, the Perennial allele always decreased the trait value, except in the case of CMV-6. Finally, the model combining the additive and epistatic effects of the three QTLs explained 57.82 and 50.88 % of the trait variation for PVY and CMV, respectively.

The number of infection foci induced by PVY correlates with the number of local lesions induced by ToMV

Eight DH lines were selected according to the QTLs they carried. We chose one DH line without QTL, one DH line with one QTL (PVY-6), two DH lines with two QTLs (PVY/CMV-7 and PVY-6), two others DH lines with two QTLs (PVY/CMV-7 and PVY/CMV-12) and two DH lines with three QTLs (PVY/CMV-7, PVY/CMV-12 and CMV-6). The eight DH lines were mechanically inoculated with *Tomato mosaic virus* (ToMV; genus *Tobamovirus*, family *Virgaviridae*). At 5 days post-inoculation (p.i.), the number of local lesions observed on leaves was quantified. The mean number of lesions ranged from 3.13 to 26.15 among the DH lines, with a mean number of 10.66±8.34. The variable was significantly correlated with the number of foci caused by PVY (Pearson's r=0.86, P=5.60×10^{-3}, Fig. 5a), but was not significantly correlated with the number of lesions caused by CMV (Pearson's r=0.56, P=0.15, Fig. 5b).

DISCUSSION

One key parameter of virus evolution is N_e, which corresponds to the number of virus individuals that pass their genes to the next generation [6]. As a viral population infects a plant, it will experience several bottlenecks that will reduce N_e. A low value of N_e will generally have a negative impact on the fitness of a viral population, even if some exceptions exist [15]. A low N_e will increase genetic drift, reduce genetic diversity and possibly lead to fitness decline by losing the most adapted variants. The repeated bottleneck events can also be conducive to Muller's ratchet processes. In this situation, the population will accumulate slightly deleterious mutations at each generation. The individuals without deleterious mutations are definitively lost by genetic drift and the final consequence can be the extinction of the population if the deleterious mutations continue to accumulate [22]. Therefore, it is of primary importance to better understand the evolutionary constraints imposed by the plant that reduce the N_e of the pathogen.

Link between the number of primary infection foci or local lesions and the effective population size

The inoculation step consists of a particularly narrow bottleneck for viruses: only a few individuals from the inoculum source succeed in initiating infection of new plants [10, 15]. Previous studies have reported very low values of N_e at this step. For example, depending on the virion dose, N_e of

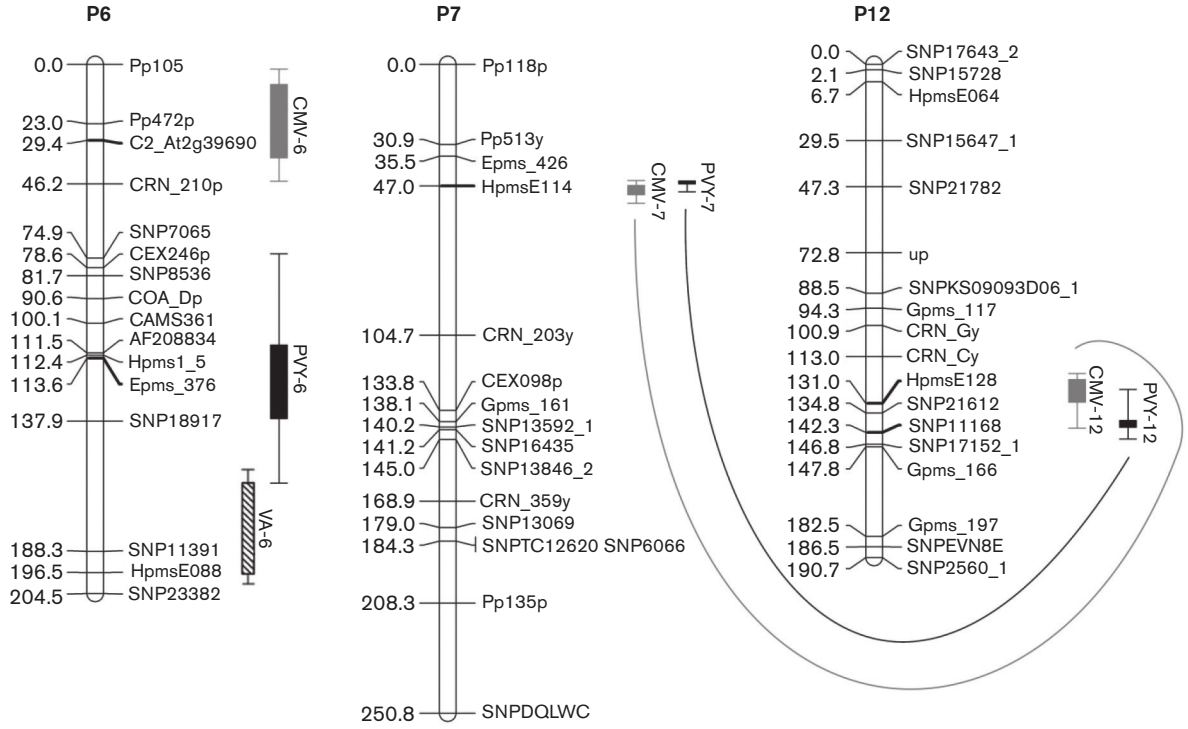


Fig. 3. QTL map for the effective population size at inoculation for PVY (black) and CMV (grey). A previously detected QTL (VA-6) controlling PVY accumulation is also mapped (hatched). For each QTL, confidence intervals obtained using the 1-LOD drop and 2-LOD drop methods are indicated. The lines represent epistatic effects between loci.

Tobacco etch virus (TEV) was estimated to have between 1 and 50 viral particles following mechanical inoculation in *Nicotiana tabacum* [21]. Considering aphid transmission, the number of virus particles transmitted was estimated to be between 0.53 and 3.24 for PVY [10] and between 0.5 and 35.8 for CMV [11]. In our study, the number of infection foci (for PVY) or local lesions (for CMV) are the minimum values for N_e at the inoculation step, since they are initiated by at least one virus particle. They would correspond to exact N_e values if and only if each infection focus/local lesion was initiated by exactly one virus particle. As proposed by Zwart *et al.* [21], we tested this hypothesis by co-inoculating *C. annuum* leaves with PVY expressing two different reporter proteins showing either green or red fluorescence, respectively. We found that the frequency of infection foci showing both green and a red fluorescence was very low, with a mean frequency of 0.64 % per DH line, and we estimated that the mean number of viral particles initiating a focus ranged from 1.01 to 1.02. Thus, we could conclude that the huge majority of foci were initiated by a single viral particle and that the number of primary infection foci was a precise estimate of N_e at the inoculation step.

The same approach could not be undertaken for CMV or ToMV. However, the quantification of local lesions has been widely used to investigate the relationship between the virus inoculum dose and the number of lesions [23, 24]. Combining models and experimental data, it has been shown that most of the local lesions are initiated by a single virus particle when considering low to moderate inoculum concentration [25, 26]. More recently, Sacristán *et al.* [12] provided an estimate of the number of founders that start an infection by contact transmission. They measured the number of necrotic local lesions induced by two *Tobacco mosaic virus* (TMV; genus *Tobamovirus*) genotypes on tobacco plants and found that, on average, one viral particle initiated each local lesion. These results suggest that the number of local lesions is also an accurate estimate of N_e for CMV and ToMV. Therefore, in the next sections of the paper, we will treat N_e at inoculation as being synonymous with the number of primary infection foci or local lesions.

Finally, since N_e is dose-dependent [21], the inoculum dose used in our study is also an important parameter in the estimation of N_e. Our goal was to use an inoculum dose that maximizes the differences in the number of foci/lesions between the DH lines in order to perform an efficient QTL detection. This objective was achieved, since the number of foci per cotyledon for PVY-GFP ranged from 1.15 to 44.05 for the DH lines and was 0.10 for the Perennial parent (Fig. 2a). The result was similar for CMV-N, with between 0.20 and 95.00 lesions/leaf (Fig. 2b). For the DH lines showing the highest numbers of foci, higher inoculum doses would have produced too many foci for them to be

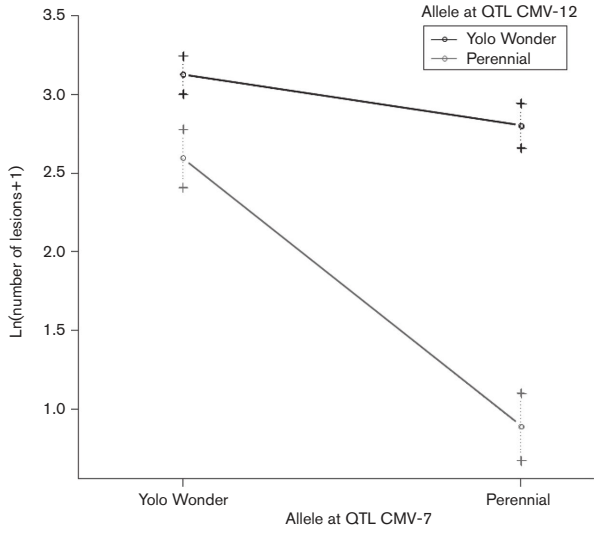


Fig. 4. Epistasis effect between QTLs CMV-7 and CMV-12. The average number of lesions induced by CMV after ln +1 transformation is plotted as a function of the parental alleles (Yolo Wonder or Perennial) at QTL CMV-7 and QTL CMV-12. Error bars represent the standard error of the mean. Similar results (i.e. synergistic epistasis) were obtained for PVY.

visualized without ambiguity, i.e. we would be at a 'technical saturation' point of the inoculum concentration density of the infection foci curve. Clearly, we do not encounter this type of saturation in our work. First, the curves representing the number of foci/lesions among the DH lines increase continuously and do not reach a plateau, both for PVY and CMV, suggesting that these numbers are not the highest possible for the vast majority of the lines (Fig. 2a, b). Since there are only small differences between the cotyledon surface areas of the different DH lines, the curve representing the foci density also increases continuously among the lines and does not reach a plateau (Fig. S3). This demonstrates that the number of foci per cotyledon does not reach the maximum value which is technically achievable given the cotyledon size. The reason could be either that the inoculum dose is not saturating or that the maximum number of virus entry sites is very low for some of the DH lines.

Common and virus-specific QTLs control the effective population size of PVY and CMV at inoculation

For both PVY and CMV, we identified three QTLs controlling N_e at inoculation and localized on chromosomes 6, 7 and 12 (Table 1 and Fig. 3). The QTLs on chromosome 7, PVY-7 and CMV-7, were detected at the same location in the genome (the strongest association was with marker HpmsE114), and the QTLs on chromosome 12, PVY-12 and CMV-12, were identified at very close positions (139.1 and 125.7 cM). On each chromosome, the confidence intervals of the two QTLs overlapped to a large extent. Moreover, the phenotypic variation explained by the QTLs was similar, with PVY-7 and CMV-7 explaining 34.73 and 31.53 % of the trait variation, and PVY-12 and CMV-12 explaining 26.22 and 21.67 % of the trait variation. We also found that for both viruses there was an epistasis between the QTLs on chromosomes 7 and 12, and that the epistasis had a similar explanatory power for both viruses (11.46 % for PVY and 9.38 % for CMV). Therefore, the same QTLs on chromosomes 7 and 12 control N_e for PVY and CMV, and the same genetic factor(s) may be responsible for this dual effect. Note that in a previous study, Caranta *et al.* [27] detected three pepper QTLs controlling the number of necrotic local lesions caused by the CMV-N strain of Fulton. At the time, the pepper linkage map coverage was incomplete, making the comparison difficult. However, CMV-12 seems to correspond to one of the three QTLs, as they are located in the same genomic region.

Table 1. QTLs detected for the effective population size N_e of PVY and CMV at inoculation

QTL*	Chr†	Position (cM)	Closest marker	LOD score	2-LOD support interval	Variation explained (%)		Estimated effect of Perennial QTL allele‡	h^2
						QTL	Model		
PVY-7	7	45.6	HpmsE114	20.71	45.1–49.3	34.73	57.82	−0.803	0.93
PVY-12	12	139.1	SNP11168	15.30	125.7–144.8	26.22		−0.652	
PVY-6	6	123.8	Epms_376	4.93	73.2–161.9	6.28		−0.320	
PVY-7 × PVY-12	7/12	–	HpmsE114/ SNP11168	8.48	–	11.46		−0.906	
CMV-7	7	48.4	HpmsE114	16.36	44.9–53.7	31.53	50.88	−1.200	0.98
CMV-12	12	125.7	HpmsE128	12.06	119.5–140.7	21.67		−1.091	
CMV-6	6	31.5	C2_At2g39690	6.77	1.8–45.2	11.18		0.712	
CMV-7 × CMV-12	7/12	–	HpmsE114/ HpmsE128	5.77	–	9.38		−1.397	

*QTL name includes the virus used for plant inoculation followed by the number of the chromosome carrying the QTL.

†Chromosome number.

‡Estimated effect of the Perennial allele on the value of the trait studied. For the pairwise epistasis effects, it corresponds to the difference in the effects at the first QTL considering the two alleles at the other QTL separately.

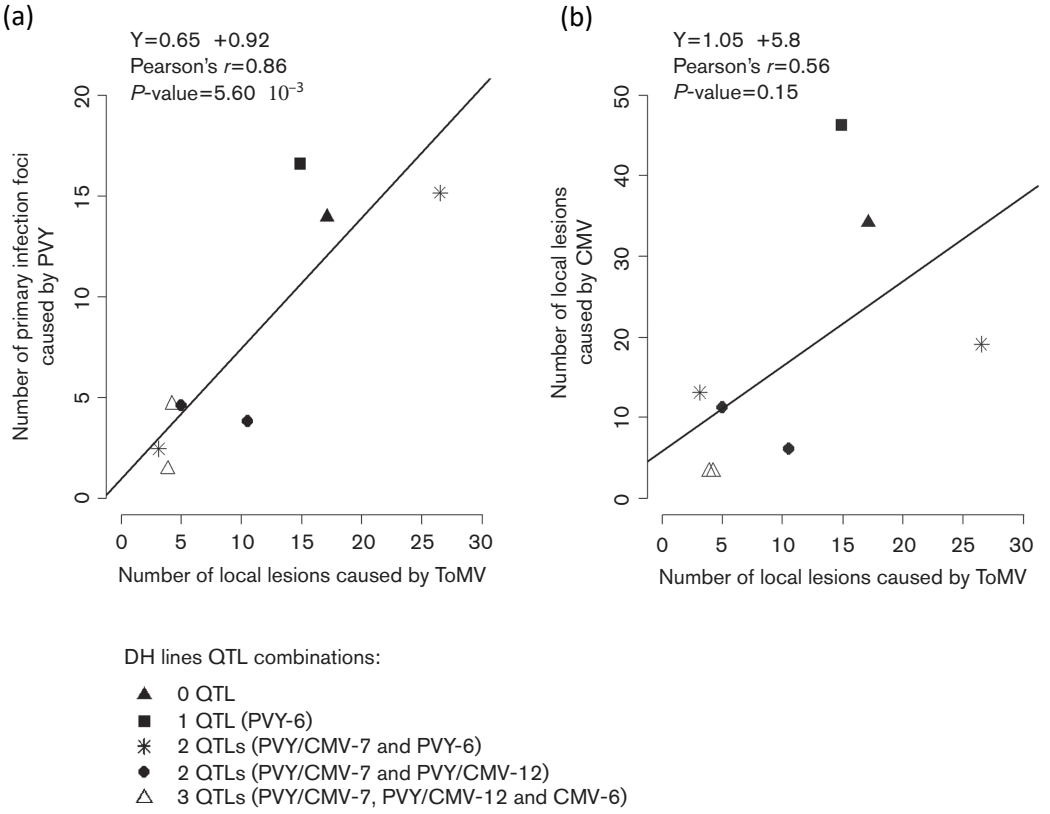


Fig. 5. Correlation between the number of lesions induced after inoculation by ToMV and PVY (a) or ToMV and CMV (b) on eight DH lines carrying different QTL combinations. The linear regression equations, Spearman's correlation coefficients and *P*-values of the correlations are indicated.

In contrast to the QTLs PVY/CMV-7 and PVY/CMV-12, the two QTLs detected on chromosome 6, PVY-6 and CMV-6, differed according to the virus. PVY-6 was localized at 123.8 cM and associated with the marker Epms_376, whereas CMV-6 was positioned at 31.5 cM and associated with the marker C2_At2g39690. Further, the QTL effects showed opposite directions, since the Perennial allele decreased the value of the trait for PVY-6 and increased its value for CMV-6. Even if the two viruses are quite different, our study highlights that two common QTLs control the effective population size of both viruses, indicating that general mechanisms underlie this trait. However, we also found one specific QTL for each virus, demonstrating that virus-specific mechanisms also act.

Hypothesis on the mechanisms of action of the QTLs

We have identified two QTLs that are shared between PVY and CMV, and two virus-specific QTLs that control N_e. Leaves were chosen to inoculate CMV because the number of CMV lesions in cotyledons is usually very low and the necrotic lesions are quite large at the moment of their detection. In contrast, cotyledons were chosen to inoculate PVY-GFP because the number of fluorescent infection foci on leaves is usually very high. Consequently, since the inoculations have not been performed on the same plant organs for the two viruses, we do not know if the QTLs detected for a given virus will still act in another plant organ. To answer this question, we inoculated 10 pepper lines on either one leaf or one cotyledon, both for PVY and CMV. The number of foci induced after PVY inoculation was significantly correlated between the leaves and the cotyledons (Pearson's r=0.93, P=9.72×10^{-5}, data not shown), as was the number of lesions caused by CMV (Pearson's r=0.75, P=0.01, data not shown). Therefore, as the N_e at inoculation is significantly correlated between leaves and cotyledons for both PVY and CMV, we can reasonably assume that the mechanisms controlling N_e at inoculation are not organ-specific.

To investigate the spectrum of action of the QTLs more deeply, we used a third virus, a strain of ToMV, which also induces local lesions on the leaves. The plant L^1 allele required for the expression of hypersensitive necrotic local lesions upon ToMV inoculation was only present in half of the DH lines. Consequently, this small number of DH lines was insufficient to conduct QTL mapping with ToMV. Therefore, we inoculated ToMV to a representative subset of eight DH lines carrying different QTL combinations. We then counted the local lesions on the leaves and compared

the results to those of PVY and CMV. We found that the values of N_e at inoculation were significantly correlated between ToMV and PVY (Fig. 5a). In contrast, no significant correlation was observed between ToMV and CMV (Fig. 5b). These findings support our previous results that found that some of the QTLs have a broad spectrum of action, including ToMV and PVY in this case. It also confirms the inference that some QTLs have a narrower spectrum of action, potentially explaining the differences between ToMV and CMV. We can hypothesize that the broad-spectrum QTLs are generalists because they act during the primary steps of the infection, before the establishment of narrow molecular interactions between the virus and intra-cellular plant components. These broad-spectrum QTLs could affect penetration of the virus in the cotyledon or leaf cells by controlling traits such as tissue thickness or sensitivity to wounding. The narrow-spectrum QTLs could act slightly later during the infection, playing on the probability for a given virus to successfully initiate the infection focus once it has penetrated into the first cells. These two mechanisms would have contrasting effects according to the virus inoculation mode, i.e. manual or with aphids. QTLs acting early on virus penetration may have little effect on aphid inoculation, since the aphid stylet will penetrate more deeply into the leaf tissues and will bypass the action of these QTLs. In contrast, QTLs acting later, once the virus is within the cell, are expected to have a similar effect in mechanical and aphid inoculations.

These findings are also interesting from an agronomic viewpoint because ToMV is a contact-transmitted virus. In this situation, plants that reduce N_e as soon as the inoculation step occurs could be particularly helpful for reducing or stopping plant-to-plant transmission and epidemics in the field. In contrast, PVY and CMV are naturally transmitted by aphids. Thus, further experiments are required to determine whether our observations of N_e variation between DH lines after mechanical inoculation can be extended to aphid transmission.

Relationship between QTLs of effective population size and plant resistance

The results from previous studies revealing the phenotypic and genetic factors involved in pepper resistance to PVY and CMV can be compared to our results. With the same DH population, Quenouille *et al.* [19] identified the pepper QTLs controlling the durability of the major PVY resistance gene *pvr2*^{3} and PVY accumulation. In theory, we could expect that a lower N_e at inoculation would delay plant infection and therefore decrease virus accumulation. We could also expect that a lower N_e at inoculation would help to lose well-adapted virus variants by genetic drift, therefore increasing resistance durability. However, the genetic evidence for links between N_e at inoculation, virus accumulation and resistance durability is weak in our case. The only co-localization with the QTLs controlling the variation of N_e at inoculation would concern a QTL on chromosome 6, named VA-6, which affects PVY accumulation (Fig. 3). QTL VA-6 showed an epistatic interaction with RB-3, a QTL on chromosome 3 controlling the virus capacity to break down *pvr2*^{3} resistance. The Perennial allele at QTL PVY-6 decreases N_e, while at QTL VA-6 it increases resistance durability and decreases within-host virus accumulation, which is consistent with our expectations. However, even if the confidence interval of VA-6 includes a part of the confidence interval of PVY-6, the two QTLs are not localized at exactly the same position. A lack of markers in this genomic region does not allow precise QTL mapping and this hypothesis should be evaluated further by adding new molecular markers.

Regarding CMV, a breeding programme has been developed in order to increase CMV resistance in pepper. The reduction of the number of local lesions at inoculation has been used as a resistance trait that was introgressed into pepper breeding lines also expressing resistance to CMV systemic movement and accumulation within the plant. The final pepper lines were close to virus immunity, which was not the case for pepper lines that only carry resistance to CMV systemic movement and/or accumulation [28]. This suggests that reducing N_e at inoculation increases plant resistance substantially, at least when combined with resistance to systemic movement and accumulation.

In this study, we showed that N_e at inoculation is a well-diversified and highly heritable trait among a DH population of *C. annuum*. We then identified and mapped the pepper QTLs controlling N_e for both PVY and CMV. Finally, comparison with previous results showed that one of these QTLs may also decrease virus adaptation to a major-effect resistance gene. From an agricultural point of view, the use of these QTLs could be even more beneficial, because some of them could act against several viruses belonging to distant groups. Therefore, our results suggest that these plant genetic factors could correspond to general mechanisms and be efficient against multiple pathogens.

METHODS

Plant and virus material

A DH population of *C. annuum* was obtained from the F_1 hybrid between Yolo Wonder, a line susceptible to PVY isolates, and Perennial, a cultivar carrying the PVY resistance allele *pvr2*^{3}. A genetic map comprising 190 molecular markers was previously built for this progeny [19]. From this population, we phenotyped 152 DH lines carrying *pvr2*^{3} and differing in their genetic background.

The first virus used was derived from a cDNA clone of PVY isolate SON41p. It carried the 115K substitution in the VPg, which allows it to overcome the *pvr2*^{3} resistance [29]. The virus was also tagged with the GFP reporter gene. The PVY-GFP was constructed by duplicating the NIa protease cleavage site at the C-terminus of the NIb cistron and inserting the GFP gene between the two sites,

allowing the virus NIa protease to cleave the GFP. For the purpose of a control experiment, the same PVY infectious clone carrying the mCherry reporter gene (expressing a red fluorescent marker) instead of the GFP gene was used. The second virus was the CMV-N strain of Fulton, a variant of CMV causing necrotic local lesions on the leaves. A third virus, ToMV (isolate Vi76), was used for the inoculation of eight representative DH lines. ToMV induces necrotic local lesions on the leaves of pepper plants carrying the L^1 resistance allele, harboured by the Yolo Wonder parent.

Evaluation of the N_e estimation method

A control experiment similar to the one developed by Zwart *et al.* was performed to assess the reliability of the N_e estimation method [21]. We used a mixture of the PVY variants carrying either the GFP or the mCherry fluorescent marker, which had been multiplied separately in *Nicotiana clevelandii* plants beforehand. Twenty-eight days after sowing, the 1 : 1 mixture was inoculated on the first true leaf of 13 DH lines selected for showing contrasting numbers of foci induced by the PVY-GFP variant alone (Fig. S1). The number of foci showing green and/or red fluorescence was estimated under a specific light wavelength (450–490 nm) at 4 and 6 days p.i. Finally, the surface area of the cotyledons was measured using ImageJ software, with 10 cotyledons for each DH line.

Counting the primary infection foci and local lesion numbers induced by PVY, CMV and ToMV in a DH population

All of the DH lines were mechanically inoculated with either PVY or CMV for the QTL detection. Based on the results of the QTL detection, a subset of eight DH lines was selected and inoculated with ToMV in order to compare the previous results with a third virus.

The PVY-GFP cDNA clone was first inoculated in *N. clevelandii* plants by DNA-coated tungsten particle bombardment. In order to obtain the inoculum, extracts of these plants were then used to propagate the virus in *Nicotiana tabacum* cv. Xanthi plants. Leaves from infected Xanthi plants were ground in phosphate buffer (0.03 M Na_2HPO_4, 0.2 % sodium diethyldithiocarbamate; 4 ml buffer per gramme of leaves) supplemented with active charcoal (90 mg ml^{-1}) and carborundum (90 mg ml^{-1}) to obtain the final inoculum. A single inoculum was used to mechanically inoculate 10 plants on their two cotyledons for each pepper DH line, 3 weeks after sowing. At 6 days p.i., the number of primary infection foci on each inoculated cotyledon was counted under a specific light wavelength (450–490 nm) (Fig. 1a). All the plants were grown under greenhouse conditions.

The CMV-N strain of Fulton was propagated on *Catharanthus roseus* plants. From extracts of these plants, 10 pepper plants per DH line were mechanically inoculated on their first leaf, 4 weeks after sowing. At 5 days p.i., the number of necrotic local lesions per inoculated leaf was counted (Fig. 1b). The experiment was realized in a climate-controlled room (20–22 °C, 12 h light day^{-1}).

The ToMV test was performed in a detached leaf assay to avoid contamination, since ToMV is readily transmitted by contact. Five weeks after sowing, the first leaf from 12 plants per DH line was excised, mechanically inoculated and kept in a humid plastic box. At 5 days p.i., the number of necrotic local lesions per inoculated leaf was counted.

Statistical analyses

The statistical analyses were performed using R software (http://www.r-project.org/). For the two phenotypic traits (number of infection foci and local lesions initiated by PVY or CMV), broad-sense heritability was estimated using the formula $h^2 = \sigma^2_G / (\sigma^2_G + \sigma^2_E / n)$, where σ^2_G corresponds to the genotypic variance, σ^2_E corresponds to the environment variance and n corresponds to the number of replicates (n=20 for PVY and 10 for CMV). An ln(x+1) transformation was applied to the two traits to approximate a normal distribution.

QTL analysis

QTLs detection was performed with the R/qtl software package [30]. A preliminary analysis was realized by using a standard interval mapping approach. In addition, a two-dimensional genome scan was performed to identify potential interactions between QTLs. Multiple QTL mapping (MQM) was then performed, using the markers previously identified as the initial set of cofactors. Finally, the positions and the effects of the QTLs were refined in the context of a multiple QTL model. The significance LOD threshold was calculated by performing a permutation test with 10 000 replicates. The LOD threshold was set at 3.79 for the foci induced by PVY and 3.18 for the lesions induced by CMV (P=0.05). The confidence intervals for the location of each QTL were determined by using the 1-LOD and 2-LOD drop-off methods. The graphical representation of the QTLs was generated using MapChart version 2.3 [31].

Funding information
L.T.'s PhD is supported by the BAP (Biologie et Amélioration des Plantes) department and the SMaCH (Sustainable Management of Crop Health) metaprogramme of INRA, and by the Région Provence-Alpes-Côte d'Azur (PACA). The experimental work was supported by the SMaCH metaprogramme.

Acknowledgements
We thank the greenhouse staff for their support with plant experimentation.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Mcdonald BA, Linde C. Pathogen population genetics, evolutionary potential, and durable resistance. *Annu Rev Phytopathol* 2002;40:349–379.
2. Rojas MR, Gilbertson RL. Emerging plant viruses: a diversity of mechanisms and opportunities. In: *Plant Virus Evolution*. Berlin, Heidelberg: Springer; 2008. pp. 27–51.

3. Elena SF, Bedhomme S, Carrasco P, Cuevas JM, de La Iglesia F *et al.* The evolutionary genetics of emerging plant RNA viruses. *Mol Plant Microbe Interact* 2011;24:287–293.
4. Brown JK. Durable resistance of crops to disease: a Darwinian perspective. *Annu Rev Phytopathol* 2015;53:513–539.
5. García-Arenal F, Fraile A, Malpica JM. Variability and genetic structure of plant virus populations. *Annu Rev Phytopathol* 2001; 39:157–186.
6. Elena SF, Sanjuán R. Virus evolution: insights from an experimental approach. *Annu Rev Ecol Evol Syst* 2007;38:27–52.
7. Sacristán S, Malpica JM, Fraile A, García-Arenal F. Estimation of population bottlenecks during systemic movement of tobacco mosaic virus in tobacco plants. *J Virol* 2003;77:9906–9911.
8. Rouzine IM, Rodrigo A, Coffin JM. Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiol Mol Biol Rev* 2001;65:151–185.
9. Li H, Roossinck MJ. Genetic bottlenecks reduce population variation in an experimental RNA virus population. *J Virol* 2004;78: 10582–10587.
10. Moury B, Fabre F, Senoussi R. Estimation of the number of virus particles transmitted by an insect vector. *Proc Natl Acad Sci USA* 2007;104:17891–17896.
11. Betancourt M, Fereres A, Fraile A, García-Arenal F. Estimation of the effective number of founders that initiate an infection after aphid transmission of a multipartite plant virus. *J Virol* 2008;82: 12416–12421.
12. Sacristán S, Díaz M, Fraile A, García-Arenal F. Contact transmission of tobacco mosaic virus: a quantitative analysis of parameters relevant for virus evolution. *J Virol* 2011;85:4974–4981.
13. González-Jara P, Fraile A, Canto T, García-Arenal F. The multiplicity of infection of a plant virus varies during colonization of its eukaryotic host. *J Virol* 2009;83:7487–7494.
14. Gutiérrez S, Pirolles E, Yvon M, Baecker V, Michalakis Y *et al.* The multiplicity of cellular infection changes depending on the route of cell infection in a plant virus. *J Virol* 2015;89:9665–9675.
15. Zwart MP, Elena SF. Matters of size: genetic bottlenecks in virus infection and their potential impact on evolution. *Annu Rev Virol* 2015;2:161–179.
16. Gutiérrez S, Michalakis Y, Blanc S. Virus population bottlenecks during within-host progression and host-to-host transmission. *Curr Opin Virol* 2012;2:546–555.
17. Acosta-Leal R, Xiong Z. Complementary functions of two recessive R-genes determine resistance durability of tobacco 'Virgin A Mutant' (VAM) to Potato virus Y. *Virology* 2008;379:275–283.
18. Marczewski W, Flis B, Syller J, Schäfer-Pregl R, Gebhardt C. A major quantitative trait locus for resistance to Potato leafroll virus is located in a resistance hotspot on potato chromosome XI and is tightly linked to N-gene-like markers. *Mol Plant Microbe Interact* 2001;14:1420–1425.
19. Quenouille J, Paulhiac E, Moury B, Palloix A. Quantitative trait loci from the host genetic background modulate the durability of a resistance gene: a rational basis for sustainable resistance breeding in plants. *Heredity* 2014;112:579–587.
20. Quenouille J, Montarry J, Palloix A, Farther MB. Slower, stronger: how the plant genetic background protects a major resistance gene from breakdown: mechanisms of polygenic resistance durability. *Mol Plant Pathol* 2013;14:109–118.
21. Zwart MP, Daròs JA, Elena SF. One is enough: in vivo effective population size is dose-dependent for a plant RNA virus. *PLoS Pathog* 2011;7:e1002122.
22. Chao L. Fitness of RNA virus decreased by Muller's ratchet. *Nature* 1990;348:454–455.
23. Bald JG. The use of numbers of infections for comparing the concentration of plant virus suspensions: dilution experiments with purified suspensions. *Ann Appl Biol* 1937;24:33–55.
24. Kleczkowski A. Interpreting relationships between the concentrations of plant viruses and numbers of local lesions. *J Gen Microbiol* 1950;4:53–69.
25. Lauffer MA, Price WC. Infection by viruses. *Arch Biochem* 1945;8: 449–450.
26. García-Arenal F, Fraile A. Population dynamics and genetics of plant infection by viruses. In: *Recent Advances in Plant Virology*. Norfolk, UK: Caister Academic Press; 2011.
27. Caranta C, Palloix A, Lefebvre V, Daubèze AM. QTLs for a component of partial resistance to cucumber mosaic virus in pepper: restriction of virus installation in host-cells. *Theor Appl Genet* 1997;94:431–438.
28. Djian-Caporalino C, Lefebvre V, Sage-Daubèze A-M, Palloix A. Capsicum. In: Singh RJ (editor). *Genetic Resources, Chromosome Engineering, and Crop Improvement: Vegetable Crops*. Florida, USA: CRC Press; 2006. pp. 185–243.
29. Ayme V, Souche S, Caranta C, Jacquemond M, Chadoeuf J *et al.* Different mutations in the genome-linked protein VPg of P*otato virus Y* confer virulence on the *pvr2*^{3} resistance in pepper. *Mol Plant Microbe Interact* 2006;19:557–563.
30. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 2003;19:889–890.
31. Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 2002;93:77–78.

Supplementary Material

Table S1: Mean number and frequency (%) of foci expressing green fluorescence only (GFP), red fluorescence only (mCherry) or both (GFP & mCherry) for each DH line. The last column shows the average estimated number of viral particles initiating an infection focus, according to the formulas of Sacristán et al. [7]. The last row indicates the average values among the 49 plants.

DH line	Number of plants	Mean number of foci				Mean frequency of foci (%)			Mean number of viral particles initiating a focus
		Total	GFP	mCherry	GFP & mCherry	GFP	mCherry	GFP & mCherry	
2400	2	3.0	1.5	1.5	0.0	50.0	50.0	0.0	1.00
219	5	6.4	2.0	4.4	0.0	31.3	68.8	0.0	1.00
2256	2	7.0	3.0	4.0	0.0	42.9	57.1	0.0	1.00
2321	1	14.0	4.0	10.0	0.0	28.6	71.4	0.0	1.00
2344	4	24.0	6.8	17.3	0.0	28.1	71.9	0.0	1.00
221	2	56.5	12.5	44.0	0.0	22.1	77.9	0.0	1.00
2173	5	50.2	12.6	37.4	0.2	25.1	74.5	0.4	1.01-1.01
2349	5	51.2	7.8	43.0	0.4	15.2	84.0	0.8	1.01-1.03
2264	8	15.0	3.4	11.5	0.1	22.5	76.7	0.8	1.01-1.02
2426	2	46.0	10.5	35.0	0.5	22.8	76.1	1.1	1.02-1.03
2367	3	59.7	16.7	42.3	0.7	27.9	70.9	1.1	1.02-1.02
2430	8	50.9	12.5	37.5	0.9	24.6	73.7	1.7	1.03-1.04
2328	2	42.5	9.0	32.5	1.0	21.2	76.5	2.4	1.04-1.06
All	***49***	***32.80***	***7.86***	***24.64***	***0.29***	***27.87***	***71.50***	***0.64***	***1.01-1.02***

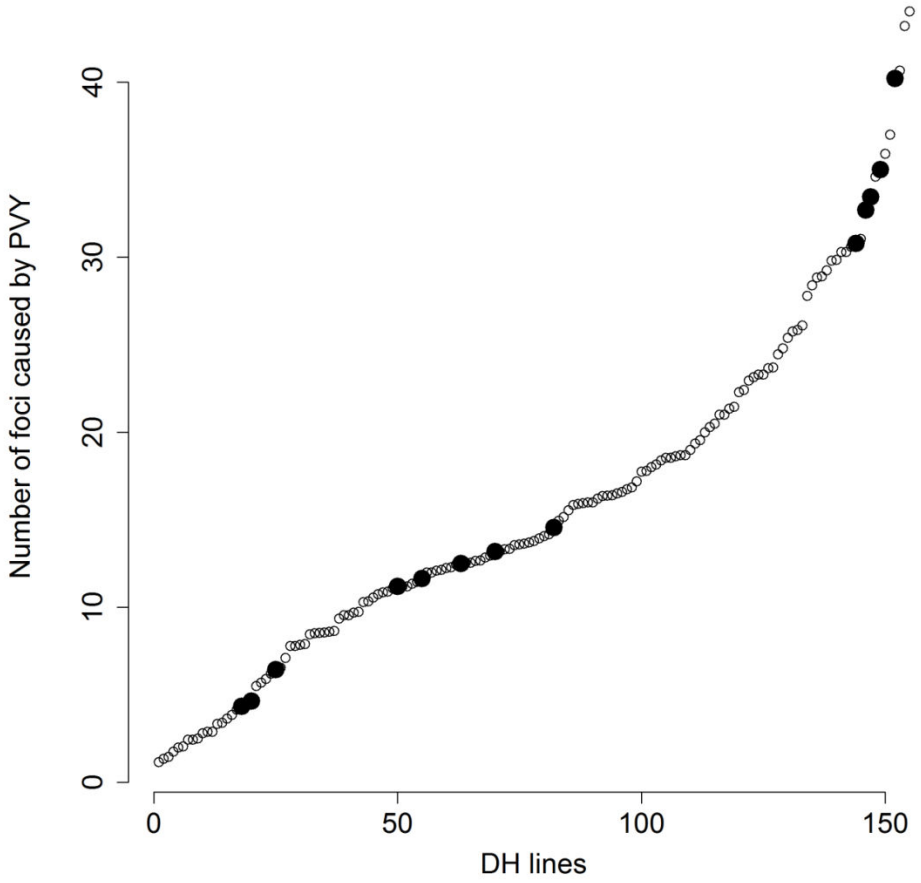


Figure S1: Distribution of the 152 DH lines according to the mean number of primary infection foci caused by PVY-GFP. The 13 DH lines used for the evaluation of the N_e estimation method with PVY-GFP and PVY-mCherry are represented by a filled black circle.

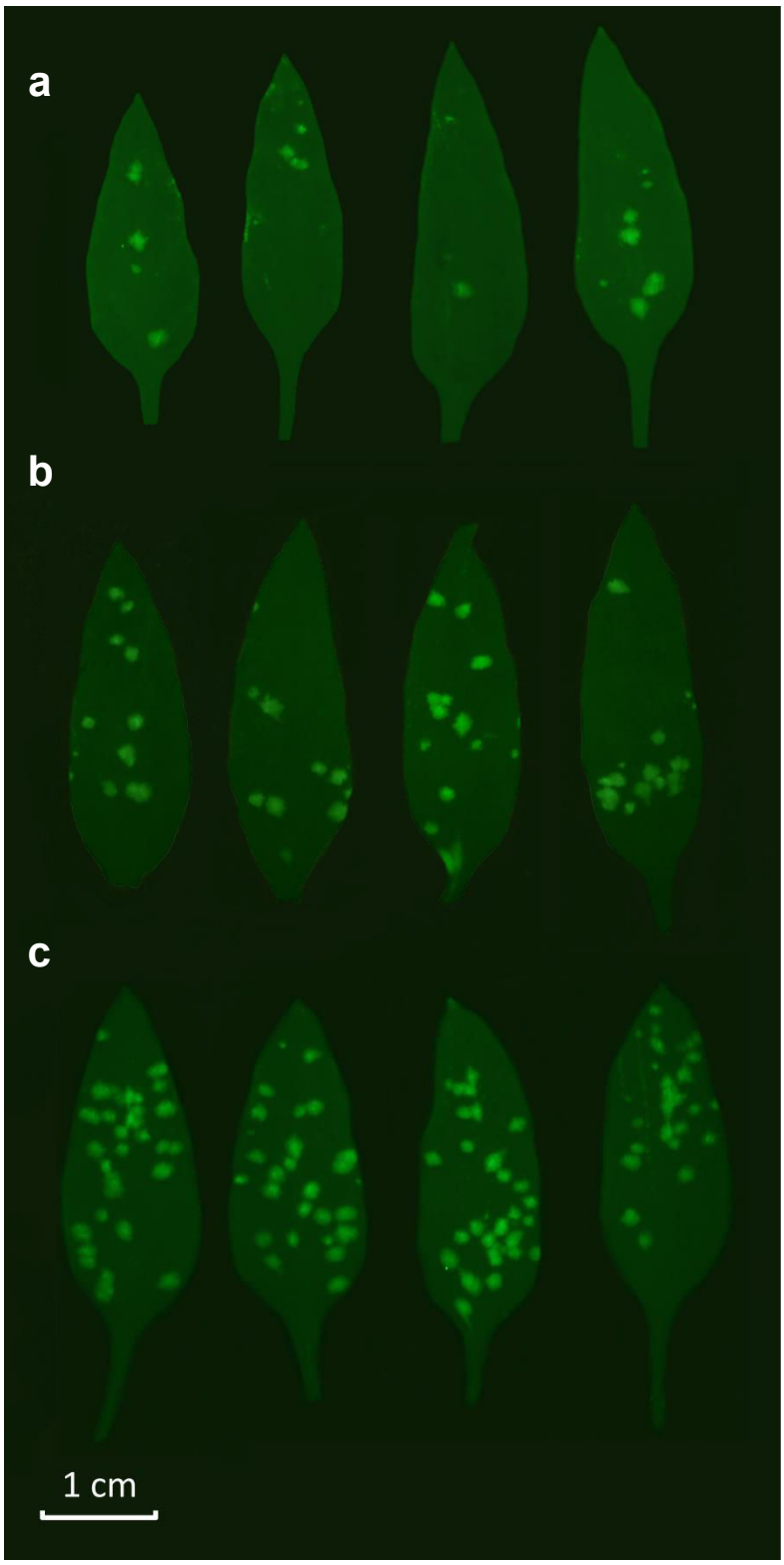


Figure S2: Primary infection foci induced by PVY-GFP on pepper cotyledons of three DH lines: HD290 showing a few number of foci (a), HD2295 showing an average number of foci (b) and HD221 showing a high number of foci (c).

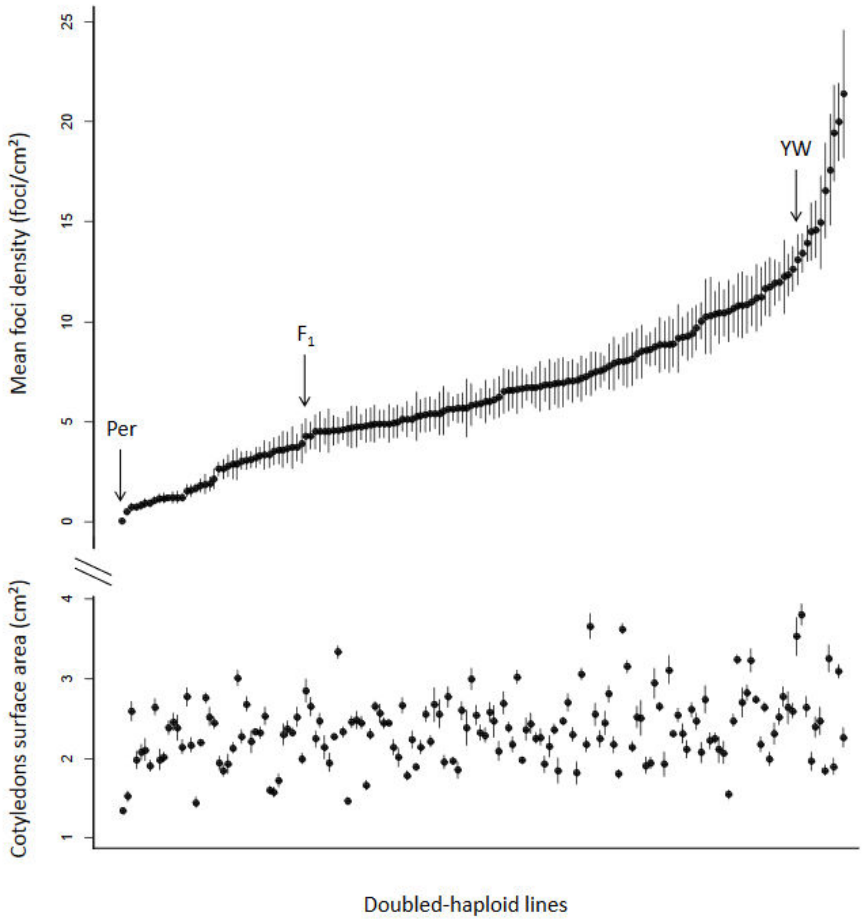


Figure S3: Mean foci density induced by PVY-GFP (foci/cm²) and cotyledon surface area (cm²) among the DH lines. Error bars represent the standard error of the mean. The position of the parental lines Perennial (Per) and Yolo Wonder (YW) and of the F1 hybrid are indicated.

Chapitre 2

Genome-wide association mapping of QTLs implied in *Potato virus Y* population sizes in the pepper germplasm

Résumé de l'article :

Objectifs : Le premier objectif était de sélectionner une core-collection de piment représentative des ressources génétiques disponibles dans la collection entière et de génotyper cette core-collection afin d'obtenir des marqueurs de qualité et suffisamment nombreux pour effectuer de la génétique d'association. Le second objectif était d'identifier les régions génomiques responsables de la taille efficace de population (N_e) du virus à l'inoculation et de l'accumulation virale dans la plante entière. Enfin, le troisième objectif était de comparer les résultats avec ceux précédemment obtenus pour les mêmes traits phénotypiques par détection de QTL dans une population haploïde doublée issue d'un croisement biparental.

Stratégie : Sur la base de marqueurs microsatellites qui avaient été utilisés pour effectuer le génotypage de la collection entière de piment, un algorithme maximisant la diversité génétique a permis de sélectionner une core-collection. Cette core-collection a été phénotypée pour deux traits : le N_e des populations virales à l'inoculation et l'accumulation virale dans la plante entière. Les accessions de piment ont été inoculées sur les cotylédons avec un clone infectieux de PVY exprimant la protéine fluorescente GFP. Pour chaque accession, le nombre de foyers fluorescents a été quantifié et utilisé comme proxy de N_e. L'accumulation virale a quant à elle été estimée en inoculant les accessions avec un clone infectieux de PVY et en mesurant sa concentration dans la plante par ELISA quantitatif. Une approche de génotypage par séquençage (GBS) a été utilisée pour identifier des SNPs le long du génome du piment. Finalement, les régions génomiques liées aux traits d'intérêts ont été identifiées par génétique d'association.

Résultats principaux : Une core-collection de 276 accessions de piment a été échantillonnée et un total de 10308 marqueurs SNP a été obtenu. La génétique d'association a permis de mettre en évidence 6 SNP et 3 régions génomiques impliqués dans le contrôle d'au moins un des deux traits étudiés. Ces régions sont en partie similaires à celles détectées pour les mêmes traits chez une population haploïde doublée par cartographie de QTL. Les résultats révèlent également l'existence d'une région génomique sur le chromosome 6 impliquée à la fois dans le contrôle du N_e des populations virales à l'inoculation et de l'accumulation virale.

Genome-wide association mapping of QTLs implied in *Potato virus Y* population sizes in the pepper germplasm

Tamisier, L. (1,2), Nemouchi, G. (1), Szadkowski, M. (1), Duboscq, R. (1), Santoni, S. (3), Sarah, G. (3), Sauvage, C. (1), Moury, B. (2), Palloix, A. (1).

(1) GAFL, INRA, 84140 Montfavet, France

(2) Pathologie Végétale, INRA, 84140 Montfavet, France

(3) UMR AGAP, INRA, F-34060 Montpellier, France

Keywords: Genome-wide association, genotyping-by-sequencing, effective population size, viral accumulation, *Capsicum annuum*, *Potato virus Y*, quantitative resistance.

Abstract

The capacity of pathogen populations to adapt to plant resistance genes is mostly influenced by the plant genetic background. In a pepper population obtained from a biparental cross, genetic factors decreasing the number of viruses entering within the plant (*i.e.* effective population size at inoculation) and the virus accumulation (*i.e.* census population size) have been detected and shown to be involved in the protection of a major resistance gene from breakdown. To evaluate the representativeness of these factors, we looked at the availability of genetic factors affecting the same traits in the pepper germplasm. We used a genotyping-by-sequencing approach in a core-collection of 276 accessions of *Capsicum annuum* and obtained 10,308 single-nucleotide polymorphism (SNP) markers. Genome-wide association analysis detected 6 SNPs significantly associated with the virus population size at inoculation and/or the census virus population size. Interestingly, the same associated SNPs were found on chromosome 6 for both traits, highlighting that the same genomic region controls in part both population sizes, potentially through a pleiotropic effect. These 2 SNPs are also localized within the confidence interval of a QTL controlling virus census population size previously identified with the biparental population. Two SNPs on chromosome 4 also associated to both PVY population sizes map close to the major resistance gene *pvr2* encoding an eIF4E. Finally, one SNP on chromosome 12 was associated to PVY effective population size at inoculation and localized close to a QTL controlling the same trait in the biparental population. Altogether, this

study (i) shows the efficiency of GBS and GWAS in *C. annuum*, (ii) indicates highly consistent results between GWAS and classical QTL mapping and (iii) suggests that PVY resistance QTL identified with a biparental population are representative of a much larger collection of *C. annuum* accessions and widespread in that collection.

1 Introduction

Cultivars carrying a major resistance gene have been extensively deployed to protect crops from pathogens. They have been widely used because they provide an almost complete resistance against pathogens, they are environmentally friendly and they can be easily introgressed by backcrossing. The major limit of this type of genetic control relies on the evolutionary potential of the pathogen population (McDonald and Linde 2002; García-Arenal and McDonald 2003). Pathogen populations showing features such as high mutation rate, large effective population size and/or strong gene flow produce a lot of genetic diversity and pose the highest risk of evolution. At the same time, the widespread deployment of resistance genes in genetically uniform monocultures imposes a strong directional selection on the pathogen population, which can lead to the selection of better adapted pathogens and to the breakdown of the resistance gene. Several strategies have been proposed to control the pathogen evolution and limit its adaptation to the resistant plant, including pyramiding of several resistance genes or rotation of crops carrying different resistance genes (Mundt 2002; Pink 2002; Zhan et al. 2015). Among these strategies, the combination of major resistance genes with quantitative resistance factors (*i.e.* polygenic resistance) seems to be a promising alternative to insure a durable crop protection. A greater durability of polygenic resistance compared to monogenic resistance has been demonstrated in three pathosystems involving virus, fungus and nematode (Palloix et al. 2009; Brun et al. 2010; Fournet et al. 2013). All these studies have observed a higher breakdown of the resistance gene when it was introgressed into a susceptible genetic background compared to a partially-resistant one, probably because of the protective effect of the partially-resistant genetic background on the major gene.

In the *Potato virus Y* (PVY) – pepper (*Capsicum annuum*) system, the greater durability of the polygenic resistances is mostly due to the additional level of resistance conferred by the genetic background, which reduces viral accumulation (Quenouille et al. 2013). Smaller effective population sizes and greater genetic drift imposed by the genetic background were also probably involved in this higher resistance level. It means that, in small viral populations,

the frequencies of the different viral strains within the population will randomly fluctuate over time independently of their fitness, allowing the loss of the most adapted virus variants by chance. Several QTL controlling virus accumulation or the effective population size have been identified using a pepper doubled-haploid (DH) population of 153 lines. All the lines were carrying the *pvr2*^{3} major gene, which confers resistance against PVY, but were segregating for the genetic background. Two QTLs controlling PVY accumulation in the plant were detected and shown to be linked to the breakdown frequency of the major gene (Quenouille et al. 2014). Three QTLs controlling the effective population size of the virus during the inoculation of the leaf were identified, one of them potentially colocalizing with a QTL controlling virus accumulation on chromosome 6 (Tamisier et al. 2017). However, the confidence intervals of the two QTLs were large and barely overlapped, and it was therefore impossible to determine if the same genomic region was involved in both traits. Indeed, even if it is a powerful approach, QTL detection using a biparental population has two major limits. Firstly, for each locus, the explored diversity is low and corresponds to the polymorphisms distinguishing the parents (for example, a maximum of two alleles may differ in progenies of 2 parental inbred lines for diploids). Secondly, the few recombination events in the population genealogy can limit the mapping resolution and conduce to large QTL support intervals (Korte and Farlow 2013). This is especially true for DH populations which are the result of only one efficient meiosis. In this regard, Genome-Wide Association Studies (GWAS) can be a complementary approach. Since GWAS are usually applied to large collections of theoretically unrelated individuals, the genetic diversity is supposed to be high and new alleles can be discovered. Furthermore, the high number of meioses that have occurred in the genealogy of the GWAS population can allow a precise QTL mapping (Hamblin et al. 2011). For several plant species, the use of both GWAS and QTL mapping approaches have provided good results to dissect the genetic architecture of traits such as plant agronomic features in soyabean (Sonah et al. 2015) or resistance to *Plum pox virus* (PPV) in *Arabidopsis thaliana* (Pagny et al. 2012).

Recently, the availability of genotypes showing a partially-resistant genetic background in the pepper germplasm has been proved (Quenouille et al. 2015). Among a collection of 20 pepper accessions, a high diversity of resistance levels conferred by the genetic background was observed, including for virus accumulation. In this context, the aim of the present study was to (i) select a core-collection of pepper accessions and genotype them with a genotyping-by-sequencing (GBS) approach, (ii) perform GWAS for PVY effective population size at

inoculation and PVY accumulation and (iii) compare the results with those obtained by QTL mapping with a biparental DH population.

2 Results

2.1 A core-collection representative of the pepper germplasm

The pepper germplasm (*Capsicum* spp.) maintained in the Institut National de la Recherche Agronomique (INRA) in Avignon is composed of 1,352 non-redundant accessions, including 11 cultivated and wild species, which are originated from 89 countries and 5 continents (Sage-Palloix et al. 2007). The majority of the pepper collection (78.6%) is composed of *C. annuum* inbred lines that can be divided into the cultivated species *C. annuum* var. *annuum*, which represents more than 90% of the *C. annuum* accessions, and the wild species *C. annuum* var. *glabriusculum*. All the accessions have been previously genotyped with a set of 28 SSR (Simple Sequence Repeats) markers (Nicolaï et al. 2012; Nicolaï et al. 2013). They have also been tested for their resistance to 3 pathotypes of PVY, characterized according to their capacity to infect plants carrying different alleles at the *pvr2* resistance gene. The pathotype PVY-0 infects accessions carrying the susceptibility allele *pvr2+*. The pathotype PVY-0,1 infects plants carrying the *pvr2+* or $pvr2^1$ alleles, and the pathotype PVY-0,1,2,3 infect plants carrying the *pvr2+*, $pvr2^1$, $pvr2^2$ or $pvr2^3$ alleles.

Our goal was to select a core-collection of *C. annuum* maximizing the genetic and PVY resistance diversity. From the whole collection, we excluded all the *C. annuum* accessions resistant to pathotype PVY-0,1,2,3 in order to get rid of the effect of *pvr2*-mediated resistance during the phenotyping and to measure the resistance conferred by the genetic background only. In addition, this choice ensured that the huge majority of the selected accessions would be infected during phenotyping and that we could measure quantitative resistance. Genetically-redundant accessions displaying the same alleles at every SSR locus were also removed from the analysis. Eventually, we kept a final set of 25 SSR in order to have less than 25% of missing data for each accession. The core-collection was then built using the MStrat software v.4.1 (Gouesnard et al. 2001). We first estimated the optimal size for the core-collection by calculating the increase of allelic richness for different core-collection sizes. The plateau of this curve indicates the minimum number of accessions needed to keep the same genetic diversity as the germplasm. The plateau was reached for 370 accessions but the allelic richness was very

close for 310 accessions, with an allelic richness of 317 and 316 respectively (Figure S1). We then applied the Maximization strategy algorithm and obtained a core-collection of 310 accessions which captured 91% of the alleles of the *C. annuum* collection (Table 1). The average number of alleles per locus in the core-collection remains high relative to the whole collection (> 12) as well as Nei's unbiased gene diversity index (He) which is constant (0.59). The observed heterozygosity remains unchanged and low (0.035), a result explained by the multiplication through selfing of the accessions and the preferential autogamy of *C. annuum*. Previously, the structure of the whole *Capsicum* spp. collection has revealed 6 distinct clusters (Nicolaï et al. 2013). The cultivated *C. annuum* var. *annuum* accessions were distributed across the clusters 1, 2 and 3 while the wild species *C. annuum* var. *glabriusculum* accessions were split in several clusters.

Table 1: Comparison of the genetic diversity of the whole collection of *C. annuum* and the core-collection obtained with the Maximization strategy algorithm. For each collection, the table provides the number of accessions, the percentage of SSR alleles represented in the collections, the mean number of alleles observed per locus, Nei's unbiased gene diversity (He), observed heterozygosity (Ho) and the percentage of accessions belonging to each cluster defined by STRUCTURE 2.3.4 software based on SSR markers (Nicolaï et al. 2013).

Sample	Sample size	% SSR alleles	Allele number	He	Ho	Distribution in the *C. annuum* clusters			
						Cluster 1	Cluster 2	Cluster 3	Clusters 4-6
Whole collection of *C. annuum*	887	100	12.57	0.59	0.035	36% (314)	21% (190)	41% (367)	2% (16)
Core-collection of *C. annuum*	310	91	12.07	0.59	0.035	30% (93)	23.6% (73)	42.9% (133)	3.5% (11)

The proportion of accessions belonging to each cluster was not significantly different between the whole and core collections (Fisher exact test, $p = 0.13$, Table 1), attesting that the core-collection sampled with the Maximization strategy algorithm is an accurate representation of the *C. annuum* collection. Finally, we were able to perform genotyping-by-sequencing (GBS) on 276 accessions among the core-collection and to phenotype this final core-collection in order to carry out the GWAS approach (Figure S1).

2.2 Distribution of SNPs in the pepper genome

The genotyping of the core-collection was performed by double-digest restriction associated DNA sequencing (ddRADseq) (Peterson et al. 2012). This method uses two restriction enzymes to reduce the genome complexity, one of them is a rare cutter and the other cleaves DNA fragments more frequently. Then, the cleaved DNA fragments that are surrounded by two different restriction sites (each one belonging to one of the two restriction enzymes) are selected according to their size and sequenced. This protocol was chosen because it offers two major advantages. Firstly, since a small proportion of DNA fragments is expected to be surrounded by the two different restriction sites and to fall in the select size, it reduces the probability of sequencing both sides of the same restriction site and increases the read depth. Secondly, because of the specificity of the DNA fragments selected, the same genomic regions will probably be sequenced between different accessions, which helps to perform alignments and polymorphism detection (Peterson et al. 2012).

After applying filters (see Materials and Methods), we have identified 10,308 SNPs including 680 InDels (Table 2). An average depth of 140x was obtained. Among the SNPs, the transition/transversion (Ts/Tv) ratio was 1.46. A bias toward transition substitutions of the same order (1.35) has been previously observed in another study on 222 *C. annuum* genotypes (Taranto et al. 2016). An average of 4 SNPs/Mb was found in the entire genome, varying between 3.18 SNPs/Mb on chromosome 10 to 6.01 SNPs/Mb on chromosome 2 (Table 2). Finally, a rapid linkage disequilibrium (LD) decay was observed across the genome, reaching 40 kb at $r^2 = 0.2$ (Figure S2).

2.3 Population structure and kinship relationships

The population structure of the core-collection was inferred with STRUCTURE 2.3.4 software. After applying Evanno et al. (2005) correction, the analysis revealed that the pepper accessions were divided into 4 clusters (K=4), with some admixture between the clusters (Figure 1a). Principal component analysis (PCA) confirmed this result and also showed the presence of 4 groups within the core-collection (Figure 1b). A neighbour-joining tree based on pairwise genetic distances between SNPs was constructed. The clusters were quite well separated within the tree which confirms the previous results (Figure 1c).

Table 2: Number and density of SNPs identified by ddRADseq in the pepper core-collection.

Chromosome	Length (bp)	Number of SNPs	SNPs/Mb
1	261,560,226	1027	3.93
2	166,118,313	998	6.01
3	241,745,451	1279	5.29
4	206,470,299	722	3.50
5	223,151,943	731	3.28
6	217,864,955	960	4.41
7	227,551,634	839	3.69
8	134,909,690	587	4.35
9	247,983,219	850	3.43
10	227,301,773	723	3.18
11	246,428,986	807	3.27
12	232,591,935	785	3.38

The first 3 clusters (green, blue and red) included all the cultivated species *C. annuum* var. *annuum*, while the fourth (yellow) was composed of the wild species *C. annuum* var. *glabriusculum* only. The genetic structure of the 4 clusters was strongly correlated to the phenotype of the accessions, a clustering pattern already observed by Nicolaï et al. (2013) with SSR markers on the same pepper germplasm. This observation was also reported by Taranto et al. (2016) using SNP markers on another pepper collection. In our study, the clusters differed significantly for several plant features such as the fruit shape and length, the fruit pericarp thickness, the flowering date or the number of leaves. The first cluster was mainly composed of sweet and large-fruited peppers, the second cluster was characterized by an early flowering and triangle and/or elongated fruits and the third cluster included small and elongated fruits. These differences result of the long-term selection imposed by farmers in multiple environments for quality traits but also for resistance to biotic and abiotic stresses. Finally, the fourth cluster showed very tiny ovoid fruits, a distinctive feature of the wild accessions belonging to this group.

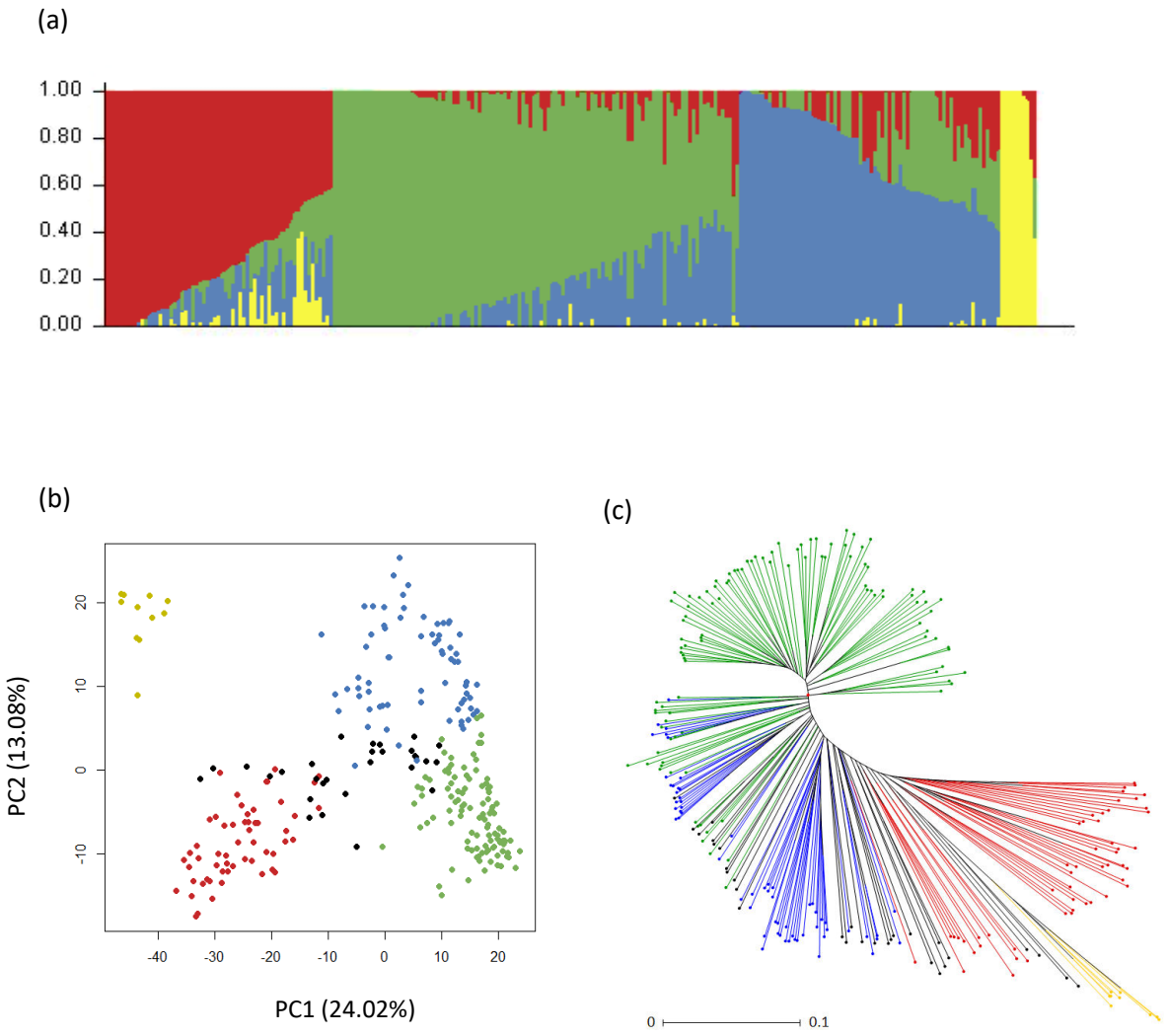


Figure 1: Population structure and genetic diversity of the *Capsicum annuum* germplasm core-collection (276 accessions) on the basis of SNPs. (a) Classification of the core-collection using STRUCTURE 2.3.4 software. Each vertical bar represents one pepper accession. (b) Principal component analysis of the core-collection. (c) Neighbor-joining phylogenetic tree of the core-collection. For the three plots the colors green, blue, red and yellow stand for the 4 clusters defined by STRUCTURE software. For the last two plots, the accessions in black display strong admixture (membership coefficient < 50% in each group).

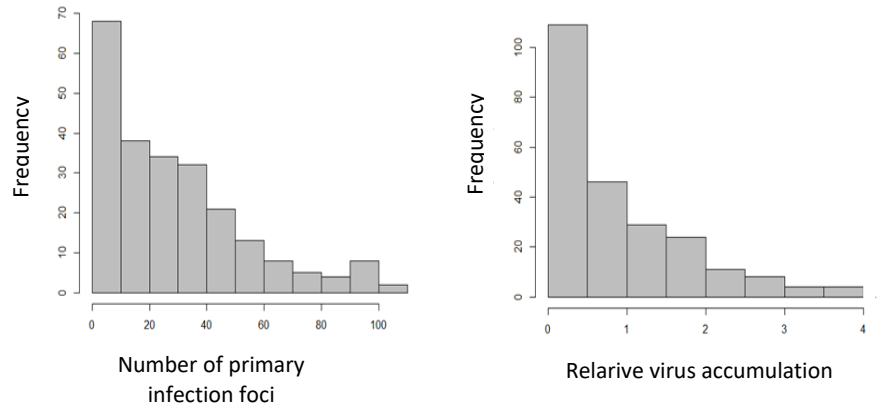
The relatedness among individuals was assessed using the algorithm of VanRaden (2008) implemented in GAPIT (Lipka et al. 2012) to calculate the kinship matrix. Across the 276 accessions, the pairwise kinship estimates were low, with a mean relatedness coefficient of 0.076. Fifty-four percent of the accessions showed a pairwise kinship estimate of 0 and only

1.8% of pairwise kinship coefficients were higher than 0.5, indicating that most of the accessions are unrelated (Figure S3).

2.4 Variation in the number of primary infection foci and the virus accumulation among the pepper core-collection

Two traits were phenotyped among the core-collection: the number of primary infection foci induced by a green fluorescent protein (GFP)-tagged PVY clone on the cotyledons and the relative virus accumulation (VA) of a variant of PVY isolate SON41p in the apical leaves. Both traits were highly variable and showed significant differences between the accessions ($P < 0.001$, Kruskal–Wallis test), attesting of the large diversity of quantitative resistances available in the pepper germplasm. The number of primary infection foci ranged from 0 to 104.25 with an overall mean number of 28.59±25.74 (mean ± standard deviation) (Figure 2a). The relative VA ranged from 0 to 3.72 with an overall mean number of 0.52±0.55. The two traits showed a weak but significant positive correlation (Spearman $\rho = 0.46$, $P < 0.001$). They also both showed a high broad-sense heritability, with h^2=0.98 for the number of primary infection foci and h^2=0.80 for the virus accumulation. The four genetic clusters previously identified were significantly contrasted for the number of foci and the level of virus accumulation (Figure 2b). The accessions belonging to the second cluster produce in average more infection foci than the others, while the accessions of the third cluster show in average a lower virus accumulation.

(a)



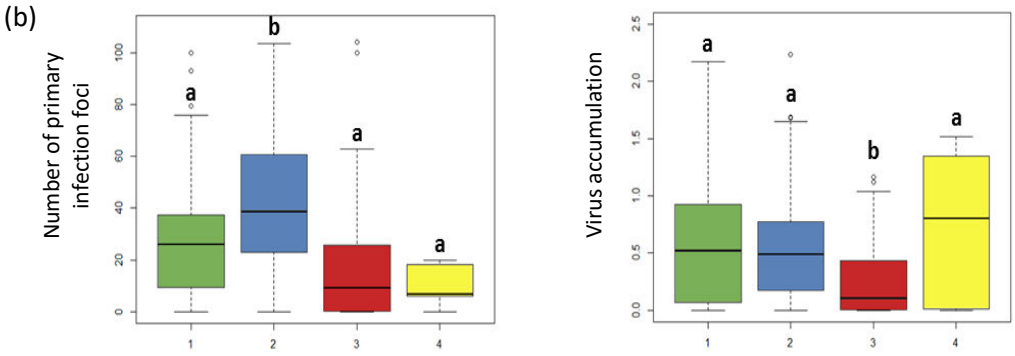


Figure 2: Phenotyping data of the PVY resistance in the pepper core-collection. (a) Frequency distribution of the number of primary infection foci caused by PVY-GFP and of the relative PVY accumulation in the *Capsicum annuum* core-collection. (b) Distribution of the number of primary infection foci and of the relative PVY accumulation among the 4 clusters determined by the structure analysis. The clusters are composed of 109, 71, 56 and 11 accessions, respectively. The letters a and b indicate the different groups obtained after pairwise comparisons using Nemenyi test ($P < 0.05$).

2.5 Genome-wide association mapping of pepper resistance to PVY

A GWAS was performed for the number of primary infection foci and the virus accumulation. Five SNPs localized on chromosomes 4, 6 and 12 were significantly associated with the number of foci (false discovery rate < 0.05) (Table 3, Figure 3a). Four SNPs localized on chromosomes 4 and 6 showed significant associations with virus accumulation. Among these markers, two SNPs on chromosome 6 and one SNP on chromosome 4 were detected for both traits. For the most significant SNP of each chromosome, the allelic effects on both traits were estimated (Figure 3b).

Table 3: SNPs identified with a genome-wide association study and associated with the number of primary infection foci induced by PVY-GFP and PVY accumulation in pepper.

Trait	Chromosome number	Position (bp)	R^2 of model without SNP	R^2 of model with SNP	P-value
Number of infection foci	6	234,143,013	0.212	0.366	2.40E-10
	6	234,142,995	0.212	0.365	2.74E-10
	4	1,151,249	0.212	0.293	2.94E-06
	4	340,333	0.212	0.28	1.58E-05
	12	235,513,719	0.212	0.289	5.15E-06
Virus accumulation	4	1,151,249	0.16	0.24	2.98E-06
	4	1,151,254	0.16	0.238	4.34E-06
	6	234,143,013	0.16	0.231	1.06E-05
	6	234,142,995	0.16	0.228	1.71E-05

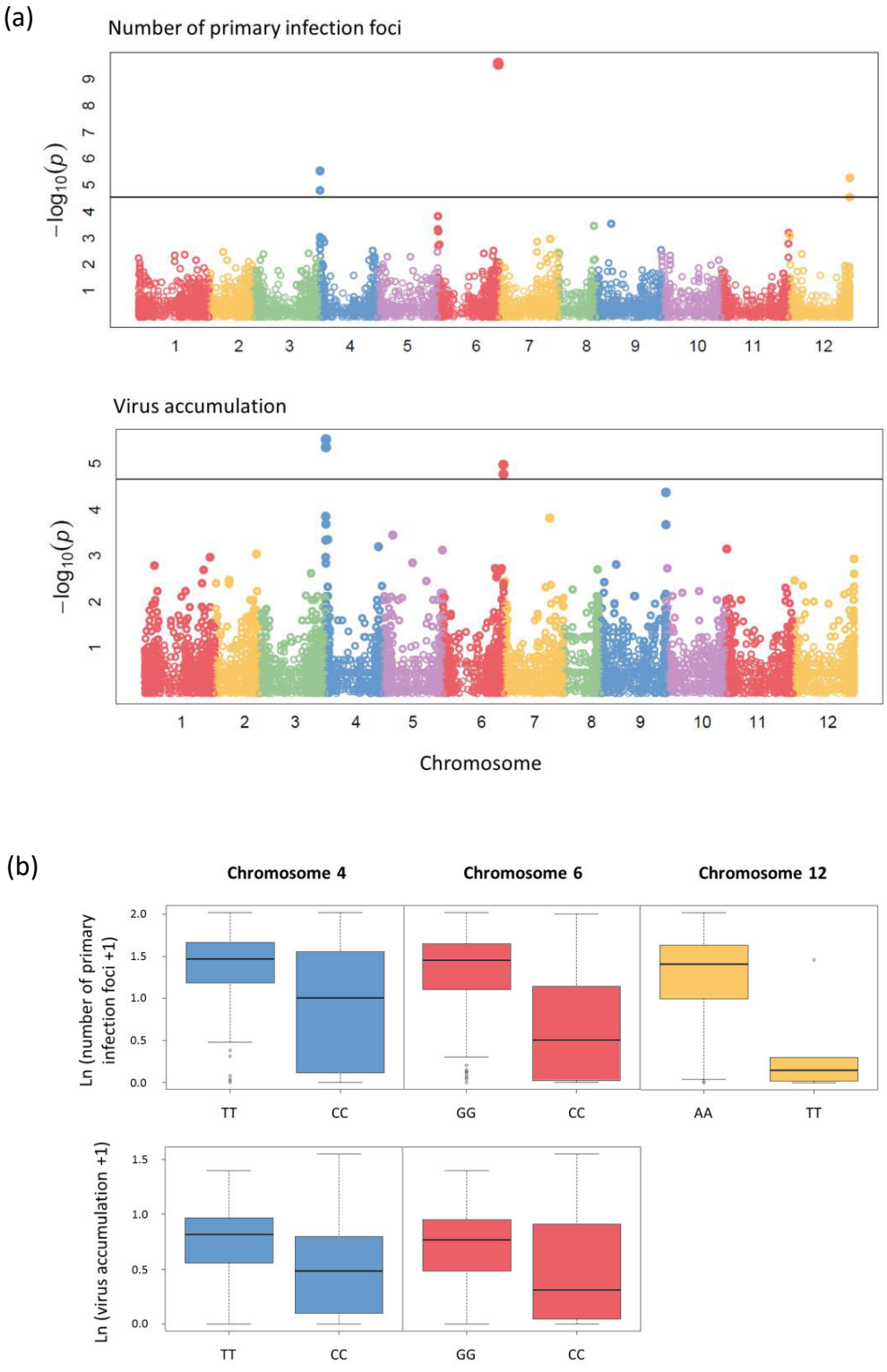


Figure 3: SNPs detected for the number of primary infection foci and the virus accumulation. (a) Manhattan plot of genome wide association study for the number of primary infection foci induced by PVY-GFP and the virus accumulation. Negative $\log_{10}$(*P*-values) from a genome-wide scan are plotted against SNP positions on each of the 12 chromosomes. (b) Allelic effect for the most significant associated markers on chromosome 4 and 6 for both traits, and on chromosome 12 for the number of foci.

These results were then compared with those previously obtained by QTL mapping with the biparental DH progeny for the same traits. For the number of primary infection foci, three QTLs have been previously detected on chromosomes 6 (PVY-6), 7 (PVY-7) and 12 (PVY-12) (Tamisier et al. 2017). No significant association was found on chromosome 7 with GWAS. The SNPs detected on chromosomes 12 and 6 were distant from approximately 10 Mb and 20 Mb of the QTLs PVY-12 and PVY-6, respectively (Figure 4, 5). Regarding virus accumulation, two QTLs have been previously identified: one QTL on chromosome 3 (VA-3), which colocalized with the *pvr6* gene, and one QTL on chromosome 6 (VA-6) (Quenouille et al. 2014). No significant association was found on chromosome 3 with GWAS. On chromosome 6, the confidence interval of VA-6 lies between the physical positions 214,459,454 and 235,745,825 bp of the CM334 reference genome. Both SNPs associated with the virus accumulation on this chromosome are included within this region (Figure 5). Finally, no QTL has been previously identified on chromosome 4. However, the *pvr2* locus is mapped between positions 1,193,194 and 1,197,686 bp on this chromosome. The SNP detected for both traits on chromosome 4 is localized at position 1,151,249 bp, which is the closest marker of the *pvr2* genomic region among all the SNPs (Figure 6).

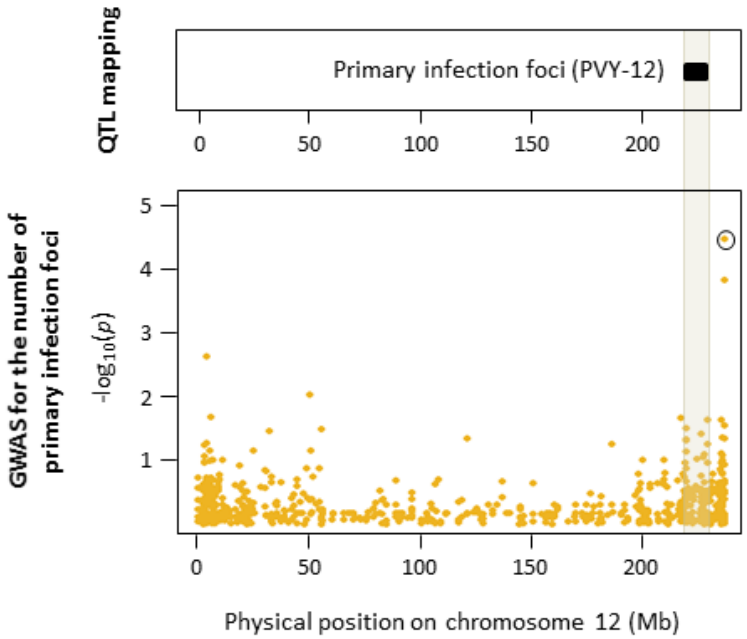


Figure 4: Genomic regions controlling the number of primary infection foci on chromosome 12. The PVY-12 (black) QTL detected by analysis of a biparental DH progeny in Tamisier et al. (2017) is mapped physically on pepper chromosome (top box). SNPs are represented in Manhattan plots displaying the $-\log_{10}(P\text{-values})$ over genomic positions (bottom boxes). SNPs showing significant association with the traits are surrounded by a black circle. The shaded gray area delimits the boundaries of the PVY-12 QTL.

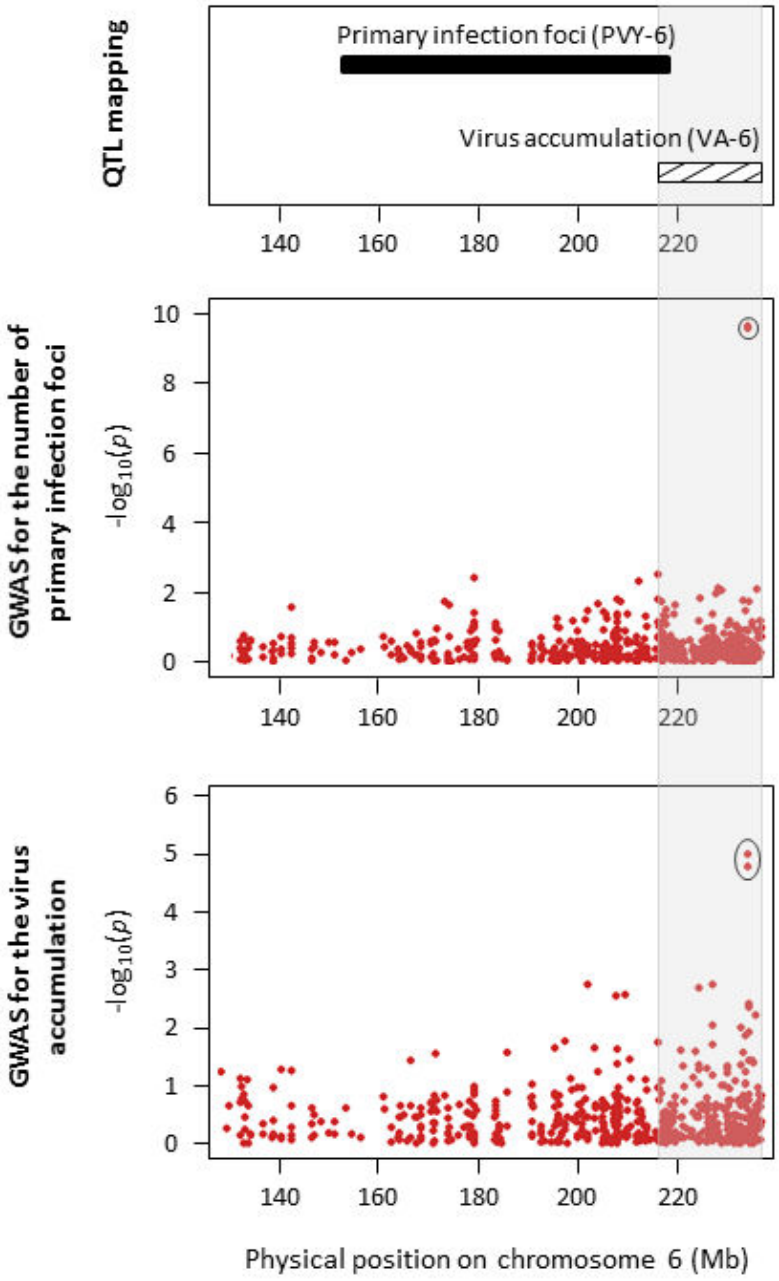


Figure 5: Genomic regions controlling the number of primary infection foci and the virus accumulation at the bottom of chromosome 6. The PVY-6 (black) and VA-6 (hatched) QTLs detected by the analysis of a biparental DH progeny in Tamisier et al. (2017) and Quenouille et al. (2014) are mapped physically on pepper chromosome (top box). For both traits, SNPs are represented in Manhattan plots displaying the $-\log_{10}(P\text{-values})$ over genomic positions (bottom boxes). SNPs showing significant association with the traits are surrounded by a black circle. The shaded gray area delimits the boundaries of the VA-6 QTL.

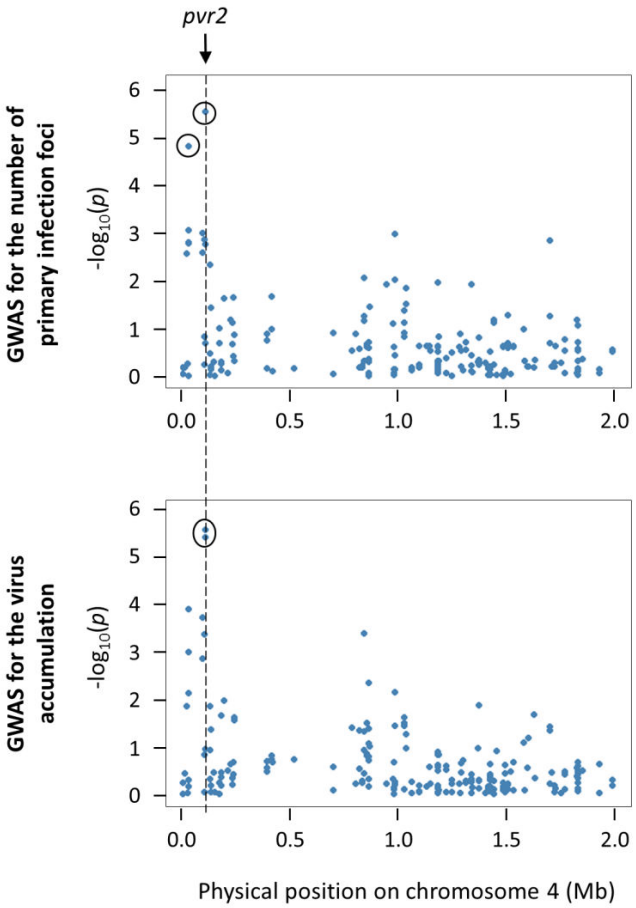


Figure 6: Comparison between SNPs mapped at the beginning of chromosome 4 (represented in a Manhattan plot displaying the $-\log_{10}(P\text{-values})$ over genomic positions) and the position of the major *pvr2* resistance gene on the CM334 reference genome (indicated with a dotted line). The SNPs showing significant association with the number of primary infection foci and/or virus accumulation are surrounded by a black circle.

3 Discussion

3.1 Benefits and limits of the pepper core-collection to perform genome-wide association

The aim of the present study was to estimate the availability and diversity of genetic factors controlling PVY population sizes in pepper resources. Our strategy was to select a core-collection composed of 276 pepper inbred lines representative of the genetic diversity of the pepper germplasm (Figure S1). The core-collection has then been phenotyped for two traits linked to PVY quantitative resistance, genotyped using GBS and association mapping has been performed.

The two traits phenotyped among the pepper accessions were the number of PVY infection foci on inoculated pepper cotyledons and the PVY accumulation in the whole plant. The number of foci is a proxy of the effective population size (N_e), while the virus accumulation (VA) is a proxy of the census population size. N_e can be defined as the number of viral particles that pass their genes to the next generation (Elena and Sanjuán 2007). It is a parameter of interest when studying plant resistance durability because it is directly related to the evolution of the viral population. Indeed, the strength of genetic drift within the plant is linked to N_e: when N_e is small, genetic drift is strong and virus variants can be lost independently of their fitness, while when N_e is large, genetic drift is weak and the most adapted variants will increase in frequency within the plant because of the action of selection. Previous studies have demonstrated that during the inoculation of virus on the leaf, a primary infection focus is induced by only one viral particle (Zwart et al. 2011; Tamisier et al. 2017). Therefore, quantifying the number of primary infectious foci on a leaf is a direct estimation of N_e at inoculation. Contrarily to N_e, VA is related to the total number of individuals in the population. Therefore, it is an accurate estimator of the quantitative resistance conferred by the plant: the lower VA, the higher the quantitative resistance. N_e at inoculation and VA were highly variable among the pepper accessions (Figure 2) and showed high broad-sense heritability (h^2=0.98 for N_e and h^2=0.80 for VA), as previously reported for these traits on a DH pepper population (Quenouille et al. 2014; Tamisier et al. 2017). Regarding these results, the phenotyping was optimal to perform GWAS on both traits.

The core-collection has been genotyped using ddRADseq (Peterson et al. 2012) and a total of 10,308 SNPs were obtained. Model-based analysis has structured the core-collection within 4 clusters closely linked to plants features (Figure 1). This observation was already reported in previous studies on *C. annuum* collections genotyped with SSR (Nicolaï et al. 2013) or GBS data (Taranto et al. 2016). However, the wild-type accessions (*C. annuum* var *glabriusculum*) were each time distributed in multiple clusters. In our analysis, we were able to cluster these accessions in one unique group, attesting of the quality of the set of SNPs and its accuracy to estimate pepper diversity. Regarding the linkage disequilibrium, we found a rapid LD decay in the 276 accessions (40 kb at r^2 = 0.2). Taranto et al. (2016) also observed a rapid decay of LD (100 kb at r^2 = 0.2) in a collection of 222 *C. annuum* accessions. Multiple factors can cause changes in LD, such as bottlenecks, selective breeding or mating system (Flint-Garcia et al. 2003). The composition of the population can also greatly influence LD. In maize, Ersoz et al. (2007) reported that LD decays within 1 kb in land races (Tenaillon et al. 2001), in

approximately 2 kb in diverse inbred lines (Remington et al. 2001) and can go up to 100 kb in commercial elite inbred lines (Ching et al. 2002). In barley, LD within 212 kb have been measured in elite lines while it decays within 0.4 kb in wild lines (Caldwell et al. 2006). A rapid LD decay can be limiting in association mapping approach because it requires a high number of markers to cover the entire genome. Nevertheless, if the marker density is high enough, it can also be advantageous because the association signal will be closer to the causal polymorphism, leading to a high mapping resolution. In our study, the rapid LD decay can also make the comparison between classical QTL mapping and GWAS difficult. Indeed, one way to compare the results of both studies is to map the QTLs on the reference genome and to see if the confidence intervals overlap between QTL and SNPs detected in association. Therefore, it requires to be able to calculate a confidence interval around the associated markers detected in GWAS. There is no widespread method for this, and the methods proposed need a high LD block around the marker in association to draw boundaries. This is usually not the case when LD is low, which generally leads to the detection of one or two significant SNPs without a high LD block around them (Cormier et al. 2014). That was the case in our study, and we have therefore simply observed if the position of the significant SNPs were close or included to the support interval of previously-detected QTLs.

To summarize, our phenotyping was optimal to perform GWAS because both traits are variable and highly heritable. The genotyping of the core-collection was accurate and described efficiently the pepper diversity. However, the rapid LD decay could be limiting to perform association mapping and compare the results with the previously-detected QTLs.

3.2 Common genetic factors control the effective population size at inoculation and the virus accumulation

The association mapping has revealed 6 different SNPs in association with N_e at inoculation and/or VA (Table 3, Figure 3a). On chromosome 4, a common SNP was associated to both traits and all significant SNPs were close to the position of the *pvr2* resistance gene on the CM334 reference genome (around 40 kb for the closest marker) (Figure 6). To select the core-collection, we have eliminated the pepper accessions known to be resistant to the pathotype PVY-0,1,2,3 in order to measure only the action of the genetic background and exclude the effect of *pvr2*-mediated resistance. However, it is likely that some accessions still carried resistance alleles at this locus which were missed during this previous phenotyping.

Indeed, around 20% of the accessions have shown no infection focus and/or were not infected with PVY according to the quantitative ELISA. Moreover, when these accessions were discarded from the core-collection, no associated SNP was found on this chromosome but the other associations were still detected (data not shown). Therefore, we can conclude that the associations detected on chromosome 4 are most likely due to the *pvr2* locus.

On chromosome 6, the same associated SNPs were identified for N_e at inoculation and VA (Table 3, Figure 5). Furthermore, the favorable allele was the same for both traits (Figure 3b). These results suggest than one or several gene(s) have a pleiotropic effect and control both N_e at inoculation and VA. Several studies have demonstrated that the number of founders entering into the plant and the virus accumulation within the plant were related. Rodrigo et al. (2014) demonstrated that diminishing N_e at inoculation induced a delay in plant infection at the systemic level and Lafforgue et al. (2012) showed that it also resulted in a lower proportion of infected cells at the systemic level. Therefore, both mechanisms probably lead to a lower VA. In our study, we provided the first proof that this relation is partly under genetic control, and that gene(s) decreasing N_e at inoculation will also decrease VA. Moreover, the SNPs lie within the support interval of VA-6, a previously detected QTL controlling the virus accumulation (Figure 5). Finally, the SNP associated with N_e at inoculation was at least 20 Mb apart from the support interval of the previously detected QTL PVY-6. Therefore, we can assume that at least two different genomic regions on chromosome 6 control N_e at inoculation.

On chromosome 12, one SNP significantly associated with N_e at inoculation was detected (Table 3, Figure 3a). This SNP was approximately 10 Mb away from the QTL PVY-12 previously detected for this trait on this chromosome (Figure 4). Two genomic regions could therefore control this trait. However, the DH population used to detect PVY-12 showed a lack of markers in the genomic region of the associated SNP (Tamisier et al. 2017). Consequently, the mapping of PVY-12 may not be precise and the associations detected with both methods could be caused by the same genomic region. Adding markers on the genetic map of the DH population could help to better interpret these results.

To finish the comparison of the results with previous studies, two QTLs identified in a biparental population were not found with the association mapping approach: the VA-3 QTL on chromosome 3 controlling virus accumulation (Quenouille et al. 2014) and the PVY-7 on pepper controlling N_e at inoculation (Tamisier et al. 2017). A limit of GWAS is its lack of power

to detect rare alleles (Brachi et al. 2011). We can assume that the frequencies of the alleles at these two QTLs were too low in the core-collection to be detected. It is probably the case for the VA-6 QTL. Indeed, Quenouille et al. (2014) have shown that VA-6 co-located tightly with the *pvr6* gene encoding eIFiso4E, an isoform of eIF4E (eukaryotic initiation factor 4E) encoded by the *pvr2* gene. In the biparental population, the *pvr6*+ allele encodes a functional eIF(iso)4E protein, while the other *pvr6* allele is a 'natural' knockout (KO) of the gene which encodes a truncated, non-functional protein (Ruffel et al. 2006). The pepper lines carrying the *pvr6*+ allele displayed low VA while the lines carrying the KO allele displayed high VA. Within a subset of 20 accessions representative of the genetic diversity of the pepper germplasm, Quenouille et al. (2015) have found that only 3 accessions carried the KO *pvr6* allele. If the KO *pvr6* allele frequency is still low at the core-collection scale, it could explain why we did not detect this gene by association mapping. Given that the LD decays rapidly, another hypothesis would be that the marker density was not high enough to detect all the genomic regions controlling the traits.

In this study, we have explored plant genetic diversity to evaluate the availability of genetic factors controlling the number of viruses entering in the plant during the inoculation step and the virus accumulation. The genomic regions detected in association with both traits were similar to the ones identified in a pepper population obtained from a biparental cross. On chromosome 6, the same genomic regions were implicated in the control of both traits. Even if a correlation between the number of founders at inoculation and the virus accumulation within the plants have already been reported for other viruses (Lafforgue et al. 2012; Rodrigo et al. 2014), it is the first proof that a common genetic factor influences these traits. Moreover, a decrease in resistance breakdown frequency has been demonstrated in a subset of 20 pepper accessions when the resistance gene was introgressed into a resistance genetic background which decreases the virus accumulation (Quenouille et al. 2015). Our results suggest that the higher level of durability observed could also be due to a smaller N_e at inoculation, and consequently a higher level of genetic drift. The use of such genetic factors in resistance breeding programs is promising and could help to achieve resistance durability by controlling pathogen evolution.

4 Materials and methods

4.1 Pepper core-collection sampling

The core-collection of *C. annuum* was built using the MStrat software v.4.1 (Gouesnard et al. 2001). We first estimated the minimum number of accessions needed to keep the same allelic richness as the whole germplasm. To do so, we performed sampling simulations for different sizes of core-collection with the Maximization (M) strategy algorithm and calculated the allelic richness for all these core-collections. We then applied the M strategy to select the final core-collection, with 20 replicates and 30 iterations per replicate. The most prevalent accessions among the 20 replicates were included in the final core-collection. The number of alleles, the Nei's unbiased gene diversity index (He) and the observed heterozygosity (Ho) of the core-collection were calculated using GenAlEx 6.5 software (Peakall and Smouse 2012).

4.2 Phenotyping of the core-collection

4.2.1 Number of PVY primary infection foci

The core-collection of 310 accessions was sown in greenhouse conditions with 20 plants per accession. The virus used for the phenotyping was derived from a cDNA clone of PVY isolate SON41p. It carried the 115K substitution in the VPg which allows it to overcome the *pvr2*^{3} resistance, belonged to pathotype (0,1,2,3) and was tagged with the GFP reporter gene. The virus was first propagated in *Nicotiana tabacum* cv. Xanthi plants in order to multiply the inoculum. Twenty-eight days after sowing, the two cotyledons of each plant were mechanically inoculated. Only one experimenter performed all the inoculations. Five days post inoculation, the number of foci showing green fluorescence was quantified under specific light wavelength (450-490 nm) for 20 cotyledons per accession. The number of primary infection foci has previously been proved to be an accurate estimation of the number of founders entering in the leaf (Zwart et al. 2011; Tamisier et al. 2017) and is therefore used as a proxy of N_e at inoculation.

4.2.2 Virus accumulation

The core-collection was sown a second time in greenhouse conditions with 10 plants per accession. The virus used was also derived from the SON41p cDNA clone and belonged to

pathotype (0,1,2,3). It carried the 119N substitution in the VPg, which allows it to overcome the *pvr2*^{3} resistance. The virus was first propagated in *Nicotiana tabacum* cv. Xanthi plants. Twenty-eight days after sowing, the two cotyledons of 10 plants per accession were mechanically inoculated with the PVY inoculum. One month post inoculation, 1 gram from 3 uninoculated apical leaves was sampled from each inoculated plant. Samples from each plant were separately ground in a phosphate buffer (0.03 M Na_2HP0_4, 0.2% sodium diethyldithiocarbamate, 4 mL buffer per gram of leaves). To reduce the number of samples and make the virus accumulation estimation achievable, the 10 samples of grounded leaves per accession were pooled into two groups of 4 samples and one group of 2 samples, reducing the experiment from 10 measures of virus accumulation per accession to 3. A control experiment had been performed previously on 30 plants to confirm that the estimation of virus accumulation was accurate when using pooled samples (Figure S4). A quantitative DAS-ELISA (Double Antibody Sandwich - Enzyme Linked Immunosorbent Assay) was performed on each pooled sample as described by Ayme et al. (2006). The mean virus concentration per DH line was assessed using serial dilutions of the pooled samples of infected plants and calculated relatively to a common control sample added to each ELISA plate.

4.3 SNP detection

The core-collection was sown a third time in greenhouse conditions. One plant per accession was used for DNA extraction. DNA was isolated from 80 mg of frozen young leaves with the DNeasy Plant Mini Kit (QIAGEN) and the extracted DNA was quantified using a Qubit fluorometer (Invitrogen). The genome complexity was reduced by ddRADseq (Peterson et al. 2012). The two restriction enzymes used were *Pst*I, a rare-cutting restriction enzyme sensitive to methylation, and *Mse*I, a common-cutting restriction enzyme. Adapters were then ligated to restriction fragments, the samples were pooled and PCR amplifications were performed. All the libraries were constructed at the Cirad facilities (Montpellier, France). Next-generation sequencing was performed in a 150-bp single-read mode using three lanes on a HiSeq3000 platform (Illumina) at Genotoul (Toulouse, France).

The FASTQ files were demultiplexed using a python script that searches for the adapters in 5' ends and for the expected restriction sites following the adapters. It also removes the adapter sequences in 5' (https://github.com/timflutre/quantgen-/blob/master/demultiplex.py). Then, adapters were removed in 3' ends and the sequence reads were filtered for quality (quality

trimming > 20 and minimum read length = 35 pb) using cutadapt. The reads were aligned against the reference genome of *C. annuum* cv. CM334 (Kim et al. 2014) using the Burrows-Wheeler Aligner tool and the algorithm BWA-MEM. Variant calling was performed using GATK haplotypecaller. A raw HapMap file was produced with TASSEL 5.2.39 (Bradbury et al. 2007) and several in-house R scripts were used to filter SNPs. Imputation of missing SNPs was performed with the Random Forest Regression Imputation Model implemented in the R package 'missForest' (Stekhoven and Bühlmann 2011).

4.4 Population structure and linkage disequilibrium estimations

Population structure was estimated using the Bayesian model-based clustering method implemented in STRUCTURE 2.3.4 software (Pritchard et al. 2000). The admixture model was used with 100,000 replicates for burn-in and Markov Chain Monte Carlo (MCMC) iteration. Five runs were performed for each value of clusters (K), ranging from 1 to 10. The optimal K value was inferred from the log probability of the data (LnP(D)) and delta K. These values were calculated with Structure Harvester (Earl 2012). A neighbour-joining tree based on pairwise genetic distances between SNPs was constructed using UPGMA procedure implemented in the DARwin 6.0 software (Perrier and Jacquemoud-Collet 2006). Linkage disequilibrium (LD) analysis was performed with TASSEL 5.2.39 (Bradbury et al. 2007) using the LD squared allele frequency correlation (r^2) estimated from pairwise comparisons between SNPs. The pairwise LD heatmap was drawn using the R package 'snp.plotter' (Luna and Nicodemus 2007).

4.5 Genome-wide association study

The GWAS analyses were performed with the compressed mixed linear model (CMLM) (Zhang et al. 2010) implemented in the Genomic Association and Prediction Integrated Tool (GAPIT) R package (Lipka et al. 2012). The CMLM accounts for the structure of the population by including principal components as fixed effects and for relatedness between accessions by including a random-effect kinship matrix (K matrix), where kinship estimates are calculated between pairs of groups. The false discovery rate (FDR) correction was used to account for multiple testing and determine a corrected significance cutoff.

4.6 Statistical analyses

Statistical analyses were performed using the R software (http://www.r-project.org/). For the two phenotypic traits (number of infection foci initiated by PVY and virus accumulation), broad-sense heritability was estimated using the formula $h^2 = \sigma^2_G / (\sigma^2_G + \sigma^2_E/n)$, where σ^2_G corresponds to the genotypic variance, σ^2_E to the environment variance and n to the number of replicates (n = 20 and 3, respectively). An ln(x+1) transformation was applied to the two traits to approximate a normal distribution.

Funding information

L. Tamisier's PhD was supported by the BAP (Biologie et Amélioration des Plantes) department and SMaCH (Sustainable Management of Crop Health) metaprogramme of INRA and by the Région Provence-Alpes-Côte d'Azur (PACA). The experimental work was supported by the SMaCH metaprogramme.

Acknowledgments

The authors would like to thank AM Sage-Palloix and G Nemouchi for providing pepper genetic resources. We also thank the greenhouse staff for support in plant experimentation, S Santoni and G Sarah for their advices in GBS and bioinformatic, and HM Clause for its support.

5 Supplementary material

Figure S1: List of the 276 *Capsicum* accessions of the core-collection and clustering of these accessions based on population structure.

PM	Name	Variety	Structure cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4
PM0001	Grosse Douce	*annuum*	1	1.00	0.00	0.00	0.00
PM0005	Dolce Di Bergamo	*annuum*	Admixed	0.44	0.43	0.14	0.00
PM0008	Giallo Dolce Di Nocera	*annuum*	1	0.76	0.24	0.00	0.00
PM0009	Rosso Di Spagna	*annuum*	1	0.61	0.36	0.04	0.00
PM0031	Yolo Wonder	*annuum*	1	1.00	0.00	0.00	0.00
PM0032	Antibois	*annuum*	1	0.79	0.17	0.01	0.03
PM0035	Gros Carre Jaune	*annuum*	1	0.88	0.05	0.03	0.04
PM0037	Piment Sucette	*annuum*	1	0.57	0.32	0.11	0.00
PM0038	Cerise	*annuum*	Admixed	0.29	0.49	0.21	0.01
PM0044	Piment De Cayenne Érigé	*annuum*	Admixed	0.31	0.44	0.20	0.05
PM0057	Doux D'Alger	*annuum*	1	0.68	0.27	0.05	0.00
PM0065	Doux D'Espagne	*annuum*	1	0.56	0.36	0.08	0.00
PM0076	Doux Long Des Landes	*annuum*	Admixed	0.37	0.46	0.17	0.00
PM0093	Gros Carre Doux	*annuum*	1	0.97	0.03	0.00	0.00
PM0135	Calwonder	*annuum*	1	1.00	0.00	0.00	0.00
PM0139	Oalaview Wonder	*annuum*	1	1.00	0.00	0.00	0.00
PM0141	World Beater	*annuum*	1	0.84	0.16	0.00	0.00
PM0145	Plat Rouge	*annuum*	1	0.87	0.07	0.06	0.01
PM0147	Long Jaune	*annuum*	1	0.77	0.17	0.05	0.01
PM0148	Niora 1	*annuum*	1	0.51	0.44	0.05	0.00
PM0159	Vinedale	*annuum*	1	0.78	0.20	0.01	0.00
PM0162	Yolo Y = Yrp10	*annuum*	1	1.00	0.00	0.00	0.00
PM0174	Roley	*annuum*	3	0.06	0.18	0.75	0.01
PM0216	Malgache	*annuum*	3	0.00	0.07	0.93	0.00
PM0217	Piment 493-1 Pi201234	*annuum*	Admixed	0.24	0.20	0.42	0.13
PM0224	Peperone Calozi	*annuum*	1	0.87	0.13	0.00	0.00
PM0226	Pimiento Bolla	*annuum*	1	0.54	0.38	0.07	0.01
PM0236	Cerutti	*annuum*	1	0.52	0.34	0.08	0.05
PM0251	Szegdi Aprocseresznije	*annuum*	2	0.21	0.60	0.17	0.02
PM0278	Largo De Reus	*annuum*	1	0.72	0.28	0.00	0.00
PM0280	Morro De Vaca	*annuum*	1	0.90	0.10	0.00	0.00
PM0285	Tomaten-Fruchtiger	*annuum*	1	0.86	0.13	0.00	0.01
PM0298	Zitavska	*annuum*	2	0.00	0.91	0.06	0.03
PM0308	Szegedi Nagycseresznye	*annuum*	2	0.00	0.93	0.04	0.03

Figure S1: Continued

PM	Name	Variety	Structure cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4
PM0318	Kubanskij Rannij 70/60	*annuum*	1	0.51	0.49	0.00	0.00
PM0321	Astrachanskij A60	*annuum*	2	0.00	0.92	0.06	0.03
PM0322	Velikan Nr-97	*annuum*	2	0.00	0.97	0.00	0.03
PM0323	Maikop 470	*annuum*	2	0.42	0.58	0.00	0.00
PM0324	Yellow Oshkosh	*annuum*	1	0.72	0.22	0.04	0.02
PM0329	Janos	*annuum*	2	0.39	0.61	0.00	0.00
PM0333	Wiener Calvill	*annuum*	2	0.16	0.84	0.00	0.00
PM0336	Yellow Naharia	*annuum*	1	0.50	0.50	0.00	0.00
PM0338	Giallo Longo Dolce Di Marconi	*annuum*	1	0.65	0.33	0.02	0.00
PM0342	De Arad	*annuum*	2	0.00	0.98	0.01	0.01
PM0343	Collectivist 1962	*annuum*	1	1.00	0.00	0.00	0.00
PM0360	Petit Fruit Conique	*annuum*	3	0.05	0.00	0.94	0.00
PM0363	H 28-17-60	*annuum*	1	0.74	0.24	0.03	0.00
PM0367	Tendre De Chateaurenard	*annuum*	1	0.79	0.21	0.00	0.00
PM0384	Doux De Valence	*annuum*	1	0.76	0.22	0.00	0.02
PM0396	Gros Carre De Robion	*annuum*	1	0.75	0.25	0.00	0.00
PM0398	Fresno Chili	*annuum*	1	0.59	0.23	0.16	0.02
PM0416	Saltenito	*annuum*	1	0.79	0.20	0.00	0.01
PM0424	Kupos	*annuum*	2	0.19	0.80	0.00	0.01
PM0425	Fertodi	*annuum*	2	0.49	0.51	0.00	0.00
PM0434	66 M 7-1	*annuum*	Admixed	0.44	0.48	0.06	0.01
PM0439	Liverpool	*annuum*	2	0.00	0.84	0.16	0.00
PM0448	Vinette	*annuum*	1	0.96	0.04	0.00	0.00
PM0453	Sivrija	*annuum*	2	0.23	0.77	0.00	0.00
PM0460	Bela Kapija 1	*annuum*	2	0.14	0.86	0.00	0.00
PM0478	Dzuljunska Sipka	*annuum*	2	0.00	0.94	0.05	0.02
PM0481	Kalinkov	*annuum*	1	0.51	0.49	0.00	0.00
PM0492	39/A	*annuum*	1	0.64	0.25	0.10	0.00
PM0499	Capsicum Grossum	*annuum*	2	0.22	0.67	0.08	0.03
PM0534	Petit Rond	*annuum*	3	0.02	0.00	0.98	0.00
PM0537	Chillies Chinise Giant	*annuum*	1	1.00	0.00	0.00	0.00
PM0573	Csokros Felallo	*annuum*	2	0.12	0.88	0.00	0.00
PM0574	Chad Assif Adom	*annuum*	1	0.99	0.00	0.00	0.01
PM0579	Ujmajori	*annuum*	2	0.11	0.89	0.00	0.00
PM0582	Decoratif Hongrie 2	*annuum*	2	0.00	0.65	0.34	0.01

Figure S1: Continued

PM	Name	Variety	Structure cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4
PM0588	Ikeda	*annuum*	1	0.85	0.09	0.03	0.03
PM0590	Puerto Rico Wonder	*annuum*	1	1.00	0.00	0.00	0.00
PM0613	Pasilla 73- Chi-15201	*annuum*	2	0.05	0.71	0.13	0.11
PM0623	Truhart Linea 27 A	*annuum*	1	0.90	0.09	0.01	0.00
PM0625	Bouquet 3	*annuum*	2	0.06	0.94	0.00	0.00
PM0630	Pimex 2	*annuum*	Admixed	0.15	0.00	0.36	0.49
PM0639	Chiapas Br 203 70	*glabriusculum*	4	0.00	0.03	0.02	0.96
PM0641	Turrialba (C81)	*glabriusculum*	4	0.00	0.00	0.00	1.00
PM0644	Antigua (390)	*glabriusculum*	4	0.04	0.00	0.00	0.96
PM0646	Rama	*glabriusculum*	4	0.00	0.00	0.00	1.00
PM0648	Turtle Mound (4675)	*glabriusculum*	4	0.00	0.00	0.05	0.95
PM0651	Cs-Zh Gepi	*annuum*	2	0.44	0.56	0.00	0.00
PM0653	D.Cecei	*annuum*	2	0.10	0.90	0.00	0.00
PM0654	Javitott Cecei	*annuum*	2	0.08	0.92	0.00	0.00
PM0655	Tochigisantaka	*annuum*	3	0.00	0.00	1.00	0.00
PM0659	Perennial	*annuum*	3	0.01	0.01	0.92	0.06
PM0662	S20-1 Singh4	*annuum*	3	0.07	0.00	0.91	0.02
PM0663	Chihuahua (504 Bp)	*glabriusculum*	4	0.00	0.07	0.00	0.93
PM0665	Nuevo Leon C 70- 13	*glabriusculum*	4	0.00	0.00	0.23	0.77
PM0668	Tonosi C 70- 1	*glabriusculum*	4	0.00	0.00	0.00	1.00
PM0669	Panama City	*glabriusculum*	4	0.00	0.00	0.00	1.00
PM0670	Pitalito	*glabriusculum*	4	0.04	0.00	0.00	0.96
PM0671	Piment De Szentes	*annuum*	2	0.19	0.81	0.00	0.00
PM0675	Moga	*annuum*	1	0.99	0.01	0.00	0.00
PM0676	Bak Louti	*annuum*	1	0.62	0.25	0.13	0.00
PM0679	N° 77-80-3	*annuum*	3	0.01	0.00	0.93	0.06
PM0681	N° 77-81-4	*annuum*	3	0.39	0.00	0.61	0.00
PM0684	Y 1	*annuum*	1	1.00	0.00	0.00	0.00
PM0690	Gabes Iv P 5B	*annuum*	1	0.55	0.24	0.21	0.00
PM0693	Hatvani	*annuum*	2	0.00	0.93	0.07	0.00
PM0697	Soliman 2	*annuum*	1	0.69	0.22	0.09	0.00
PM0702	Criollo De Morelos 334	*annuum*	Admixed	0.19	0.24	0.44	0.13
PM0707	D. Cecei Li / Li	*annuum*	2	0.15	0.85	0.00	0.00
PM0708	Feherozon Li / Li	*annuum*	2	0.34	0.66	0.00	0.00
PM0713	X P H 832 (F1)	*annuum*	1	1.00	0.00	0.00	0.00

Figure S1: Continued

PM	Name	Variety	Structure cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4
PM0721	Morales Prv121 'Mora'	*annuum*	3	0.15	0.00	0.65	0.20
PM0723	D'Hirat P3 -P15	*annuum*	Admixed	0.33	0.48	0.19	0.00
PM0731	Golden Bell	*annuum*	1	0.91	0.08	0.00	0.02
PM0743	Ata 100	*annuum*	2	0.03	0.97	0.00	0.00
PM0744	Cetinel 150	*annuum*	2	0.00	1.00	0.00	0.00
PM0748	Eskisehir Carliston	*annuum*	2	0.07	0.93	0.00	0.00
PM0752	C 75 - 3	*annuum*	Admixed	0.30	0.17	0.39	0.14
PM0764	Export	*annuum*	1	0.68	0.32	0.00	0.00
PM0765	Aroma	*annuum*	1	0.73	0.27	0.00	0.00
PM0766	Titan	*annuum*	2	0.06	0.94	0.00	0.00
PM0768	Spendid	*annuum*	1	0.66	0.34	0.00	0.00
PM0770	Da -Tong Early	*annuum*	1	0.91	0.08	0.01	0.00
PM0771	Yong Jiu 3	*annuum*	1	0.97	0.03	0.00	0.00
PM0775	Ben Xi	*annuum*	1	0.71	0.09	0.20	0.00
PM0776	Nanjing Early Pepper	*annuum*	3	0.00	0.19	0.75	0.06
PM0800	Florida Vr4	*annuum*	1	0.94	0.05	0.01	0.00
PM0805	P 44 Marako Alaba	*annuum*	2	0.06	0.69	0.24	0.01
PM0809	Early Spring Green	*annuum*	3	0.13	0.36	0.51	0.00
PM0810	Big Star (13 - 02)	*annuum*	3	0.16	0.33	0.52	0.00
PM0811	83 - 130	*annuum*	2	0.00	0.69	0.26	0.05
PM0818	Tong Feng 16	*annuum*	1	1.00	0.00	0.00	0.00
PM0819	Chao Tian Jiav	*annuum*	3	0.00	0.00	1.00	0.00
PM0822	Klenot	*annuum*	2	0.47	0.52	0.00	0.00
PM0833	Ju Sp 1980 - 08	*annuum*	2	0.22	0.65	0.13	0.00
PM0838	Pi 323 314	*annuum*	3	0.15	0.20	0.65	0.00
PM0841	Shinpyung 1	*annuum*	3	0.00	0.17	0.83	0.00
PM0850	Deco Vert Orange	*annuum*	3	0.06	0.19	0.75	0.00
PM0852	Shuang Feng (F1)	*annuum*	1	0.96	0.00	0.04	0.00
PM0853	Zhong Jiao 2	*annuum*	1	0.88	0.06	0.06	0.00
PM0862	Capsicum Annuum 83 - 175	*annuum*	1	0.93	0.00	0.07	0.00
PM0863	Capsicum Annuum 83 - 181	*annuum*	1	0.76	0.21	0.02	0.00
PM0870	Hunan Pepper	*annuum*	3	0.02	0.15	0.72	0.10
PM0877	V 178	*annuum*	2	0.49	0.51	0.00	0.00
PM0879	Feherozon Synthétic	*annuum*	2	0.46	0.54	0.00	0.00
PM0883	Autona	*annuum*	1	1.00	0.00	0.00	0.00

Figure S1: Continued

PM	Name	Variety	Structure cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4
PM0889	Haricot Rouge	*annuum*	2	0.00	1.00	0.00	0.00
PM0894	Bajio 1020	*annuum*	2	0.22	0.54	0.23	0.00
PM0895	Rpc 1 I	*annuum*	1	0.85	0.05	0.09	0.01
PM0897	P 1000	*annuum*	1	1.00	0.00	0.00	0.00
PM0899	Permagreen	*annuum*	1	1.00	0.00	0.00	0.00
PM0900	Rubin	*annuum*	1	0.58	0.42	0.00	0.00
PM0901	Jubilantka	*annuum*	1	0.81	0.19	0.00	0.00
PM0910	Tatli Sivri	*annuum*	2	0.04	0.96	0.00	0.00
PM0912	Aci Sivri	*annuum*	2	0.00	1.00	0.00	0.00
PM0913	Kandil A	*annuum*	2	0.49	0.51	0.00	0.00
PM0917	Ariane	*annuum*	1	1.00	0.00	0.00	0.00
PM0918	Poivron De Cuneo	*annuum*	1	0.90	0.10	0.00	0.00
PM0919	(F1) Shuang Feng	*annuum*	1	0.90	0.06	0.04	0.00
PM0921	Qie Men	*annuum*	1	1.00	0.00	0.00	0.00
PM0930	Fleurs Violettes	*annuum*	3	0.04	0.06	0.89	0.00
PM0931	Red Chile	*annuum*	2	0.00	0.69	0.28	0.04
PM0940	Zefir	*annuum*	1	0.87	0.13	0.00	0.00
PM0944	Stano	*annuum*	2	0.49	0.51	0.00	0.01
PM0947	187314	*annuum*	Admixed	0.25	0.46	0.29	0.00
PM0953	224433	*annuum*	Admixed	0.15	0.00	0.38	0.47
PM0955	267734	*annuum*	Admixed	0.39	0.18	0.37	0.06
PM0956	267736	*annuum*	Admixed	0.39	0.35	0.25	0.02
PM0960	F1 Zhongjiao N°4	*annuum*	1	1.00	0.00	0.00	0.00
PM0973	Carre D'Asti Jaune	*annuum*	1	0.95	0.05	0.00	0.00
PM0974	Laichi-2	*annuum*	3	0.06	0.03	0.91	0.00
PM0976	Indonesian Sel.	*annuum*	3	0.00	0.00	0.96	0.04
PM0977	L.25	*annuum*	2	0.08	0.92	0.00	0.00
PM0983	Peza	*annuum*	2	0.15	0.85	0.00	0.00
PM0984	Szentesi Piacos	*annuum*	2	0.48	0.52	0.00	0.00
PM0990	F1 14-10	*annuum*	2	0.45	0.55	0.00	0.00
PM0991	Cuba 1	*annuum*	1	0.63	0.29	0.08	0.00
PM0993	Cuba 3	*annuum*	2	0.26	0.74	0.00	0.00
PM0998	H6	*annuum*	Admixed	0.46	0.39	0.15	0.00
PM1011	E5/3	*annuum*	1	0.58	0.22	0.20	0.00
PM1013	Flora	*annuum*	1	0.67	0.24	0.08	0.00

Figure S1: Continued

PM	Name	Variety	Structure cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4
PM1015	Bousso 3	*annuum*	1	0.62	0.00	0.33	0.05
PM1018	Bousso 6	*annuum*	3	0.20	0.20	0.58	0.02
PM1019	Bousso 7	*annuum*	3	0.19	0.23	0.54	0.05
PM1024	Cacho De Cabra	*annuum*	1	0.50	0.43	0.07	0.00
PM1051	Golden Calwonder California	*annuum*	1	0.94	0.02	0.04	0.00
PM1055	Karlo	*annuum*	2	0.26	0.74	0.00	0.00
PM1080	Numex Sweet	*annuum*	2	0.04	0.75	0.20	0.00
PM1081	Numex Joe E.Parker	*annuum*	2	0.14	0.68	0.17	0.00
PM1083	Cajun Za-Peto Increase	*annuum*	2	0.17	0.51	0.23	0.09
PM1089	Katmandou	*annuum*	3	0.11	0.18	0.71	0.00
PM1090	Bagdaon	*annuum*	3	0.19	0.15	0.63	0.03
PM1094	Chile Dulce (Merida)	*annuum*	Admixed	0.25	0.25	0.38	0.12
PM1107	Kahramanmaras	*annuum*	2	0.32	0.50	0.19	0.00
PM1115	118 I-1-5-8-6	*annuum*	3	0.03	0.31	0.63	0.03
PM1116	135B	*annuum*	2	0.16	0.59	0.20	0.04
PM1118	135F	*annuum*	3	0.03	0.36	0.62	0.00
PM1121	171D	*annuum*	3	0.02	0.37	0.61	0.00
PM1124	2174B	*annuum*	3	0.08	0.36	0.55	0.00
PM1139	3872	*annuum*	1	0.91	0.00	0.09	0.00
PM1140	3898	*annuum*	1	0.89	0.00	0.11	0.00
PM1141	3899	*annuum*	1	0.86	0.00	0.14	0.00
PM1144	2060-10	*annuum*	1	0.97	0.00	0.02	0.01
PM1148	Et1147/792	*annuum*	1	0.97	0.02	0.00	0.01
PM1149	Et1147/795	*annuum*	1	0.63	0.37	0.00	0.00
PM1154	Tl299/347	*annuum*	1	0.66	0.30	0.04	0.00
PM1164	Donor	*annuum*	1	0.76	0.24	0.00	0.00
PM1166	Early California Wonder	*annuum*	1	1.00	0.00	0.00	0.00
PM1182	Negev	*annuum*	3	0.09	0.08	0.82	0.01
PM1196	Topepo	*annuum*	1	0.75	0.15	0.09	0.01
PM1211	Pi197409	*annuum*	Admixed	0.21	0.10	0.41	0.27
PM1221	D'Jbel Sarhro	*annuum*	Admixed	0.39	0.44	0.17	0.00
PM1223	Shalhevet	*annuum*	Admixed	0.36	0.35	0.22	0.07
PM1233	Anais 17-12-3-2	*annuum*	1	1.00	0.00	0.00	0.00
PM1237	Piment De Thaïlande	*annuum*	3	0.00	0.00	1.00	0.00
PM1241	Jaïpur - Pendjab	*annuum*	3	0.08	0.20	0.71	0.00

Figure S1: Continued

PM	Name	Variety	Structure cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4
PM1246	Napfeny	*annuum*	2	0.00	0.93	0.04	0.03
PM1249	Viktoria	*annuum*	2	0.23	0.51	0.27	0.00
PM1253	Negral	*annuum*	Admixed	0.47	0.42	0.11	0.00
PM1274	Ex Pi 273 429	*annuum*	Admixed	0.21	0.06	0.27	0.46
PM1275	Ex Pi 439 413-1 Ex Cha13	*annuum*	3	0.11	0.13	0.61	0.14
PM1337	G 3	*annuum*	3	0.06	0.03	0.89	0.02
PM1340	U1042/778	*annuum*	1	0.54	0.46	0.00	0.00
PM1345	K90	*annuum*	2	0.02	0.94	0.04	0.01
PM1356	Szededi 179	*annuum*	2	0.00	0.88	0.08	0.04
PM1361	Pyc22	*annuum*	Admixed	0.49	0.03	0.39	0.08
PM1371	Pa427	*annuum*	Admixed	0.38	0.08	0.35	0.19
PM1374	L 491 A	*annuum*	2	0.44	0.56	0.00	0.00
PM1399	Espanol 2	*annuum*	1	0.65	0.12	0.18	0.05
PM1400	P 709	*annuum*	3	0.06	0.07	0.70	0.17
PM1401	P 1514	*annuum*	2	0.00	0.71	0.29	0.00
PM1408	Milord	*annuum*	1	1.00	0.00	0.00	0.00
PM1412	Bastidon	*annuum*	1	1.00	0.00	0.00	0.00
PM1415	Chilitecpin	*glabriusculum*	4	0.00	0.06	0.00	0.94
PM1417	Himalaya	*annuum*	3	0.11	0.01	0.86	0.02
PM1423	Piment Soudan	*annuum*	3	0.00	0.01	0.94	0.04
PM1426	Yantaï 985	*annuum*	1	1.00	0.00	0.00	0.00
PM1429	Ms 77013 B (Restaurateur)	*annuum*	1	0.86	0.00	0.14	0.00
PM1430	Pikuti	*annuum*	3	0.01	0.24	0.72	0.03
PM1432	Pa 94 - 1416	*annuum*	1	1.00	0.00	0.00	0.00
PM1434	Pa 94 - 2014	*annuum*	Admixed	0.30	0.44	0.26	0.00
PM1442	Mexico 3	*annuum*	Admixed	0.21	0.37	0.42	0.00
PM1452	Dulce Italiano	*annuum*	1	0.68	0.23	0.08	0.01
PM1456	Clara	*annuum*	1	0.89	0.04	0.06	0.00
PM1459	7 Roc	*annuum*	3	0.05	0.21	0.65	0.10
PM1477	Xiang Jan 10	*annuum*	3	0.09	0.21	0.70	0.01
PM1482	Max 3	*annuum*	1	0.98	0.00	0.02	0.00
PM1483	Dalat	*annuum*	3	0.15	0.03	0.76	0.06
PM1498	L 491 A=Jaune	*annuum*	1	0.61	0.39	0.00	0.00
PM1501	Chalockchari	*annuum*	2	0.43	0.57	0.00	0.00
PM1504	Saint Nicolas	*annuum*	2	0.42	0.58	0.00	0.00

Figure S1: Continued

PM	Name	Variety	Structure cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4
PM1509	Xi Liu De Shandong	*annuum*	2	0.05	0.94	0.00	0.00
PM1526	Erectus	*annuum*	2	0.38	0.62	0.00	0.00
PM1535	Piment d'Espelette C	*annuum*	Admixed	0.34	0.46	0.20	0.00
PM1537	Yangoon 1	*annuum*	3	0.00	0.00	0.91	0.08
PM1539	Dora Jaune (F1)	*annuum*	1	0.64	0.35	0.00	0.00
PM1540	Mendoza	*annuum*	3	0.26	0.04	0.52	0.18
PM1541	Bell Boy	*annuum*	1	1.00	0.00	0.00	0.00
PM1546	Arceaux	*annuum*	3	0.14	0.13	0.74	0.00
PM1552	Roloubok	*annuum*	2	0.09	0.91	0.00	0.00
PM1566	Muraille 2	*annuum*	3	0.00	0.23	0.71	0.06
PM1573	Wad Medani	*annuum*	3	0.00	0.00	1.00	0.00
PM1580	Ca 8	*annuum*	3	0.20	0.00	0.76	0.04
PM1583	Orias Kossarou	*annuum*	2	0.15	0.67	0.19	0.00
PM1588	Tulum	*annuum*	Admixed	0.24	0.30	0.47	0.00
PM1602	Ise	*annuum*	3	0.01	0.22	0.78	0.00
PM1604	Ishii Midori	*annuum*	3	0.14	0.14	0.72	0.00
PM1606	Tosa Shishitoh	*annuum*	3	0.20	0.18	0.62	0.00
PM1609	Fushimi Amanaga	*annuum*	3	0.06	0.29	0.65	0.00
PM1612	Jaipur	*annuum*	3	0.09	0.09	0.82	0.00
PM1631	Alicante	*annuum*	1	0.55	0.33	0.12	0.00
PM1636	Nantes 6	*annuum*	1	0.88	0.04	0.08	0.00
PM1637	All Big Jaune	*annuum*	1	0.97	0.00	0.03	0.00
PM1646	Pritty In Purple Hot Ornemental	*annuum*	3	0.28	0.01	0.64	0.07
PM1654	Tunis 1	*annuum*	2	0.00	0.80	0.20	0.00
PM1655	Tunis 2	*annuum*	2	0.21	0.71	0.08	0.00
PM1668	San Sepulcro	*annuum*	3	0.22	0.19	0.59	0.00
PM1670	Cherry Bomb	*annuum*	1	0.69	0.31	0.00	0.00
PM1671	Elena 1	*annuum*	Admixed	0.47	0.48	0.05	0.00
PM1678	Padron	*annuum*	2	0.34	0.52	0.14	0.00
PM1684	Santos Flamme	*annuum*	3	0.20	0.13	0.67	0.00
PM1685	Purple Flash	*annuum*	3	0.01	0.15	0.84	0.00
PM1687	Andratx	*annuum*	Admixed	0.20	0.37	0.43	0.00

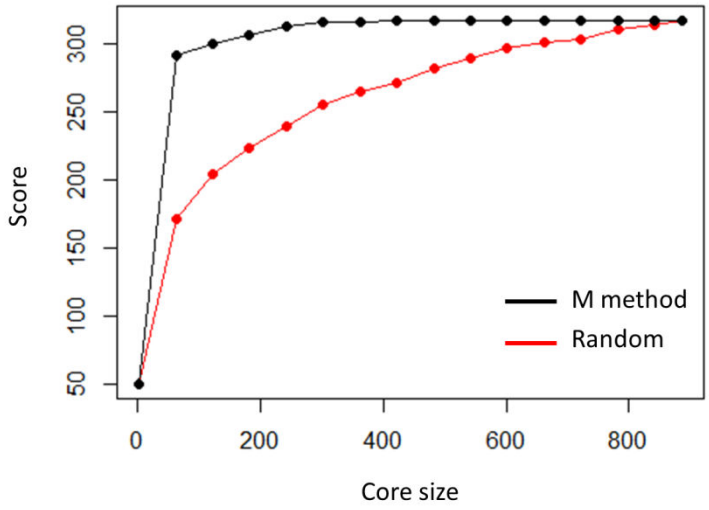


Figure S1: Allelic richness (score) capture for different core collection sizes. A random (red) and a maximization (black) sampling strategies are represented.

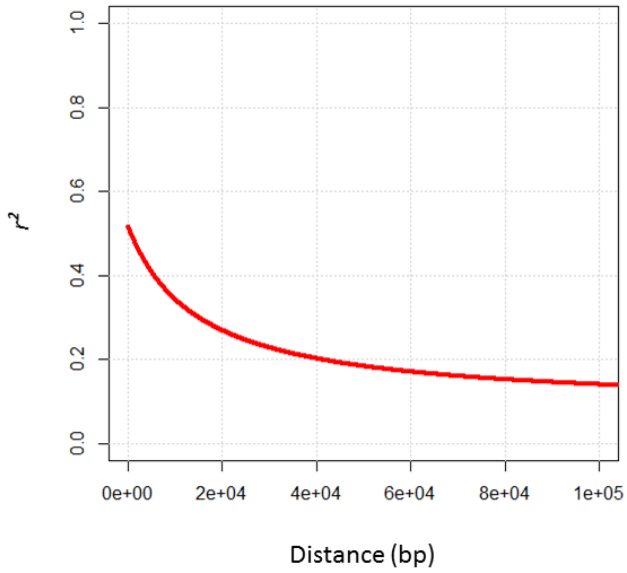


Figure S2: Average linkage disequilibrium decay (r^2) against the genetic distance (bp) throughout the CM334 pepper reference genome.

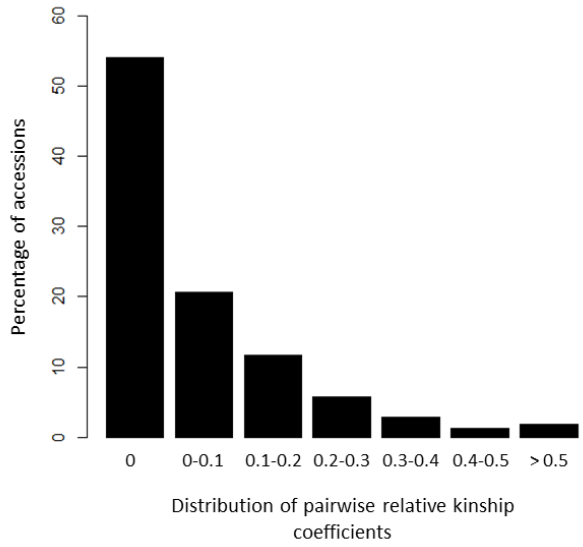


Figure S3: Distribution of pairwise kinship estimates across the core-collection.

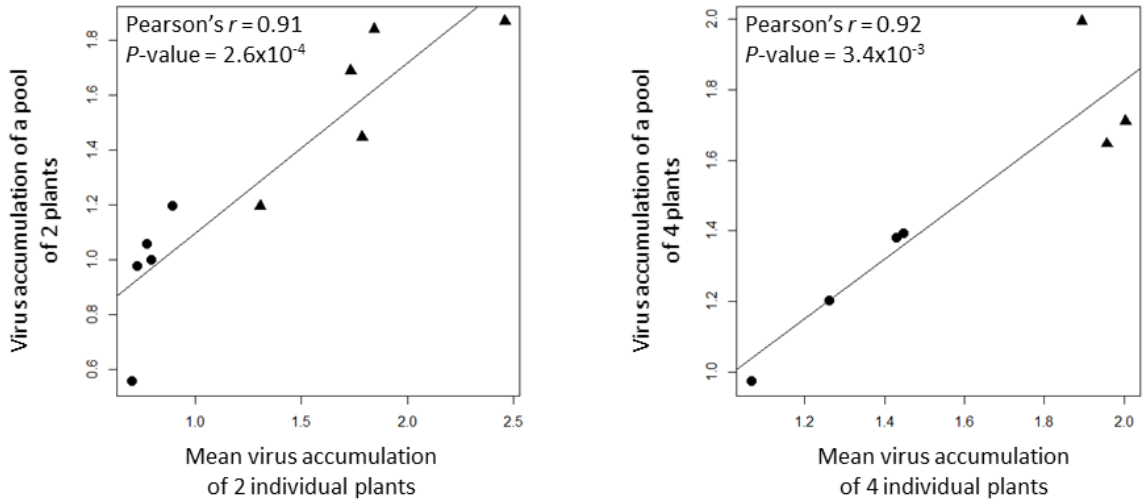


Figure S4: Control experiment for the accuracy of the virus accumulation measured with pooled samples. Two pepper accessions with fifteen plants per accession have been inoculated with a cDNA clone of PVY isolate SON41p. Apical leaves of each plant have been grounded in a phosphate buffer (0.03 M Na_2HP0_4, 0.2% sodium diethyldithiocarbamate, 4 mL buffer per gram of leaves) and the virus accumulation was assessed with a quantitative DAS-ELISA for each plant. Samples of grounded leaves have then been pooled. For each accession, 1.8 mL of grounded leaves belonging to 2 different plants and 0.9 mL of grounded leaves belonging to 4 different plants have been pooled together. Once again, a DAS-ELISA was performed to quantify the virus accumulation of the pooled samples. The relative virus accumulation obtained for the individual plants and the pool of (a) 2 and (b) 4 plants was compared for the two pepper accessions (illustrated by circles and triangles). Significant positive correlations were found for both types of pools.

References

Ayme, V., Souche, S., Caranta, C., Jacquemond, M., Chadœuf, J., Palloix, A. and Moury, B. (2006). Different mutations in the genome-linked protein VPg of *Potato virus Y* confer virulence on the *pvr2*^{3} resistance in pepper. *Mol. Plant. Microbe Interact.* **19**, 557–563.

Brachi, B., Morris, G. P. and Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* **12**, 232.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635.

Brun, H., Chèvre, A.-M., Fitt, B. D., Powers, S., Besnard, A.-L., Ermel, M., Huteau, V., Marquer, B., Eber, F., Renard, M., et al. (2010). Quantitative resistance increases the durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. *New Phytol.* **185**, 285–299.

Caldwell, K. S., Russell, J., Langridge, P. and Powell, W. (2006). Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics* **172**, 557–567.

Ching, A., Caldwell, K. S., Jung, M., Dolan, M., Smith, O., Tingey, S., Morgante, M. and Rafalski, A. J. (2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet.* **3**, 19.

Cormier, F., Le Gouis, J., Dubreuil, P., Lafarge, S. and Praud, S. (2014). A genome-wide identification of chromosomal regions determining nitrogen use efficiency components in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **127**, 2679–2693.

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361.

Elena, S. F. and Sanjuán, R. (2007). Virus Evolution: Insights from an Experimental Approach. *Annu. Rev. Ecol. Evol. Syst.* **38**, 27–52.

Ersoz, E. S., Yu, J. and Buckler, E. S. (2007). Applications of linkage disequilibrium and association mapping in crop plants. *Genomics-Assist. Crop Improv.* 97–119.

Evanno, G., Regnaut, S. and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620.

Flint-Garcia, S. A., Thornsberry, J. M. and Buckler IV, E. S. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* **54**, 357–374.

Fournet, S., Kerlan, M. C., Renault, L., Dantec, J. P., Rouaux, C. and Montarry, J. (2013). Selection of nematodes by resistant plants has implications for local adaptation and cross-virulence: Local adaptation and cross-virulence in *Globodera pallida*. *Plant Pathol.* **62**, 184–193.

García-Arenal, F. and McDonald, B. A. (2003). An analysis of the durability of resistance to plant viruses. *Phytopathology* **93**, 941–952.

Gouesnard, B., Bataillon, T. M., Decoux, G., Rozale, C., Schoen, D. J. and David, J. L. (2001). MSTRAT: An Algorithm for Building Germ Plasm Core Collections by Maximizing Allelic or Phenotypic Richness. *J. Hered.* **92**, 93–94.

Hamblin, M. T., Buckler, E. S. and Jannink, J.-L. (2011). Population genetics of genomics-based crop improvement methods. *Trends Genet.* **27**, 98–106.

Kim, S., Park, M., Yeom, S.-I., Kim, Y.-M., Lee, J. M., Lee, H.-A., Seo, E., Choi, J., Cheong, K., Kim, K.-T., et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat Genet* **46**, 270–278.

Korte, A. and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* **9**, 29.

Lafforgue, G., Tromas, N., Elena, S. F. and Zwart, M. P. (2012). Dynamics of the establishment of systemic potyvirus infection: independent yet cumulative action of primary infection sites. *J. Virol.* **86**, 12912–12922.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S. and Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399.

Luna, A. and Nicodemus, K. K. (2007). snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. *Bioinformatics* **23**, 774–776.

McDonald, B. A. and Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* **40**, 349–379.

Mundt, C. (2002). Use of multiline cultivars and cultivar mixtures for disease management. *Annu. Rev. Phytopathol.* **40**, 381–410.

Nicolaï, M., Pisani, C., Bouchet, J.-P., Vuylsteke, M. and Palloix, A. (2012). Short Communication Discovery of a large set of SNP and SSR genetic markers by high-throughput sequencing of pepper (*Capsicum annuum*). *Genet. Mol. Res.* **11**, 2295–2300.

Nicolaï, M., Cantet, M., Lefebvre, V., Sage-Palloix, A.-M. and Palloix, A. (2013). Genotyping a large collection of pepper (*Capsicum* spp.) with SSR loci brings new evidence for the wild origin of cultivated *C. annuum* and the structuring of genetic diversity by human selection of cultivar types. *Genet. Resour. Crop Evol.* **60**, 2375–2390.

Pagny, G., Paulstephenraj, P. S., Poque, S., Sicard, O., Cosson, P., Eyquard, J., Caballero, M., Chague, A., Gourdon, G. and Negrel, L. (2012). Family-based linkage and association mapping reveals novel genes affecting *Plum pox virus* infection in *Arabidopsis thaliana*. *New Phytol.* **196**, 873–886.

Palloix, A., Ayme, V. and Moury, B. (2009). Durability of plant major resistance genes to pathogens depends on the genetic background, experimental evidence and consequences for breeding strategies. *New Phytol.* **183**, 190–199.

Peakall, R. and Smouse, P. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28, 2537e2539.

Perrier, X. and Jacquemoud-Collet, J. (2006). DARwin (http://darwin.cirad.fr/darwin).

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One* **7**, e37135.

Pink, D. A. (2002). Strategies using genes for non-durable disease resistance. *Euphytica* **124**, 227–236.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

Quenouille, J., Paulhiac, E., Moury, B. and Palloix, A. (2014). Quantitative trait loci from the host genetic background modulate the durability of a resistance gene: a rational basis for sustainable resistance breeding in plants. *Heredity* **112**, 579–587.

Quenouille, J., Saint-Felix, L., Moury, B. and Palloix, A. (2015). Diversity of genetic backgrounds modulating the durability of a major resistance gene. Analysis of a core collection of pepper landraces resistant to *Potato virus Y*. *Mol. Plant Pathol.* **17**, 296–302.

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. M. and Buckler, E. S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA.* **98**, 11479–11484.

Rodrigo, G., Zwart, M. P. and Elena, S. F. (2014). Onset of virus systemic infection in plants is determined by speed of cell-to-cell movement and number of primary infection foci. *J. R. Soc. Interface* **11**, 20140555.

Ruffel, S., Gallois, J.-L., Moury, B., Robaglia, C., Palloix, A. and Caranta, C. (2006). Simultaneous mutations in translation initiation factors eIF4E and eIF (iso) 4E are required to prevent *Pepper veinal mottle virus* infection of pepper. *J. Gen. Virol.* **87**, 2089–2098.

Sage-Palloix, A.-M., Jourdan, F., Phaly, T., Némouchi, G., Lefebvre, V. and Palloix, A. (2007). Analysis of diversity in pepper genetic resources: distribution of horticultural and resistance traits in the INRA pepper germplasm. In *Progress in research on Capsicum & Eggplant*, pp. 32–42. Warsaw (Pologne).

Sonah, H., O'Donoughue, L., Cober, E., Rajcan, I. and Belzile, F. (2015). Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol. J.* **13**, 211–221.

Stekhoven, D. J. and Bühlmann, P. (2011). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112–118.

Tamisier, L., Rousseau, E., Barraillé, S., Nemouchi, G., Szadkowski, M., Mailleret, L., Grognard, F., Fabre, F., Moury, B. and Palloix, A. (2017). Quantitative trait loci in pepper control the effective population size of two RNA viruses at inoculation. *J. Gen. Virol.* **98**, 1923–1931.

Taranto, F., D'Agostino, N., Greco, B., Cardi, T. and Tripodi, P. (2016). Genome-wide SNP discovery and population structure analysis in pepper (*Capsicum annuum*) using genotyping by sequencing. *BMC Genomics* **17**, 943.

Tenaillon, O. (2014). The utility of Fisher's geometric model in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* **45**, 179–201.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423.

Zhan, J., Thrall, P. H., Papaïx, J., Xie, L. and Burdon, J. J. (2015). Playing on a Pathogen's Weakness: Using Evolution to Guide Sustainable Plant Disease Control Strategies. *Annu. Rev. Phytopathol.* **53**, 19–43.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K. and Ordovas, J. M. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360.

Zwart, M. P., Daròs, J.-A. and Elena, S. F. (2011). One Is Enough: In Vivo Effective Population Size Is Dose-Dependent for a Plant RNA Virus. *PLoS Pathog.* **7**, e1002122.

Chapitre 3

Impact of genetic drift, selection and within-host accumulation on virus adaptation to its host plants

Résumé de l'article :

Objectifs : L'objectif principal était de décrire les mécanismes mis en place par le fonds génétique de la plante pour orienter l'évolution des populations virales et protéger le gène majeur du contournement. L'impact de plusieurs forces évolutives (la dérive génétique à l'inoculation, la dérive génétique globale et la sélection) ainsi que l'impact de l'accumulation virale contrôlée par le fonds génétique ont été étudiés.

Stratégie : Plusieurs jeux de données comprenant des mesures effectuées sur 89 lignées haploïdes doublées de piment ont été utilisés pour effectuer une analyse statistique globale et essayer d'expliquer au mieux la durabilité de la résistance au PVY observée chez certaines lignées. Le premier jeu de données comprenait les mesures de taux de contournement du gène majeur pour chaque lignée de piment. Ces mesures permettaient d'évaluer la durabilité du gène majeur chez chaque lignée et ont été utilisées comme variable à expliquer. Des mesures d'accumulation virale et de dérive génétique à l'inoculation précédemment obtenues ont quant à elles été employées comme variables explicatives. Enfin, les lignées ont été phénotypées pour leur capacité à induire de la dérive génétique et de la sélection sur les populations virales au cours de l'infection jusqu'à 21 jours après inoculation, et ces données ont permis d'obtenir deux nouvelles variables explicatives. Un modèle linéaire généralisé a été appliqué pour étudier l'effet des différentes variables explicatives (accumulation virale, dérive génétique à l'inoculation, dérivé génétique globale et sélection) ainsi que de leurs interactions sur les données de contournement du gène majeur.

Résultats : Les quatre variables explicatives ainsi que leurs interactions deux à deux se sont révélées avoir un effet significatif sur la fréquence de contournement du gène majeur. L'effet de la sélection sur la durabilité s'est révélé particulièrement complexe et dépend fortement des niveaux d'accumulation virale et de taille efficace des populations virales dans la plante. Des hypothèses ont été proposées pour expliquer ces résultats.

Impact of genetic drift, selection and within-host accumulation on virus adaptation to its host plants

Rousseau, E. (1,2,3)*, Tamisier, L. (1,4)*, Fabre, F. (5), Simon, V. (1,6), Szadkowski, M. (4), Girardot, G. (1), Mailleret, L. (2,3), Grognard, F. (2), Palloix, A. (4), Moury, B. (1).

(1) Pathologie Végétale, INRA, 84140 Montfavet, France

(2) Biocore Team, INRIA, Sophia Antipolis, France

(3) Université Côte d'Azur, INRA, CNRS, ISA, France

(4) GAFL, INRA, 84140 Montfavet, France

(5) UMR SAVE, INRA, Villenave d'Ornon, France

(6) UMR BFP, INRA, Villenave d'Ornon, France

* co-first authors

Keywords : genetic drift, selection, effective population size, census population size, resistance breakdown, plant breeding

Abstract

Plant major resistance genes constitute an efficient protection against viruses, but their efficiency is limited by the emergence and spread of resistance-breaking mutants. Modulating the evolutionary forces acting on virus populations may constitute a promising way to increase the durability of major resistance genes. Here, we studied the effect of four plant traits affecting the intensity of such evolutionary forces (putative explanatory variables) on the frequency of resistance breakdown (RB) by a virus (the response variable). Two of those traits correspond to the effective virus population sizes, either at the plant inoculation step or during plant infection and are therefore

related to genetic drift undergone by the virus population. The third trait corresponds to the differential selection exerted by the plant on the virus population. Finally, the fourth trait corresponds to within-plant virus accumulation (VA) and is related to the probability of appearance of new mutations in the virus population. The levels of these traits were estimated from experiments on the *Potato virus Y* (PVY) - pepper (*Capsicum annuum*) system. A set of 89 doubled-haploid lines of *C. annuum* was analyzed. All lines carried the same major resistance gene $pvr2^3$ encoding a eukaryotic translation initiation factor eIF4E but had contrasted genetic backgrounds. They showed extensive variation for the rate of $pvr2^3$ resistance breakdown by PVY and for the four other traits of interest. A generalized linear model showed that the four explanatory variables and their two-by-two interactions had significant effects on the RB frequency. RB increased when PVY effective population sizes and VA increased. The effect of differential selection on RB was more complex because of strong interactions with the other explanatory variables. When VA (or PVY effective population sizes) was high, RB increased as selection intensity increased. An opposite relationship between RB and selection intensity was observed when VA (or PVY effective population sizes) was low. We provide hypotheses to interpret these relationships. This study provides a framework to select plants with appropriate virus-evolution-related traits to avoid or delay resistance breakdown.

1 Introduction

Resistance to pathogens, the capacity of a host to decrease its pathogen load (Restif and Koella 2004; Råberg et al. 2007), is widespread in plants. However, resistance intensity, specificity and genetic determinism are highly variable across genotypes of a given plant species. Up to date, plant breeders have mostly created resistant cultivars using resistance mechanisms showing monogenic inheritance and a high efficiency level, often called ‘qualitative resistance’. Unfortunately, the protection against pathogens conferred by such resistance genes was often poorly durable (McDonald and Linde 2002; García-Arenal and McDonald 2003). In the case of viruses, a maximum of ten cropping seasons with plant cultivars carrying a given resistance gene are usually sufficient for the counter-adaptation of the targeted virus, impairing resistance efficiency (García-Arenal and McDonald 2003). Usually, such ‘resistance breakdowns’ involve the selection by the plants of virus variants carrying mutation(s) in one, or a small number of, specific genes.

Four main evolutionary forces drive the evolution of virus populations at the within-plant scale: mutation, recombination, selection and genetic drift. Mutation creates new variants in the virus population, some of them potentially adapted to a hitherto efficient resistance gene. Recombination

(and reassortment in the case of viruses with a segmented genome) can also increase genetic diversity within virus populations by the exchange of genome parts between variants harboring nucleotide polymorphisms. Then, selection favors the variants with highest fitness, increasing their frequency from one generation to the next, at the expense of the weakest ones. This deterministic force is usually evaluated with the selection coefficient *s*, defined as the difference in fitness, *i.e.* relative growth rates when dealing with the within-plant scale, between two variants. By contrast, genetic drift acts in the same way on all variants of the population, introducing random fluctuations in the dynamics of variant frequencies (Charlesworth 2009). This stochastic force is commonly evaluated with the effective population size, defined as the size of an idealized population (*i.e.* a panmictic population of constant size with discrete generations) that would show the same degree of randomness in the evolution of variant frequencies as the population under consideration (Wright 1931; Kimura and Crow 1963).

The breakdown of a major resistance gene recently introgressed into commercial plant cultivars and used by growers can be schematically divided into three major steps that must be successfully achieved by the virus (Gómez et al. 2009; Moury et al. 2011). If we assume that no resistance-breaking virus variant is present initially, the first step is the appearance of such resistance-breaking variants from a wild-type (WT) virus population. The second step is within-plant accumulation and colonization of the resistance-breaking variants in competition with the rest of the virus population (*i.e.* WT components). Finally, in the third step, the resistance-breaking variants may have the opportunity to be transmitted to other plants and to spread in crops, allowing epidemics to develop in plant cultivars carrying the resistance gene. Different evolutionary forces rule these three steps. In the case of plant viruses, appearance of resistance-breaking variants usually involves a small number of nucleotide substitutions in the so-called avirulence factor encoded by the viral genome (Harrison 2002; Moury et al. 2011). Exceptionally, recombination may be required for appearance of resistance-breaking variants (Díaz et al. 2004; Miras et al. 2014). Then, accumulation of the resistance-breaking variants within plants depends on their relative fitness, *i.e.* on the selection coefficient, and on the intensity of genetic drift modulated by the bottlenecks that occur at multiple steps of plant infection (Gutiérrez et al. 2010, 2012; Zwart and Elena 2015). Modeling approaches have forecasted that the evolutionary forces acting at the within-plant scale, especially the mutational pathway involved in resistance breakdown and the fitness cost associated with the resistance-breaking mutation(s) accounted for about 50% of the risk of resistance breakdown in the field (Fabre et al. 2009, 2012b, 2015). Experimental data have also shown that these two factors were good predictors of the risk of resistance breakdown (Harrison 2002; Janzac et al. 2009; Fabre

et al. 2012a). The remaining 50% depended on factors related to virus epidemiology and thus mostly to step 3 of resistance breakdown.

One way to avoid or delay the breakdown of monogenic qualitative resistances is to combine the resistance gene with a suitable plant genetic background. A plant major-effect resistance gene combined with a partially-resistant genetic background showed a highly significant increase in durability compared to the same resistance gene associated with a susceptible genetic background. This was demonstrated experimentally for resistances targeting an RNA virus (Palloix et al. 2009), a fungus (Brun et al. 2010) or a nematode (Fournet et al. 2013). Indeed, the genetic background of the host can be composed of quantitative trait loci (QTLs) that affect the level of resistance to pathogens and the intensity of evolutionary forces acting on pathogen populations (Lannou 2012). When combined with a major resistance gene, the genetic background of a host plant may have three main effects, detailed thereafter from studies of (Quenouille et al. 2013, 2014, 2015). In the case of the *Potato virus Y* (PVY, genus *Potyvirus*, family Potyviridae) - pepper (*Capsicum annuum*; family Solanaceae) interaction, Quenouille et al. (2013, 2015) showed a significant correlation between the breakdown frequency of a major-effect resistance gene (the $pvr2^3$ gene) and the capacity of the virus to accumulate in the plant, *i.e.* the additional resistance level conferred by the plant genetic background. Assuming that the virus mutation rate was identical between plant genotypes, they hypothesized that within-plant virus accumulation was linked to the overall number of virus genome replications during plant infection and consequently to the probability of appearance of the resistance-breaking mutations in the virus population. Using a progeny of pepper genotypes carrying the same major resistance gene but contrasted genetic backgrounds, Quenouille et al. (2014) mapped quantitative trait loci (QTLs) controlling either within-plant virus accumulation or the frequency of breakdown of the major resistance gene in the pepper genome. This mapping revealed that the two QTLs controlling virus accumulation colocalized with QTLs controlling the frequency of breakdown of the major resistance gene, which provided a genetic explanation for the observed correlation between the two traits. By comparing two pepper genotypes carrying the same major resistance gene associated with either a partially-resistant or a susceptible genetic background, Quenouille et al. (2013) also showed that the selection of the most adapted resistance-breaking PVY mutants was slower in the former accession. The slower selection may be a consequence of the smaller selection coefficient of the adapted mutants, of a more intense genetic drift in the plants with a partially-resistant genetic background, or of both (Quenouille et al. 2013). Consequently, in the analyzed PVY - pepper system, the mutation, selection and/or genetic drift forces acting on the virus population seem to contribute to the breakdown or durability of the major-

effect resistance gene. Tu sum up, the three main hypothetical effects of the plant genetic background on the durability of a major resistance gene identified by Quenouille et al. (2013, 2015) are: (i) a decrease in the probability of appearance of the mutation(s) involved in resistance breakdown in the virus population, because of a decrease in the pathogen census population size, (ii) a slower selection of the virus mutants carrying the resistance-breaking mutations once they have appeared, because of a reduced selection coefficient between viruses carrying or not resistance-breaking mutations, or because of a reduced virus effective population size counteracting the effect of selection, and (iii) an increased extinction probability of the resistance-breaking mutants through more intense genetic drift during plant infection, because of small virus effective population size. The aim of the present study was to determine in the same PVY - pepper system which of the three factors, virus accumulation within-plants, selection coefficient between virus variants and virus effective population size, was (or were) determining the breakdown of the major resistance gene and to estimate their relative importance in order to help the breeding of future plant cultivars with durable virus resistance.

2 Results

The breakdown frequency of the *pvr2*^{3}–mediated resistance at the individual plant level (response variable 'RB') and several putative explanatory variables linked to within-plant PVY evolutionary processes (mutation, selection and genetic drift) were estimated in 89 doubled-haploid (DH) pepper lines. The terminology 'resistance breakdown' is frequently used to describe the increase of infection rates, and often of subsequent economic losses, in crops of plant cultivars carrying resistance genes following the adaptation of pathogen populations. For simplicity, we use this term to describe pathogen adaptation at the individual plant level, without any assumption about the consequences at the field scale or in terms of losses. The variable RB and the explanatory variables corresponding to within-plant virus accumulation (variable 'VA') and to the PVY effective population size at the plant inoculation step (variable 'N_e^{inoc}') were estimated previously (Quenouille et al. 2014; Tamisier et al. 2017; Table 1 and section Materials and Methods). In the present study, two more explanatory variables were estimated: the PVY effective population size during plant infection from inoculation to 21 days post inoculation (dpi) and the selection intensity exerted by the plant on the same PVY population (variables 'N_e' and 's', respectively; Table 1). Note that in the following, N_e and RB will correspond to the variables estimated following this specific experimental context, whereas we will use the expressions 'effective population size' and 'resistance breakdown', respectively, when used in a general manner. The methodology used to estimate N_e and s for each DH line was essentially identical to Rousseau et al. (2017).

Table 1: Description of the variables used in this study.

Variable	Experiment	Number of DH lines	Plants per DH line	PVY inoculum[a]	Infection stage / dpi[b]	Reference
RB	1	151	60	CI chimera	Systemic / 38 dpi	Quenouille et al. 2014
VA	2	151	10	Mutant N of CI chimera	Systemic / 36 dpi	Quenouille et al. 2014
N_e^{inoc}	3	151	20	Mutant K of SON41p-GFP	Inoculated cotyledon / 5-6 dpi	Tamisier et al. 2017
N_e	4	151 (89)[c]	8	SON41p mutants G, N, K, GK and KN	Systemic / 21 dpi	This study
s	4	151 (89)[c]	8	SON41p mutants G, N, K, GK and KN	Systemic / 21 dpi	This study

[a] One or two letter codes for mutants correspond to amino acid substitutions in PVY VPg allowing infection of plants carrying the *pvr2*[3] resistance gene.
[b] dpi: days post inoculation
[c] 151 DH lines were initially included in the experiment but only 89 were finally analyzed due to contaminations during RT-PCR revealed by the MiSeq Illumina sequencing.

2.1 Estimates of N_e and s corresponding to a PVY composite population in 89 pepper DH lines

The intensities of genetic drift and selection operating on PVY during plant infection were estimated from an experiment that consisted in analyzing the dynamics of an artificial population composed of five SON41p VPg variants carrying one or two nonsynonymous substitutions (named variants G, K, N, GK and KN) from inoculation to 21 dpi Rousseau et al. (2017). The inoculum consisted in a near-equimolar population of the five PVY variants. The frequencies of the variants in the common inoculum and in pools of three systemically-infected leaves at 21 dpi for each plant were accurately determined using MiSeq Illumina high-throughput sequencing (HTS) of the region of the VPg cistron where are located the mutations that distinguish the five variants. The effective population size of the composite PVY population during plant infection from 0 to 21 dpi (N_e) and the relative growth rates of the five PVY variants (r_i with $i \in \{G, K, N, GK, KN\}$) were jointly estimated for each DH line thanks to a method developed recently (Rousseau et al. 2017). Briefly,

this method allows to estimate jointly the N_e and r_i parameters of a multi-allelic Wright-Fisher model for haploids (Ewens 2004) and does not require neutral markers. The fitness (r_i) rank among the five PVY variants was highly similar among the 89 DH lines, as described previously on a limited set of 15 DH lines (Figure 1; Rousseau et al. 2017). In 82 of the 89 DH lines, variant 'G' was the least fit. In the seven remaining lines, variant 'N', which shows on average an intermediate fitness among the DH lines, was the least fit. The three variants possessing the 'K' mutation (*i.e.* variants 'K', 'GK' and 'KN') showed the highest fitness with small r_i differences among them. Overall, in 86 of 89 lines, one of these three latter variants had the highest fitness. This similarity of r_i ranking between the five PVY variants among the 89 DH lines justifies to estimate the selection effect s of each DH line on the PVY population by the standard deviation of the growth rates r_i of the five PVY variants. Similar, but less robust, results were obtained by considering that s equals the difference in r_i between the fittest and the least fit PVY variant (data not shown). Among the 89 DH lines, s and N_e showed continuous distributions (Figure 2). s varied from 0.037 to 0.160 and presented a near-Normal distribution (Figure 2a). N_e varied from 15.4 to 289.1 and the DH line distribution was skewed towards small values (Figure 2b). About the two thirds of lines (60 of 89) had N_e<50 and only six lines had N_e>100. Figure 2c shows a set of bar plots representing the frequencies of the five PVY variants in the 8 sampled plants and on average for four DH lines showing contrasted N_e and s values.

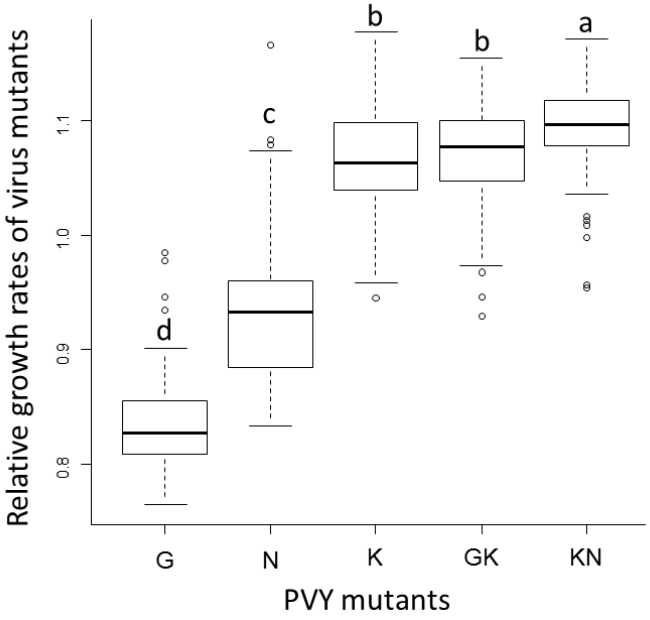


Figure 1: Relative growth rates of the five PVY variants in 89 pepper DH lines estimated using Rousseau et al.'s (2017) model. Different letters indicate significantly different groups in Mann-Whitney-Wilcoxon tests with Bonferroni correction.

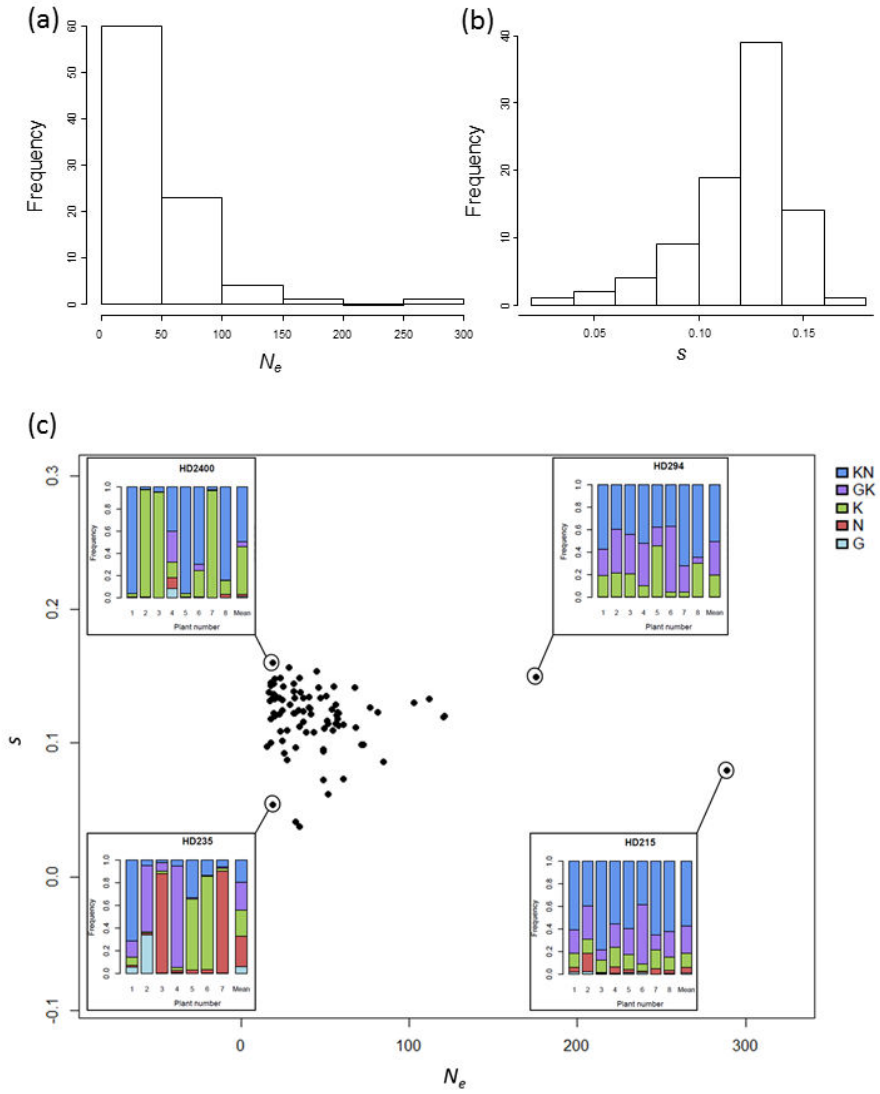


Figure 2: Distribution of (a) effective population sizes (N_e) and (b) selection effects (s) exerted on a composite PVY population during plant infection among 89 pepper DH lines. s was estimated as the standard deviation of the relative growth rates r_i of the five variants composing the PVY population. The plot in (c) represents the distribution of the 89 pepper DH lines according to N_e and s. For four DH lines showing contrasted N_e and s values, bar plots illustrating the dynamic of the PVY population used to measure N_e and s are provided. Each bar corresponds to a plant initially infected with an equimolar mixture of five PVY variants (G, N, K, GK and KN). Within each bar, the frequencies of the five PVY variants at 21 dpi is represented. The last bar indicates the mean viral composition in the infected plants.

In our experimental context, Rousseau et al.'s (2017) model was shown to adjust satisfactorily to the experimental observations (Figure 3). Indeed, the coefficients of correlation between observed and adjusted values was r=0.89 (Pearson coefficient; p-value<2.2e-16) for the mean frequencies of each of the five PVY variants among the eight plants analyzed per DH line (Figure 3a), while it was r=0.67 (Pearson coefficient; p-value=3e-13) for the standard deviations of the mean frequencies of the five variants among the eight plants (Figure 3b). The heritability of N_e and s was estimated as in Rousseau et al. (2017): for each DH line, two subsamples of four plants were

randomly defined and N_e and s were estimated with Rousseau et al.'s (2017) model separately for each subsample. Then, broad-sense heritability h^2 was calculated by considering these two subsamples as experimental replicas as in Quenouille et al. (2014). The entire process was repeated seven times with independent random subsamples. Note that h^2 calculated in this manner is an underestimate of the true heritability since it considers only four plants per DH line for N_e and s estimations whereas eight plants are used in the whole experiment. For s, heritability was medium (h^2 = 0.44; mean of the eight subsamplings), while it was poor for N_e (h^2 = 0.16). This means that 44% (respectively 16%) of the variance of s (respectively N_e) is explained by the plant genetic diversity. Given these rather low h^2 values, we created a second dataset by discarding the 38 DH lines showing the less robust N_e and s estimations over the eight replicas. This led to a 51-DH-line set with higher h^2 values (0.56 and 0.48 for s and N_e, respectively). This dataset was analyzed in the same way as the global one and gave similar results (Tables S1 and S2).

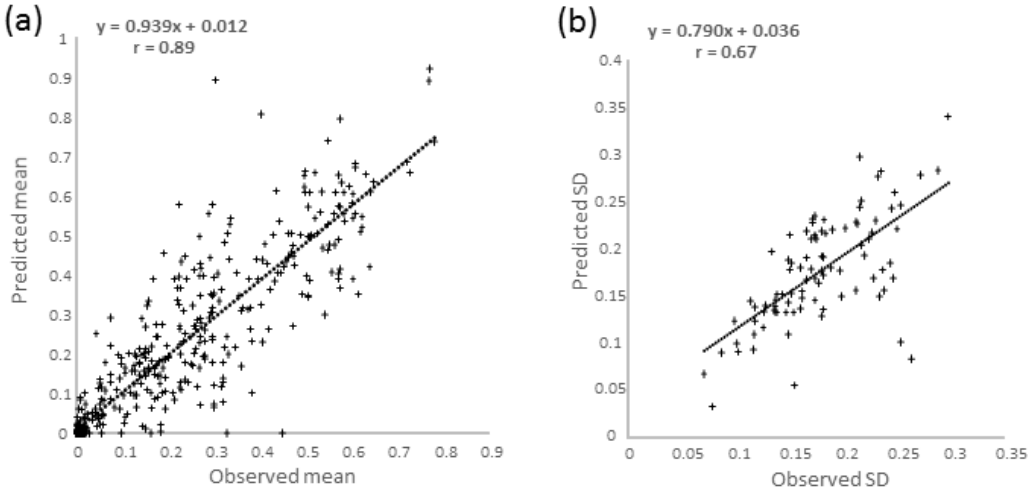


Figure 3: Ajustment of the model (Rousseau et al. 2017) used to estimate N_e and s to (a) the observed mean frequency of each of the five virus variants in the population (hence five values per DH line) and to (b) the observed standard deviation (SD) of the mean frequencies of the five virus variants (hence one value per DH line).

2.2 Correlation between putative explanatory variables and with the RB response variable

Table 2 and Supplementary Figures S1 and S2 show the relationships among the variables considered in our analyses. Weak or no correlation was observed among the explanatory variables N_e, s, N_e^{inoc} and VA. N_e and VA showed a significant positive correlation (Pearson's r=0.28, p-value=0.0067). However, this effect was mainly due to an outlier DH line showing an extreme N_e value and the highest VA value (Figure S1). After withdrawing this DH line or using the Spearman' rank correlation instead of Pearson's r, no significant correlation was observed between N_e and VA

(p-values≥0.08). Similarly, a negative and marginally significant correlation was also observed between s and N_e^{inoc} only with Pearson's r (r=-0.23, p-value=0.033). In contrast, weakly significant correlations were observed only with Spearman's ρ between N_e and s (ρ=-0.25, p-value=0.016) and between VA and s (ρ=-0.21, p-value=0.048). No correlation was detected between N_e and N_e^{inoc} or N_e^{inoc} and VA. Moreover, variance inflation factors (VIF) were calculated for each explanatory variable to check for multicollinearity. All VIF values were below 4, which ensures the lack of multicollinearity in our analysis.

Table 2: Pearson r (below the diagonal) and Spearman rank ρ (above the diagonal) correlation coefficients and associated significance (p-values; H_0: r=0 or ρ=0) between the *pvr2*^{3} resistance breakdown frequency (RB) and four explanatory variables linked to evolutionary forces exerted by plant genotypes on PVY populations for 89 pepper DH lines. Explanatory variables are the selection s (*i.e.* the standard deviation of the intrinsic growth rates of five PVY variants), the effective population size N_e corresponding to the colonization of the plant by the PVY population, the effective PVY population size at inoculation (N_e^{inoc}) and the initial viral accumulation (VA) characterizing each DH line. Correlations that are significant with both the Pearson and Spearman coefficients are underlined in gray.

	RB	***s***	N_e	N_e^{inoc}	**VA**
RB		ρ = -0.27 p = 0.012*	ρ = 0.012 p = 0.91	ρ = 0.29 p = 0.005**	ρ = 0.48 p = 2e-06***
s	r = -0.26 p = 0.015*		ρ = -0.25 p = 0.016*	ρ = -0.11 p = 0.31	ρ = -0.21 p = 0.048*
N_e	r = -0.015 p = 0.64	r = -0.145 p = 0.18		ρ = 0.09 p = 0.42	ρ = 0.19 p = 0.08
N_e^{inoc}	r = 0.35 p = 0.00083***	r = -0.226 p = 0.033*	r = -0.058 p = 0.59		ρ = 0.13 p = 0.23
VA	r = 0.36 p = 0.00048***	r = -0.173 p = 0.10	r = 0.28 p = 0.0067**	r = 0.12 p = 0.27	

In contrast, three explanatory variables were significantly and consistently correlated with the response variable RB: s, N_e^{inoc} and VA (Table 2). The correlation was positive and highly significant between RB and either N_e^{inoc} or VA. The correlation was negative and moderately significant between RB and s. Finally, no significant link was noticed between RB and N_e. Plots representing pairs of variables for the 89 DH lines did not reveal particular relationships between them other than the linear or ranking relationships revealed by the correlation analyses (Figures S1 and S2).

GLM analyses were performed to explain the response variable RB with the four explanatory variables linked to evolutionary forces exerted by pepper DH lines on PVY populations. A first model incorporating the four explanatory variables and their pairwise interactions indicated that all factors and interactions had a highly significant effect on RB (Table 3). The same model was retained after a stepwise selection procedure based on the Akaike's information criterion (AIC). The model fit was moderate with McFadden $R^2 = 0.35$ (McFadden, 1973).

Table 3: Generalized linear model (GLM) analysis of the frequency of *pvr2*^{3} resistance breakdown frequency (RB) in 89 pepper DH lines with four explanatory variables linked to evolutionary forces exerted by plant genotypes on PVY populations. Explanatory variables are the selection s (*i.e.* the standard deviation of the relative growth rates of the five PVY variants), the effective population size N_e corresponding to the colonization of the plant by the PVY population, the effective PVY population size at inoculation (N_e^{inoc}) and the initial viral accumulation (VA) characterizing each DH line. All four variables and all their two-by-two interactions were kept after performing a stepwise selection procedure and the selection of the best-fitting models using the Akaike's information criterion (AIC).

Variable (or variable pair)	Estimate	Std. Error	z value	*p*-value
Intercept	8.419	0.7131	11.806	<2e-16 ***
s	-84.95	5.164	-16.451	<2e-16 ***
N_e	-7.426e-02	1.366e-02	-5.436	5.44e-08 ***
N_e^{inoc}	-0.1819	2.082e-02	-8.734	<2e-16 ***
VA	-2.817	0.40	-7.042	1.89e-12 ***
$s \times N_e$	0.2921	7.838e-02	3.727	2e-04 ***
$s \times N_e^{inoc}$	1.428	0.1482	9.633	<2e-16 ***
s×VA	24.83	2.906	8.542	<2e-16 ***
$N_e \times N_e^{inoc}$	1.017e-03	2.832e-04	3.592	3e-04 ***
$N_e \times$ VA	1.276e-02	2.436e-03	5.237	1.64e-07 ***
$N_e^{inoc} \times$ VA	2.696e-02	8.796e-03	3.065	0.0022 **

The pairwise interactions of variables revealed their synergistic effects on RB (Table 3 and Figure 4). Indeed, VA and N_e^{inoc} had a positive effect on RB, but this effect increased with increasing values of VA or N_e^{inoc}, as attested by the higher slope value. The same trend was observed with VA and N_e or with N_e^{inoc} and N_e. For the three additional pairwise interactions, which all included the

variable s, the effect on RB was contrasted depending on the values of the variables. For example, when VA is small, RB decreased with increasing s values, whereas when VA is high RB increased with increasing s values (Figure 4). Again, the same trends were observed when considering the joined effects of s and N_e (or s and N_e^{inoc}) on RB.

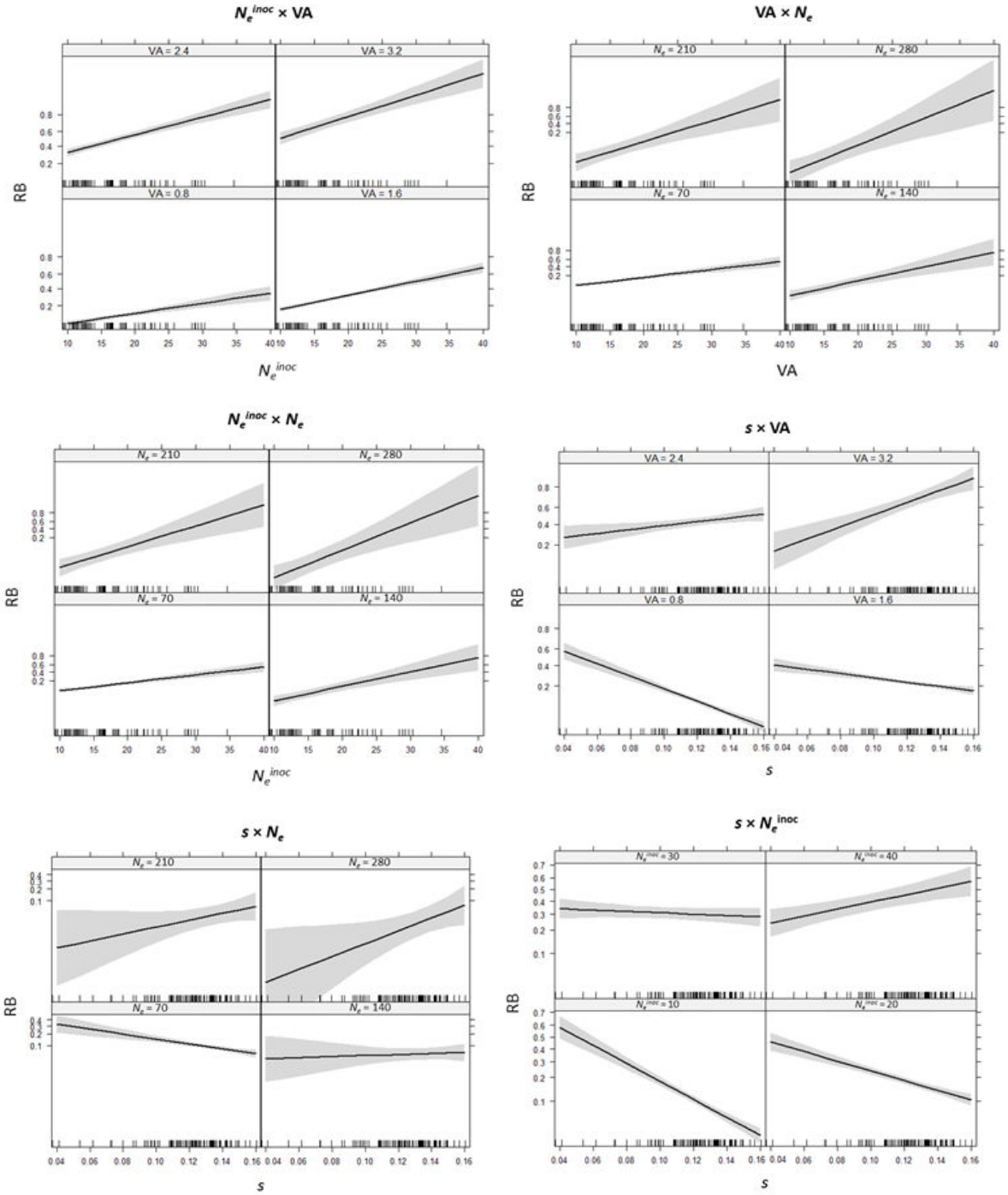


Figure 4: Pairwise effects of explanatory variables on the response variable RB, the frequency of breakdown of the *pvr2*^{3}–mediated resistance in 89 pepper DH lines, based on GLM analyses. Explanatory variables are within-plant virus accumulation (VA), effective population at the plant inoculation step (N_e^{inoc}) or during plant colonization (N_e) and selection intensity exerted by the host plant on the composite PVY population (s).

We computed conditional inference regression trees to synthesize the effect of the four explanatory variables on RB (Figure 5). Regression trees take into account the interactions between explanatory variables and indicate which combinations of variables and variable levels correspond to higher or lower RB. The first dichotomy in the regression tree for RB was linked to the variable VA, reflecting its strong influence on RB. The second and third dichotomies were linked to the variables s and N_e^{inoc}. The variable N_e was not retained in this analysis probably because of its weaker effect on RB. The best way to reduce RB is to combine a weak VA (≤0.609) and a high s (>0.117). When VA is higher (>1.474), a low N_e^{inoc} value (≤16) contributes also significantly to reduce RB.

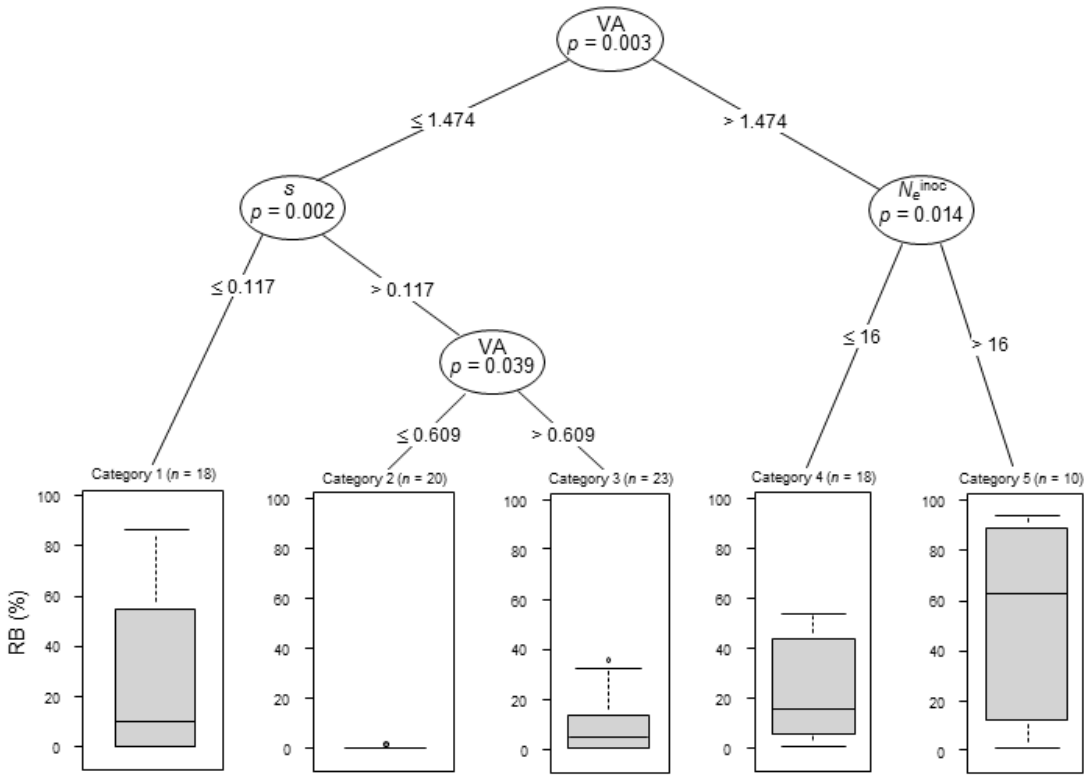


Figure 5: Conditional inference regression tree modelling the frequency of breakdown of the *pvr2*^{3}–mediated resistance (RB) in 89 pepper DH lines with three explanatory variables representing evolutionary forces exerted by the host plant on PVY populations (VA, N_e^{inoc} and s). The fourth explanatory variable (N_e) was not retained in the analysis. n: number of DH lines in each category.

3 Discussion

The main objective of this article was to study the relationships between the frequency of breakdown of the major-effect resistance gene $pvr2^3$ to PVY in pepper (variable RB) and proxy variables quantifying the main evolutionary forces exerted by the host plants on the virus population. Indeed, we assumed that some of these evolutionary forces, notably the intensities of genetic drift and selection (N_e^{inoc}, N_e and s) and the resistance efficiency (inversely related to VA), may be used to predict RB (Quenouille et al. 2013).

3.1 Lack of strong relationships between N_e^{inoc}, N_e, s and VA

Based on the literature, several correlations between the explanatory variables N_e^{inoc}, N_e, s and VA could have been expected with our dataset. First, a positive correlation could have been expected between N_e^{inoc} and N_e. Indeed, experiments conducted with *Tobacco etch virus* (TEV; genus *Potyvirus*) in pepper or tobacco plants showed that the probability that a primary infection focus causes systemic infection is ~1.0 and therefore within-plant genetic drift is mostly determined by the inoculation step and not by the colonization of inoculated or apical leaves (Zwart et al. 2011, 2012). Second, a similar positive correlation could have been expected between N_e^{inoc} and VA, because for potyviruses a low N_e^{inoc} was shown to result in a delay in plant infection at the systemic level (Rodrigo et al. 2014) or in a lower proportion of infected cells (Lafforgue et al. 2012), hence probably also in a lower VA. Contrary to these assumptions, we did not observe any relationship between N_e^{inoc} and N_e or between N_e^{inoc} and VA (Table 2; Figure S1). These discrepancies could be due to the plant genotypes used for the experiments, either highly susceptible to virus infection (Lafforgue et al. 2012; Rodrigo et al. 2014) or partially resistant (this study). Partial resistance mechanisms acting specifically on the virus systemic movement may be the cause of the lack of relationship between the density of infection foci in the inoculated leaf and the efficiency of systemic infection. This is consistent with the observation of a narrow bottleneck between 6 and 10 dpi, *i.e.* at the onset of systemic infection, in the majority of the pepper DH lines (8 of 15 tested) (Rousseau et al. 2017), whereas such a bottleneck did not seem to occur in the experiments performed with TEV and *Nicotiana tabacum* (Zwart et al. 2011). Alternatively, the links with N_e^{inoc} may have occurred earlier in the infection process and may have blurred with time until they disappeared at 21 (for N_e) or 36 dpi (for VA) (Table 1). Accordingly, in the case of TEV and *Nicotiana tabacum*, Zwart et al. (2012) showed that the effects of N_e^{inoc} on VA are mainly due to a delay in the colonization of the inoculated leaf and are transient.

Third, a positive correlation between s and N_e (or N_e^{inoc}) could also have been expected. Indeed, small effective population sizes reduce the genetic variability of populations and consequently the effectiveness of selection (Kimura 1970; Rouzine et al. 2001; Charlesworth 2009) as well as the rate of appearance and fixation of adaptive mutations (Lanfear et al. 2014). Similarly, in our experimental frame, Quenouille et al. (2014) argued that the effect of the pepper genetic background to slow down the selection of the most adapted PVY variants (the GK and KN mutants) may be a side effect of genetic drift and reduced effective population size. Analysis of the 89 DH lines showed, on the contrary, that s and N_e (or N_e^{inoc}) were negatively correlated though the link was marginally significant (Table 2; Figure 2c). An important consequence of these poor correlations is that some mechanisms act directly on s and independently of their effects on N_e or N_e^{inoc} and one may choose pepper DH lines with contrasted effects on virus effective population size and selection intensity to control virus evolution.

In contrast, we did not expect any correlation between N_e and VA and indeed the link was weak (Table 2). This is mainly due to the numerous and transient bottlenecks that viruses undergo during plant colonization (Gutiérrez et al. 2010). Indeed, for populations varying in size over time, the effective population size over a given number of generations can be approximated by the harmonic mean of the effective population sizes at each generation. Consequently, because they greatly decrease population size, these bottlenecks have a disproportionate effect on the overall value of the effective population size (N_e in our case) (Lande and Barrowclough 1987; Charlesworth 2009) but founder effects after each of the bottlenecks contribute to maintain high VA values. This means also that the two types of virus population sizes, *i.e.* the census (VA) and effective (N_e) population sizes, are complementary levers to control virus infection and evolution, while usually only VA is considered by breeders and pathologists for resistance phenotyping.

3.2 Which evolutionary forces contribute most to resistance breaking?

There are complex relationships between the number of viral mutations required for host adaptation, the probability of appearance of each of these mutations, the incurred fitness changes and genetic drift (Iwasa et al. 2004; Fabre et al. 2009; Quenouille et al. 2013; da Silva and Wyatt 2014). As a consequence, the effects of VA, s, N_e and N_e^{inoc} on RB, and complex interactions between them, were anticipated.

A strong positive correlation between VA and RB was observed earlier (Quenouille et al. 2014). It may be the consequence of the link between VA and the probability of appearance of the resistance-breaking mutations in the inoculated plants (Quenouille et al. 2013). Indeed, the

probability of appearance of a given mutation in a plant virus depends on its intrinsic mutation rate, for which only few estimates have been experimentally obtained (Malpica et al. 2002; Gago et al. 2009; Tromas and Elena 2010; de la Iglesia et al. 2012), on the type of mutation (for example the transition *vs*. transversion types of nucleotide substitutions; Ayme et al. 2006; Quenouille et al. 2013), on the population growth rate and on the generation time, which is poorly known (Khelifa et al. 2010; Martínez et al. 2011). VA is an indicator of the census population size at a given time-point, which depends on the population growth rate and generation time. Assuming that the intrinsic mutation rate is similar between DH lines, the VA level would actually be a proxy of the probability for a new mutation to occur in a given amount of time. Moreover, the link between VA and RB implies that the within-plant accumulation of the resistance-breaking variant (CI chimera carrying the 'N' substitution; Table 1), that was used to measure VA, is correlated to the residual accumulation of the WT variant (CI chimera), that was used to estimate RB. We could not measure the accumulation of the WT variant in these plants. Indeed, the *pvr2*^{3} resistance gene does not control completely PVY multiplication or even movement within the plant (Montarry et al. 2011). However, in 99% of the infected plants, these WT viruses were outcompeted by the resistance-breaking mutants, that appeared stochastically and then emerged rapidly, and they could not be observed. In addition, Montarry et al.'s (2011) results suggested that the PVY resistance-breaking mutants did not preexist in the inoculum and resistance breakdown was therefore linked to the residual accumulation of the WT PVY variant in these plants. Altogether, these results suggest that the resistance-breaking mutants appear in the inoculated plants, but less likely when VA is reduced.

The strong positive correlation between N_e^{inoc} and RB was expected if we consider, as mentioned above, that genetic drift frequently slows down adaptation. However, since the PVY resistance-breaking mutants appeared most probably in the inoculated plants, the decrease of RB when N_e^{inoc} is low is certainly not due to purging the PVY inoculum of putative resistance-breaking mutants but rather to a reduced chance of appearance and slower fixation of the resistance-breaking mutations in the early steps of infection (inoculated cotyledons).

Similarly, we anticipated a positive correlation between s and RB, since higher s values would accelerate the fixation of resistance-breaking mutations in the PVY populations. Though the overall link between s and RB was negative (Table 2), the effect of s on RB is complex due to strong interactions, especially with VA and N_e^{inoc} (Table 3). When VA is high, RB increased with increasing s values, which fits with our expectations. However and unexpectedly, when VA is small, RB decreased with increasing s values (Figures 4 and 5). This latter result may be attributable to our measure of s. We could not evaluate the selection coefficient s between a wild-type PVY and a resistance-breaking variant in plants carrying the major resistance gene because the tremendous

competition leads rapidly to the extinction of the wild-type variant (Montarry et al. 2011). As a consequence, we calculated *s* as the standard deviation in relative growth rates between five resistance-breaking variants that display different levels of adaptation, which may be not informative for RB that was measured with a wild-type variant with very low accumulation (Table 1). In that case, we would have expected no relationship between RB and *s*, but not a negative one. A further hypothesis is linked to relationships between VA, the type of resistance-breaking substitutions (transitions or transversions; in PVY, transitions are 5 to 8 times more frequent than transversions; Ayme et al. 2006) and *s*. In the experiments aiming at estimating RB, the resistance-breaking mutants were exclusively single-nucleotide mutants, with a majority of the N mutant and a minority of the G, K and a few other mutants (Ayme et al. 2006; Montarry et al. 2011). Double-mutants (GK and KN for example) need more time to appear and become fixed in the population (Quenouille et al. 2013). We can imagine what would be the effects of VA and *s* on RB in the three main groups of plants defined by the regression tree (Figure 5). When VA is high (>1.474; plant group 1), any kind of PVY single mutant is likely to appear rapidly, either issued from a transition (G and K mutants) or from a (rarer) transversion (K mutant). Thus, a resistance-breaking mutant with a high fitness like mutant K is likely to become fixed in a quite large proportion of plants. Because of the higher virus diversity expected in these plants, the selection parameter *s* is expected to amplify the effect of VA to increase RB. In contrast, when VA is low (≤1.474), the K mutation has a low probability to appear and be involved in resistance-breakdown, while the N and G mutations that are inherently more frequent would still case resistance-breakdown. In that cas, a high s value (s>0.117; plant group 2; Figure 5) is mostly due to a large fitness difference between mutants K, GK and KN on one side, and mutants G and N on the other side (Figure 6). Lower s values (s≤0.117; plant group 3) are associated to smaller fitness differences between mutants (Figure 6). As a consequence, the relative growth rates of mutants N and G, that are most likely to appear, are significantly higher for group 3 than for group 2. The same is probably true for their absolute growth rates, since VA is almost identical between the two groups (0.683 and 0.681 on average for groups 2 and 3, respectively; *p*-value>0.9, Kruskal-Wallis test). Then, starting with a wild-type virus with a low fitness as was the case in the experiment used to measure RB, the fitness gain provided by mutations G or N would be higher for plants of group 3, with lower *s* values, than for plants of group 2, explaining the higher RB observed in that case (Figure 6).

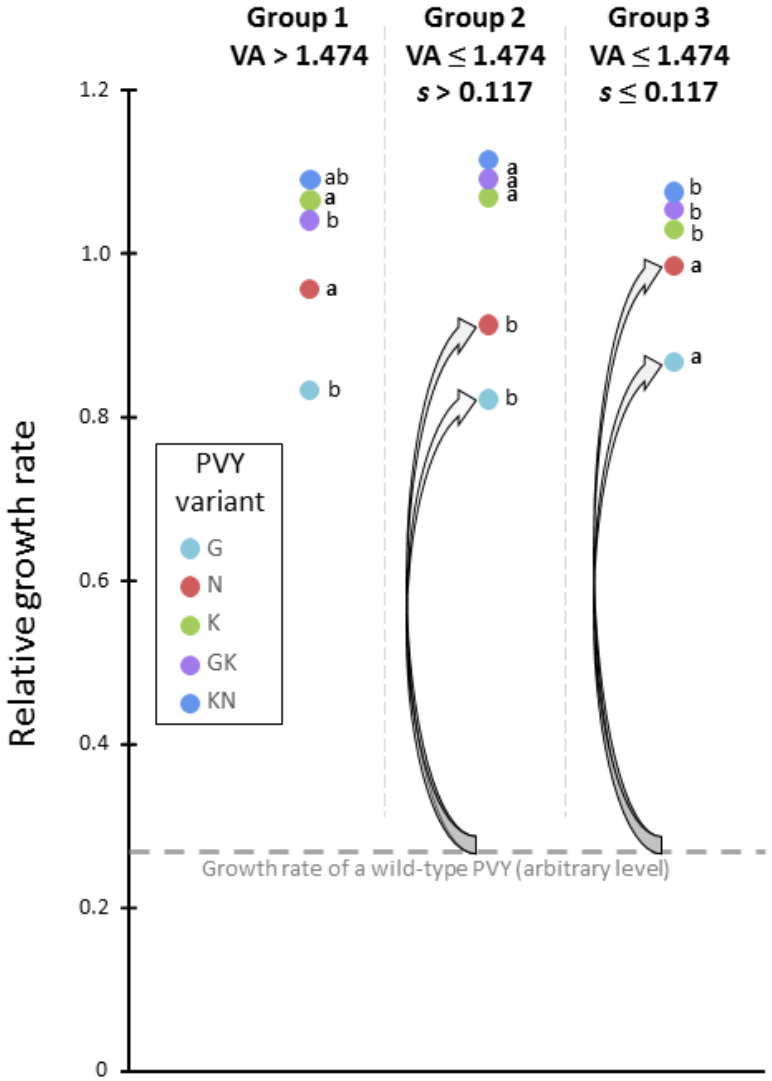


Figure 6: Interpretation of the combined effects of virus accumulation (VA) and selection intensity (s) on the frequency of resistance breakdown (RB) among three groups of DH lines defined by the regression tree (Figure 5). Represented are the mean relative growth rates of the five PVY variants used to estimate s in experiment 4 (Table 1) among each plant group. Letters represent, for each PVY variant, significant differences in relative growth rates among plant groups (Kruskal-Wallis test, p-value<0.05). Double mutants GK and KN did not appear in the resistance breaking experiments but were used for estimation of s. When VA is high (group 1), the K mutant which requires a transversion is likely to appear and confer a high fitness gain to PVY. When VA is lower, resistance breakdown involves more likely mutants which require a transition (G and N) but confer a lower fitness gain. Since RB was evaluated in experiment 1 (Table 1) with a wild-type PVY, its initial fitness in *pvr2*^{3}-carrying plants is low (fictitious broken line). Arrows indicate that the fitness gains associated with resistance-breaking mutations G and N are higher for group 3 than for group 2, despite higher s values in the latter group. This may explain the negative effect of s on RB when VA is low (Figure 4).

3.3 Applied consequences to improve the durability of major-effect resistance genes

Breeding programs aiming to create plant varieties with efficient and durable pathogen resistance consider frequently the pathogen load (VA in our case) for phenotyping. Overall, our results suggest that the intensity of genetic drift is also a relevant trait to promote the durability of plant major resistance genes to viruses (Zhan et al. 2015), as proposed more generally to limit the emergence of microbe variants adapted to drug treatments (Abel et al. 2015). In our case, this is particularly true for genetic drift acting at the inoculation step (N_e^{inoc}). We showed that the intensity of selection exerted by the plant on a composite virus population is also a relevant factor to consider. Similarly, Lê Van et al. (2013) showed that the intensity of selection exerted by apple trees on a composite fungus population was inversely correlated with the spectrum of action of the resistance, hence also inversely correlated with the durability potential of the resistance. However, the numerous, and sometimes unexpected, interactions that we observed between these factors make it difficult to choose the most appropriate combinations of factors and factor levels to enhance resistance durability. For this, the regression tree can provide decision rules (Figure 5). According to this tree, to reduce RB, the primary factor is to reduce VA. Then, if VA is high, on can still reduce RB by choosing plants with low N_e^{inoc} values. Finally, low VA values can be combined with high s values to still improve resistance durability. Note however that this latter effect should be taken with caution, because it is certainly linked to the experimental context used to estimate s. Moreover, if one can argue that N_e^{inoc} estimates and, to a lower extent, VA estimates may be representative of diverse PVY populations, at least if plant-virus genotype × genotype interactions are limited, the s estimates may vary strongly with the PVY population used (number, type and initial frequency of variants).

The interest of using these evolution-related traits in plant breeding depends also on the ultimate cause of the observed relationships, *i.e.* genetic pleiotropy or linkage. Previous results showed that QTLs corresponding to VA map to the same pepper genome regions than QTLs controlling RB, though the QTL intervals are large (Quenouille et al. 2014). Hence, pleiotropy is not excluded in this case. In contrast, the linkage hypothesis is more probable in the case of the link between N_e^{inoc} and RB (Tamisier et al. 2017) and the effect of N_e^{inoc} on RB may be incidental. However, it should be noted that the power to detect QTLs is always limited by the size of progenies analyzed, the density of molecular markers, the phenotyping precision and the QTL effect. Hence, QTLs might have been missed in these studies. Unfortunately, concerning N_e and s, the population size of 89 DH lines is too small for QTL mapping, with high risks of both false-positive and false-

negative QTLs (Barchi et al. 2009). Hence, we cannot make assumptions about the linkage or pleiotropy between these traits and RB.

As a conclusion, we provide evidence that plant traits accounting for the plant impact on virus evolution have a great potential to reduce the risks of breakdown of a major resistance gene. Phenotyping of such traits becomes feasible with recent sequencing technologies and population genetics models, and trait estimates are precise enough to ensure moderate to high trait heritabilities. Hence, conditions are met to include these, or similar, approaches in plant breeding programs aiming to improve the durability of pathogen resistance.

4 Materials and methods

4.1 Previous data

Different variants of the WT PVY clone SON41p (Moury et al. 2004) were chosen to estimate the response variable RB (breakdown frequency of the *pvr2*^{3}–mediated resistance) and the explanatory variables corresponding to within-plant PVY accumulation (VA) and PVY effective population size at inoculation (N_e^{inoc}). RB was previously evaluated after inoculation of each of the DH lines (60 plants per DH line) with the 'CI chimera', an artificial recombinant of SON41p carrying the cylindrical inclusion (CI)-coding region of PVY isolate LYE84.2 (Table 1) (Montarry et al. 2011; Quenouille et al. 2014). This variant was preferred to SON41p because of its higher ability to break the *pvr2*^{3} resistance, providing a larger range of RB values among DH lines, hence allowing a higher precision for genetic and statistical analyses. Still, RB obtained with SON41p was shown to be highly correlated with RB obtained with the CI chimera on a subset of sixteen contrasted DH lines (Quenouille et al. 2013). RB corresponds to the frequency of plants showing virus infection at the systemic level around one month after inoculation and was shown to correspond to situations where a nonsynonymous mutation in the VPg-(viral protein genome-linked) coding region became fixed in the PVY population, conferring adaptation to the *pvr2*^{3} resistance (Ayme et al. 2006; Montarry et al. 2011). In this experiment, a high heritability (h^2 = 0.87) was shown for RB (Quenouille et al. 2014).

PVY accumulation in plants was measured in infected leaves by quantitative DAS-ELISA (double antibody sandwich enzyme-linked immunosorbent assay) (Quenouille et al. 2014). Mean relative virus accumulation (VA) was assessed in pools of three systemically-infected leaves per plant at 36 days post-inoculation (dpi) in 10 plants per DH line (Quenouille et al. 2014; Table 1).

For this, a mutant of the CI chimera carrying the aspartic acid to asparagine substitution at amino acid position 119 of the VPg ('N' substitution) that allowed infection of plants carrying $pvr2^3$ was used. A fairly good heritability was observed for VA ($h^2 = 0.64$) (Quenouille et al. 2014).

Finally, effective population size at plant inoculation (N_e^{inoc}) was estimated with a SON41p variant carrying a GFP (Green Fluorescent Protein) reporter gene and a single amino acid substitution in the VPg (threonine to lysine substitution at codon position 115; 'K' substitution) which allows infection of plants carrying $pvr2^3$ (Tamisier et al. 2017; Table 1). N_e^{inoc} corresponds to the mean number of PVY primary infection foci visualized by the GFP fluorescence 5 or 6 days after mechanical inoculation of 20 pepper cotyledons per DH line. N_e^{inoc} was shown to have a high heritability ($h^2 = 0.93$).

4.2 Analysis of composite PVY populations infecting pepper DH lines

An experiment was dedicated to estimate two more explanatory variables for RB: the PVY effective population size during plant infection (N_e) and the selection intensity exerted by the host genotype on a composite PVY population (s). The experimental design was as in Rousseau et al. (2017) except two main differences: (i) the experiment comprised initially 151 pepper DH lines instead of 15 and (ii) only one plant sampling date (21 dpi) was retained instead of six, to keep the experimental size compatible with MiSeq Illumina sequencing capacity. The sampling date at 21 dpi was chosen based on the previous experiments (Rousseau et al. 2017) as a balance between the time for selection to operate on the PVY population and the risk of extinction of the less fit PVY variants if too much time was allowed. In the latter case, the lack of genetic information in the PVY populations would preclude estimation of N_e and s with Rousseau et al.'s (2017) model. All the 151 DH lines of *C. annuum* carried the PVY resistance allele $pvr2^3$ and differed in their genetic background (Quenouille et al. 2014). They were issued from the F_1 hybrid between 'Perennial', a PVY-resistant pepper line carrying the $pvr2^3$ allele, and 'Yolo Wonder', a PVY-susceptible line (Quenouille et al. 2014). The five SON41p variants, named G, N, K, GK and KN based on their amino acid substitutions in the VPg, were mixed in equimolar amounts based on quantitative DAS-ELISA and were mechanically inoculated to the two cotyledons of eight plants per DH line. Each of these mutations or mutation pairs (double-letter names) (*i.e.* the 'K' and 'N' mutations defined previously and the 'G' mutation corresponding to the serine to glycine substitution at codon position 101 of the VPg) conferred to PVY the capacity to infect of plants carrying $pvr2^3$. At 21 dpi, for each plant, all apical leaves were collected, pooled together and crushed in buffer before RNA purification as in Rousseau et al. (2017). One-step reverse-transcription polymerase chain reaction

(RT-PCR) amplification was conducted for the 1208 plants (8 plants × 151 DH lines) individually, in thirteen 96-well PCR plates. In all, the sampling, RT-PCR and sequencing procedures performed to obtain the viral sequences were essentially as described by Rousseau et al. (2017). Three of the eight primers used by Rousseau et al. (2017) were poorly efficient in PCR (corresponding to primer tags 5'-GGTCTAGTAC, 5'-GAGGCTCTAC and 5'-TGCTGATATC), and were thus replaced with primer tags 5'-CGACGACTGC, 5'-TGGAGTACGA and 5'-GGAGCGTCAC, respectively. To avoid opening of the reaction microtubes and hence cross-contamination between microtubes of the same RT-PCR plate, single-step RT-PCRs were performed on RNA extracts on a first set of 13 PCR plates. For 66 DH lines, no RT-PCR products were detected by agarose gel electrophoresis for at least 4 of the 8 plants. A two-step RT-PCR protocol was thus carried out for all the samples of these 66 DH lines in six additional 96-well PCR plates. This protocol usually provides a higher sensitivity than the first one but increases the risk of contamination (Bustin 2000). Unfortunately, MiSeq sequencing results showed that significant contaminations occurred in 4 of these 6 additional PCR plates (see below). As next-generation sequencing data may be impacted by contaminations and potential biases during PCR amplification, several controls were conducted. Potential cross-contamination among samples during the crushing step in a roller mill was evaluated thanks to plants of the pepper genotype Yolo Wonder, mechanically inoculated with the SON41p G-bis variant (aspartic acid to glycine substitution at codon position 119 of the VPg) which differs by at least two nucleotides from the five variants studied. Leaf samples were ground with a roller mill. After each series of 56 samples, leaves from plants infected by the G-bis variant were incorporated and ground similarly. The samples ground right after these leaves were used as contamination controls and tested for the presence of sequences of the G-bis variant. Twenty-three such contamination controls were analyzed in total. Illumina MiSeq sequencing revealed that no sequences belonging to the G-bis variant were detected in all contamination controls, indicating that no cross-contamination occurred among the samples during the crushing step. Additionally, to control for contaminations during RT-PCR, 8 negative controls (*i.e.* one per primer pair), where the RNA template was replaced with water, were added on each PCR plate. After Illumina MiSeq sequencing, the number of reads detected in the negative controls of the former set of thirteen PCR plates ranged from 0 to 100, with a mean number (± standard deviation) of 40 ± 24. As mentioned above, the sequencing results confirmed the occurrence of high contamination levels on four plates in the second set of six PCR plates, with 1449 ± 1917 reads in the negative controls. Consequently, all DH lines corresponding to these four PCR plates were removed from the final dataset. Finally, in total, sequences corresponding to 89 DH lines were kept and analyzed further. The number of reads per sample of these DH lines ranged from 201 to 11052, with a mean number of 4919 ± 1747.

From these reads, counts of sequences corresponding to the five inoculated PVY variants in each individual plant were obtained as in Rousseau et al. (2017). The initial inoculum was also sequenced and the following frequencies were obtained for the five PVY variants: G (22%), N (15%), K (17%), GK (18%) and KN (28%).

4.3 Inference of virus N_e and s

Estimation of the strength of genetic drift (N_e) and selection (s) acting on the composite PVY population was performed with Rousseau et al.'s (2017) model by using PVY variant counts obtained from HTS data for each plant as input. Outputs of the model are the relative growth rates r_i of each PVY variant i (G, N, K, GK and KN) in the population and the variance effective population size N_e of the whole PVY population. The selection coefficient between two variants corresponds to the difference in their relative growth rates r_1 and r_2. To summarize the selection effect s exerted by a pepper DH line on the composite PVY population, we used the standard deviation of the r_i values of these five variants. Similar, but less robust (*i.e.* lower h^2 estimates), results were obtained by considering that s equals the difference between the highest and the lowest r_i values among the five PVY variants for each DH line.

4.4 Statistical analyses of the links between variables related to the evolution of PVY populations

All statistical analyses were handled with the R software version 3.0.2 (R Core Team, 2013; (http://www.r-project.org/). We used general linear models (GLM) to study the effects of N_e, s, N_e^{inoc} and VA on the response variable RB. RB was considered as a binary variable representing the occurrence (1) or absence (0) of infection in a total of 60 individual plants per DH line (Quenouille et al. 2014). These infections resulted from the fixation in the PVY population of a single mutation in the VPg cistron that conferred the capacity to infect *pvr2*^{3}-carrying plants. As a consequence, a binomial distribution was used in the GLMs for RB. All the explanatory variables (N_e, s, N_e^{inoc} and VA) and their pairwise interactions were included in the full GLM and stepwise model selection was performed to eliminate variables. The Akaike's information criterion (AIC) was used to select the best-fitting models (Akaike 1974). The GLM analyses were done with the R packages 'lme4' and 'MASS'. The variance inflation factor (VIF) was assessed with the R package 'car'. Additionally, conditional inference regression trees were realized using the method 'ctree' implemented in the package 'party'. Such trees allow to explore the effects of the most significant explanatory variables on RB. Regression trees were computed with a minimum number of 20 DH

lines in each terminal 'leaf' of the tree and default setting for other parameters to describe the conditional distribution of RB as a function of the four explanatory variables N_e, s, N_e^{inoc} and VA.

Funding information

L. Tamisier's PhD was supported by the BAP (Biologie et Amélioration des Plantes) department and SMaCH (Sustainable Management of Crop Health) metaprogramme of INRA and by the Région Provence-Alpes-Côte d'Azur (PACA). The experimental work was supported by the SMaCH metaprogramme.

Acknowledgements

The authors would like to thank AM Sage-Palloix and G Nemouchi for providing pepper genetic resources, and N Truglio's team for technical assistance. We also thank HM Clause for its support.

5 Supplementary material

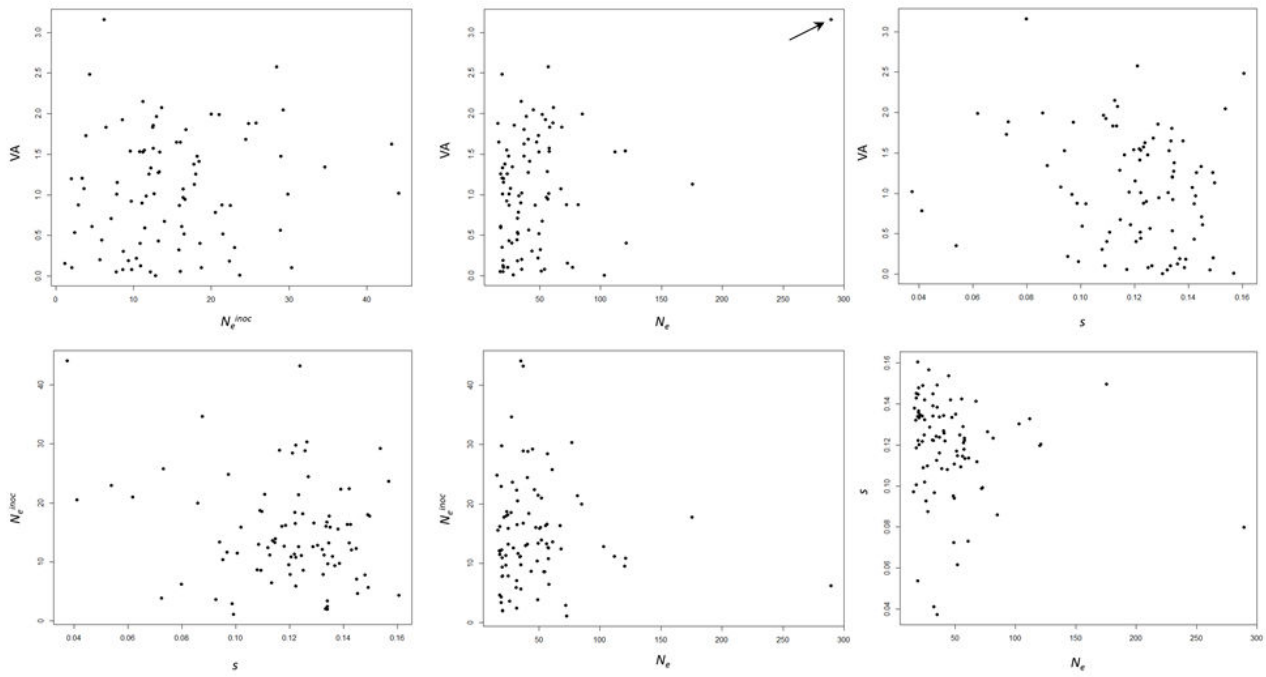


Figure S1 : Pairwise relationships between the four explanatory variables related to PVY evolution: VA (within-plant virus accumulation), s (selection intensity on the virus population), N_e (virus genetic drift during plant infection) and N_e^{inoc} (genetic drift at virus inoculation). The arrow indicates the outlier for N_e.

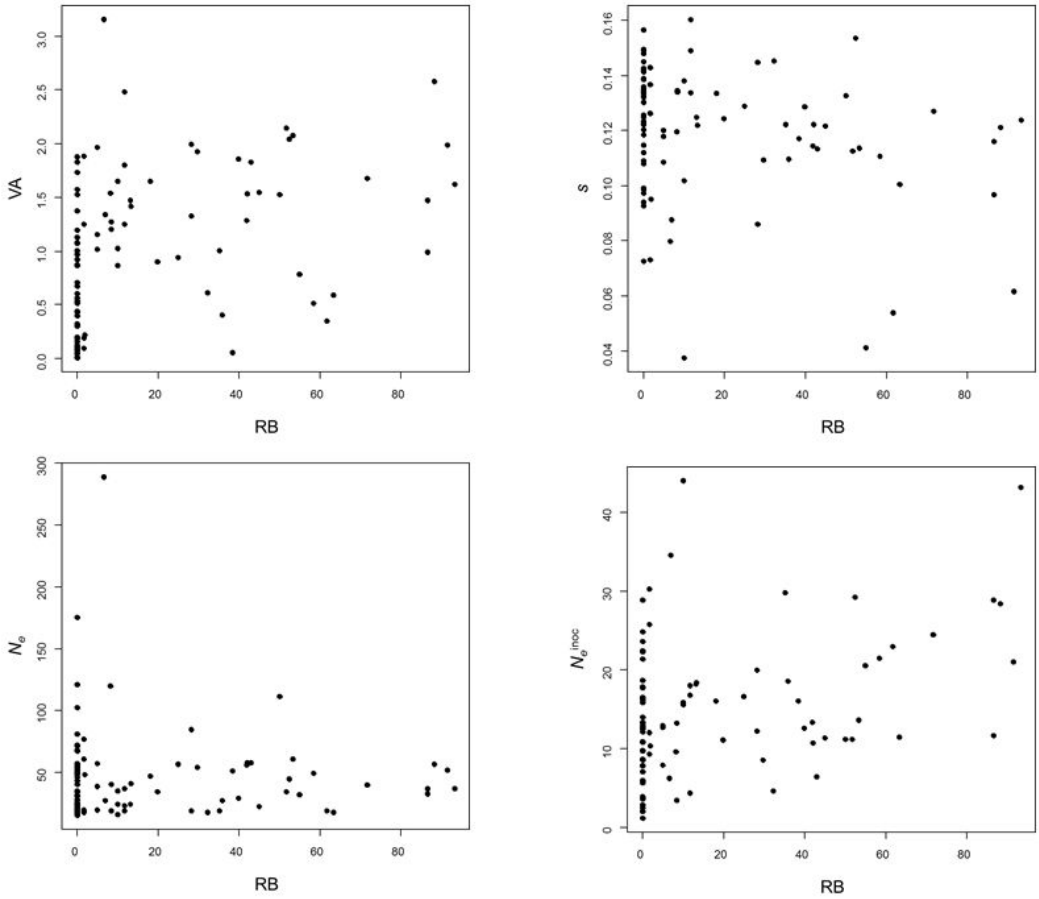


Figure S2 : Pairwise relationships between the response variable RB (frequency of breakdown of the *pvr2*^{3}-mediated resistance) and each of four explanatory variables related to PVY evolution: VA (within-plant virus accumulation), s (selection intensity on the virus population), N_e (virus genetic drift during plant infection) and N_e^{inoc} (genetic drift at virus inoculation).

References

Abel, S., Abel zur Wiesch, P., Davis, B. M. and Waldor, M. K. (2015). Analysis of Bottlenecks in Experimental Models of Infection. *PLoS Pathog.* **11**, e1004823.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723.

Ayme, V., Souche, S., Caranta, C., Jacquemond, M., Chadœuf, J., Palloix, A. and Moury, B. (2006). Different mutations in the genome-linked protein VPg of *Potato virus Y* confer virulence on the *pvr2*^{3} resistance in pepper. *Mol. Plant. Microbe Interact.* **19**, 557–563.

Barchi, L., Lefebvre, V., Sage-Palloix, A.-M., Lanteri, S. and Palloix, A. (2009). QTL analysis of plant development and fruit traits in pepper and performance of selective phenotyping. *Theor. Appl. Genet.* **118**, 1157–1171.

Brun, H., Chèvre, A.-M., Fitt, B. D., Powers, S., Besnard, A.-L., Ermel, M., Huteau, V., Marquer, B., Eber, F., Renard, M., et al. (2010). Quantitative resistance increases the durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. *New Phytol.* **185**, 285–299.

Bustin, S. A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* **25**, 169–193.

Charlesworth, B. (2009). Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205.

da Silva, J. and Wyatt, S. (2014). Fitness valleys constrain HIV-1's adaptation to its secondary chemokine coreceptor. *J. Evol. Biol.* **27**, 604–615.

de la Iglesia, F., Martínez, F., Hillung, J., Cuevas, J. M., Gerrish, P. J., Daròs, J.-A. and Elena, S. F. (2012). Luria-Delbrück estimation of *Turnip mosaic virus* mutation rate in vivo. *J. Virol.* **86**, 3386–3388.

Díaz, J. A., Nieto, C., Moriones, E., Truniger, V. and Aranda, M. A. (2004). Molecular characterization of a *Melon necrotic spot virus* strain that overcomes the resistance in melon and nonhost plants. *Mol. Plant. Microbe Interact.* **17**, 668–675.

Ewens, W. J. (2004). Mathematical population genetics. I. Theoretical introduction. Interdisciplinary applied mathematics, 27.

Fabre, F., Bruchou, C., Palloix, A. and Moury, B. (2009). Key determinants of resistance durability to plant viruses: Insights from a model linking within- and between-host dynamics. *Virus Res.* **141**, 140–149.

Fabre, F., Rousseau, E., Mailleret, L. and Moury, B. (2012b). Durable strategies to deploy plant resistance in agricultural landscapes. *New Phytol.* **193**, 1064–1075.

Fabre, F., Montarry, J., Coville, J., Senoussi, R., Simon, V. and Moury, B. (2012a). Modelling the Evolutionary Dynamics of Viruses within Their Hosts: A Case Study Using High-Throughput Sequencing. *PLoS Pathog.* **8**, e1002654.

Fabre, F., Rousseau, E., Mailleret, L. and Moury, B. (2015). Epidemiological and evolutionary management of plant resistance: optimizing the deployment of cultivar mixtures in time and space in agricultural landscapes. *Evol. Appl.* **8**, 919–932.

Fournet, S., Kerlan, M. C., Renault, L., Dantec, J. P., Rouaux, C. and Montarry, J. (2013). Selection of nematodes by resistant plants has implications for local adaptation and cross-virulence: Local adaptation and cross-virulence in *Globodera pallida*. *Plant Pathol.* **62**, 184–193.

Gago, S., Elena, S. F., Flores, R. and Sanjuán, R. (2009). Extremely high mutation rate of a hammerhead viroid. *Science* **323**, 1308–1308.

García-Arenal, F. and McDonald, B. A. (2003). An analysis of the durability of resistance to plant viruses. *Phytopathology* **93**, 941–952.

Gómez, P., Rodríguez-Hernández, A., Moury, B. and Aranda, M. (2009). Genetic resistance for the sustainable control of plant virus diseases: breeding, mechanisms and durability. *Eur. J. Plant Pathol.* **125**, 1–22.

Gutiérrez, S., Yvon, M., Thébaud, G., Monsion, B., Michalakis, Y. and Blanc, S. (2010). Dynamics of the multiplicity of cellular infection in a plant virus. *PLoS Pathog.* **6**, e1001113.

Gutiérrez, S., Michalakis, Y. and Blanc, S. (2012). Virus population bottlenecks during within-host progression and host-to-host transmission. *Curr. Opin. Virol.* **2**, 546–555.

Harrison, B. D. (2002). Virus variation in relation to resistance-breaking in plants. *Euphytica* **124**, 181–192.

Iwasa, Y., Michor, F. and Nowak, M. A. (2004). Stochastic tunnels in evolutionary dynamics. *Genetics* **166**, 1571–1579.

Janzac, B., Fabre, F., Palloix, A. and Moury, B. (2009). Constraints on evolution of virus avirulence factors predict the durability of corresponding plant resistances. *Mol. Plant Pathol.* **10**, 599–610.

Khelifa, M., Massé, D., Blanc, S. and Drucker, M. (2010). Evaluation of the minimal replication time of *Cauliflower mosaic virus* in different hosts. *Virology* **396**, 238–245.

Kimura, M. (1970). Stochastic processes in population genetics, with special reference to distribution of gene frequencies and probability of gene fixation. In *Mathematical topics in population genetics*, pp. 178–209. Springer.

Kimura, M. and Crow, J. F. (1963). The measurement of effective population number. *Evolution* **17**, 279–288.

Lafforgue, G., Tromas, N., Elena, S. F. and Zwart, M. P. (2012). Dynamics of the establishment of systemic potyvirus infection: independent yet cumulative action of primary infection sites. *J. Virol.* **86**, 12912–12922.

Lande, R. and Barrowclough, G. F. (1987). Effective population size, genetic variation, and their use in population management. *Viable Popul. Conserv.* **87**, 124.

Lanfear, R., Kokko, H. and Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends Ecol. Evol.* **29**, 33–41.

Lannou, C. (2012). Variation and Selection of Quantitative Traits in Plant Pathogens. *Annu. Rev. Phytopathol.* **50**, 319–338.

Lê Van, A., Caffier, V., Lasserre-Zuber, P., Chauveau, A., Brunel, D., Le Cam, B. and Durel, C.-E. (2013). Differential selection pressures exerted by host resistance quantitative trait loci on a pathogen population: a case study in an apple × *Venturia inaequalis* pathosystem. *New Phytol.* **197**, 899–908.

Malpica, J. M., Fraile, A., Moreno, I., Obies, C. I., Drake, J. W. and García-Arenal, F. (2002). The rate and character of spontaneous mutation in an RNA virus. *Genetics* **162**, 1505–1511.

Martínez, F., Sardanyés, J., Elena, S. F. and Daròs, J.-A. (2011). Dynamics of a plant RNA virus intracellular accumulation: stamping machine vs. geometric replication. *Genetics* **188**, 637–646.

McDonald, B. A. and Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* **40**, 349–379.

Miras, M., Sempere, R. N., Kraft, J. J., Miller, W. A., Aranda, M. A. and Truniger, V. (2014). Interfamilial recombination between viruses led to acquisition of a novel translation-enhancing RNA element that allows resistance breaking. *New Phytol.* **202**, 233–246.

Montarry, J., Doumayrou, J., Simon, V. and Moury, B. (2011). Genetic background matters: a plant-virus gene-for-gene interaction is strongly influenced by genetic contexts: Viral genetic background matters. *Mol. Plant Pathol.* **12**, 911–920.

Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C., Palloix, A. and Jacquemond, M. (2004). Mutations in *Potato virus Y* genome-linked protein determine virulence toward recessive resistances in *Capsicum annuum* and Lycopersicon hirsutum. *Mol. Plant. Microbe Interact.* **17**, 322–329.

Moury, B., Fereres, A., García-Arenal, F. and Lecoq, H. (2011). Sustainable management of plant resistance to viruses. In *Recent advances in plant virology*, pp. 219–336. Norwich, UK.

Palloix, A., Ayme, V. and Moury, B. (2009). Durability of plant major resistance genes to pathogens depends on the genetic background, experimental evidence and consequences for breeding strategies. *New Phytol.* **183**, 190–199.

Quenouille, J., Montarry, J., Palloix, A. and Moury, B. (2013). Farther, slower, stronger: how the plant genetic background protects a major resistance gene from breakdown: Mechanisms of polygenic resistance durability. *Mol. Plant Pathol.* **14**, 109–118.

Quenouille, J., Paulhiac, E., Moury, B. and Palloix, A. (2014). Quantitative trait loci from the host genetic background modulate the durability of a resistance gene: a rational basis for sustainable resistance breeding in plants. *Heredity* **112**, 579–587.

Quenouille, J., Saint-Felix, L., Moury, B. and Palloix, A. (2015). Diversity of genetic backgrounds modulating the durability of a major resistance gene. Analysis of a core collection of pepper landraces resistant to *Potato virus Y*. *Mol. Plant Pathol.* **17**, 296–302.

Råberg, L., Sim, D. and Read, A. F. (2007). Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. *Science* **318**, 812–814.

Restif, O. and Koella, J. C. (2004). Concurrent evolution of resistance and tolerance to pathogens. *Am. Nat.* **164**, E90–E102.

Rodrigo, G., Zwart, M. P. and Elena, S. F. (2014). Onset of virus systemic infection in plants is determined by speed of cell-to-cell movement and number of primary infection foci. *J. R. Soc. Interface* **11**, 20140555.

Rousseau, E., Fabre, F., Senoussi, R., Mailleret, L., Palloix, A., Simon, V., Valière, S., Moury, B. and Grognard, F. (In press). Estimating virus effective population size and selection without neutral markers. *PLoS Pathog.*

Tamisier, L., Rousseau, E., Barraillé, S., Nemouchi, G., Szadkowski, M., Mailleret, L., Grognard, F., Fabre, F., Moury, B. and Palloix, A. (2017). Quantitative trait loci in pepper control the effective population size of two RNA viruses at inoculation. *J. Gen. Virol.* **98**, 1923–1931.

Tromas, N. and Elena, S. F. (2010). The rate and spectrum of spontaneous mutations in a plant RNA virus. *Genetics* **185**, 983–989.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.

Zhan, J., Thrall, P. H., Papaïx, J., Xie, L. and Burdon, J. J. (2015). Playing on a Pathogen's Weakness: Using Evolution to Guide Sustainable Plant Disease Control Strategies. *Annu. Rev. Phytopathol.* **53**, 19–43.

Zwart, M. P. and Elena, S. F. (2015). Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annu. Rev. Virol.* **2**, 161–179.

Zwart, M. P., Daròs, J.-A. and Elena, S. F. (2011). One Is Enough: In Vivo Effective Population Size Is Dose-Dependent for a Plant RNA Virus. *PLoS Pathog.* **7**, e1002122.

Zwart, M. P., Daros, J.-A. and Elena, S. F. (2012). Effects of Potyvirus Effective Population Size in Inoculated Leaves on Viral Accumulation and the Onset of Symptoms. *J. Virol.* **86**, 9737–9747.

Chapitre 4

Taking control of virus adaptation by choosing host plant genotype

Résumé de l'article :

Objectifs : L'objectif principal était d'étudier la capacité d'adaptation du PVY par mutation lorsqu'il est confronté à des lignées de piment imposant différents niveaux de dérive génétique, de sélection et d'accumulation virale. Un objectif complémentaire était d'identifier quel génotype de plante pouvait contraindre au mieux l'évolution du virus.

Stratégie : De l'évolution expérimentale de populations de PVY a été réalisée. Durant sept mois, soixante-quatre populations virales ont été passées en série sur six lignées de piment contrastées en termes de dérive génétique, de sélection et d'accumulation virale. Après chaque passage, la région codant pour la protéine VPg du virus a été séquencée chez chaque population virale pour identifier d'éventuelles mutations. La valeur adaptative et le niveau d'agressivité des populations finales ont également été mesurés et comparés à ceux des clones viraux de départ.

Résultats : Des trajectoires évolutives très contrastées ont été obtenues en fonction des lignées de piment sur lesquelles ont évolué les populations virales, allant de l'extinction de la population à l'adaptation via des mutations parallèles. Cette étude a mis en évidence que les plantes imposant à la fois un faible niveau d'accumulation virale et une forte dérive génétique empêchent les populations virales de s'adapter durant la totalité de l'expérience d'évolution et peuvent même entraîner des extinctions, potentiellement par l'action du cliquet de Muller.

Taking control of virus adaptation by choosing host plant genotype

Tamisier, L. (1,2), Nemouchi, G. (1), Szadkowski, M. (1), Girardot, G. (2), Fabre, F. (3), Palloix, A. (1), Moury, B. (2).

(1) GAFL, INRA, 84140 Montfavet, France

(2) Pathologie Végétale, INRA, 84140 Montfavet, France

(3) UMR 1065 Santé et Agroécologie du Vignoble, INRA, Villenave d'Ornon, France

Keywords: Experimental evolution, genetic drift, selection, evolutionary forces, *Capsicum annuum*, *Potato virus Y*, virus emergence.

Abstract

Several authors have proposed to fight against the emergence of virulent pathogens by manipulating the evolutionary forces imposed by the host on the pathogen population, like genetic drift, selection and mutation. However, no experimental proof of this approach in a plant-pathogen system is available. We have analyzed the impact of pepper (*Capsicum annuum*, family Solanaceae) DH lines carrying the same major resistance gene but contrasted genetic backgrounds on the evolution of *Potato virus Y* (PVY; genus *Potyvirus*, family Potyviridae) populations. The lines were chosen for their different level of selection (*s*) and genetic drift (N_e) they exerted on the virus population, as well as the initial virus accumulation level of the PVY clones in the lines (VA). We have performed an experimental evolution by serially passaging 64 PVY populations every month on 6 contrasted pepper DH lines during 7 months. The VPg cistron of the viral populations, where most adaptive mutations are likely to occur, has been sequenced at the end of every passage and the final accumulation levels of the viral lineages have been assessed and compare to the initial inocula. We showed that highly contrasted evolutionary trajectories have occurred, including virus extinctions and fitness gain linked to parallel nonsynonymous mutations in the VPg cistron. The evolutionary trajectories were particularly well explained by the evolutionary forces imposed by the host. More specifically, N_e, VA and their synergistic interaction played a major role in the fate of PVY populations: when N_e was low, the virus fitness gain increased with VA, whereas when N_e

was high, the fitness gain decreased with VA. The synergistic effect of N_e and VA is of particular importance since it shows that combining a high resistance efficiency (low VA) and a low N_e is the best solution to increase the resistance durability potential, *i.e.* to avoid virus adaptation on the long term. These results demonstrate that N_e and VA are two highly variable and heritable traits of plant genotypes that can be selected by breeders to avoid the emergence of adapted virus variants in spite of the strong selective pressure exerted by a host major resistance gene.

1 Introduction

The evolution of plant RNA viruses is shaped by complex interactions between genetic drift, selection and mutation. Genetic drift is a stochastic force inducing random variations of viral variants frequencies from generation to generation. Its strength is defined by the effective population size (N_e), a parameter corresponding to the number of viral variants that effectively transmit their genes to the next generation (Elena and Sanjuán 2007): the lower N_e, the stronger the genetic drift. Since N_e is the key parameter to understand the effect of genetic drift on virus evolution, a lot of studies have focused on the assessment of N_e of plant RNA viruses. Most of these studies found very small values, usually much lower than the census size of the population (Zwart and Elena 2015). The main reason is that the viral population size varies considerably during the infection process due to the action of repetitive bottlenecks, which significantly affect the value of N_e. For instance, the inoculation into the plant constitutes a very narrow bottleneck. Using aphid transmission, an average of 0.5 to 3.2 infectious particles of *Potato virus Y* (PVY; genus *Potyvirus*, family Potyviridae) per aphid was assessed to be inoculated on *Capsicum annuum* (Moury et al. 2007) and similar values were estimated for *Cucumber mosaic virus* (CMV; genus *Cucumovirus*, family Bromoviridae) (Betancourt et al. 2008). Likewise, using mechanical inoculation, N_e of *Tobacco etch virus* (TEV; genus *Potyvirus*) was estimated between 1 and 50 viral particles in *Nicotiana tabacum* (Zwart et al. 2011). Small N_e have also been found during later infection steps, like leaf infection (Tromas et al. 2014) and plant colonization (French and Stenger 2003; Sacristán et al. 2003). Similar N_e have been estimated for animal viruses. For example, for *Venezuelan equine encephalitis virus* (VEEV; genus *Alphavirus*, family Togaviridae), bottleneck size was 1.1 on average during infection of individual midgut cells of *Aedes taeniorhynchus* mosquito vectors (Smith et al. 2008; Zwart and Elena 2015). Stochastic effects due to genetic drift play therefore a major role during virus evolution. Another key parameter determining the rate of

evolution of RNA viruses is selection. Contrary to genetic drift, selection is a deterministic force that increases the frequency of the fittest virus variants over time. Usually, it is measured with the selection coefficient *s*, defined as the difference in fitness between two virus variants. The action of selection overcomes the action of genetic drift when N_e is large. Even if bottlenecks are frequent and narrow during the infection cycle, the total number of viral particles in a plant can reach extremely high values at some stages of the infection. For example, between 10^{11} and 10^{12} *Tobacco mosaic virus* (TMV; genus *Tobamovirus*, family *Virgaviridae*) particles can infect a single tobacco leaf (Gibbs et al. 2008). Therefore, selection is also expected to act significantly on virus evolution. While genetic drift or selection will lead to a loss of genetic diversity with time, mutation, a third force of particular importance when considering virus evolution, will increase genetic diversity within the viral population. RNA viruses are known for their high mutation rate due to the lack of proofreading activity of their RNA-dependent RNA polymerases. Their mutation rate is estimated to be between 10^{-4} and 10^{-6} mutation per nucleotide per generation, corresponding to a rough average of one mutation per genome per replication cycle (Sanjuan et al. 2010).

The fate of all mutations occurring in the population depends both on N_e and *s*. Rouzine et al. (2001) have defined several evolutionary regimes according to the value of those parameters. A genetic drift regime occurs when $N_e << 1/s$. In this case, genetic drift predominates over selection and the fluctuations of deleterious mutation frequencies will become similar to those of neutral mutations. Some deleterious mutations could be fixed and potentially cause a loss of fitness by accumulating in the population. A selection-drift regime occurs when $1/s < N_e < 1/\mu$, where μ stands for the mutation rate. Under this regime, the evolution will be driven by either stochastic or deterministic factors, according to the parameter values. Finally, a selection regime occurs when $N_e >> 1/\mu$. In this case, selection prevails over genetic drift. As a consequence, to be able to study RNA virus evolution, we need to estimate both N_e and *s*. Using the pepper (*Capsicum annuum*, family Solanaceae) - PVY pathosystem, Rousseau et al (2017) have developed a mechanistic-statistical model estimating simultaneously those two parameters. The novelties of this model are that it applies (i) to variants that differ only at non-neutral markers and (ii) in a wide range of N_e and *s* values, particularly suited to viruses. With this model, they have characterized pepper doubled-haploid (DH) lines according to the intensity of genetic drift and selection that they imposed on a PVY population composed of 5 variants. To do so, they compared the changes in frequencies of the 5 variants composing the population in the inoculum and in infected plants at different infection time points up to one month after inoculation. Importantly, N_e and

s values were highly contrasted among pepper lines and both N_e and *s* showed a high heritability (h^2 = 0.63 and 0.94, respectively) from the plant side. These DH lines can therefore be used as a tool to test hypotheses on virus adaptability under different evolutionary constraints. For this purpose, we have performed an experimental evolution by serially passaging PVY populations every month on 6 contrasted pepper DH lines during 7 months. We have selected the 6 DH lines according to 3 parameters. First, the DH lines showed contrasted values of N_e and *s* that fall in the selection-drift regime. Then, we expected that PVY adaptability could depend on the initial adaptation of the virus to the DH line, *i.e.* its level of accumulation (VA, initial virus accumulation) in plants. Therefore, we selected both pepper lines and PVY variants with different levels of initial accumulation to perform the experimental evolution.

The goal of this experimental evolution was to answer two major questions. Firstly, can the virus take different evolutionary trajectories when it evolves in different plant lines of the same species? A lot of studies have involved experimental evolution to test the adaptation ability of viruses or other microorganisms in hosts belonging to different species (Bedhomme et al. 2012; Guidot et al. 2014; Hillung et al. 2014). In contrast, in our experiment, the virus evolved in closely-related pepper lines sharing the same major-effect resistance gene and differing only in their genetic background (see Materials and Methods). Even if the lines were selected for their contrasted effects on virus variant frequencies, we do not know if such effects will significantly impact virus adaptation through *de novo* mutations during 7 months of evolution. Secondly, if different evolutionary trajectories occur, can we explain these differences by taking into account the evolutionary forces imposed by the host on virus populations? If we are able to explain the evolution of the PVY populations according to the values of N_e, *s* and/or VA, it will prove that our measures of these variables are efficient to predict the evolutionary trajectories of the viruses. From an agronomic point of view, selecting varieties based on these values could help us to take control of virus evolution by manipulating the evolutionary forces acting on virus populations within the plant.

2 Results

2.1 Divergent evolutionary trajectories among PVY lineages: extinction, *status quo* or parallel fixation of mutations

In order to study the effects of different evolutionary constraints corresponding to different host plant environments on virus evolution, we conducted an experimental evolution of PVY on closely-related pepper DH lines derived from the same F_1 hybrid between two parental lines. Sixty-four independent PVY evolutionary lineages have been obtained by serial passages in 6 DH lines contrasted for N_e, s and VA (Figure 1, Table S1 and Materials and Methods).

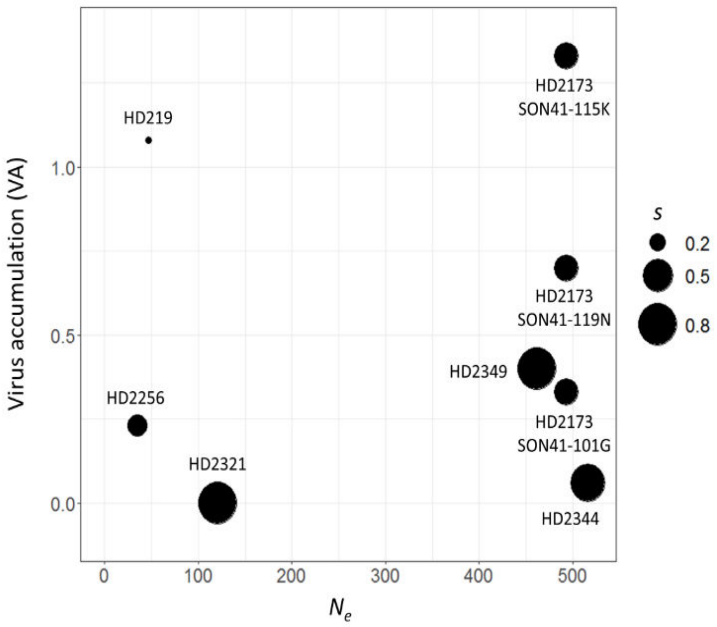


Figure 1: Features of the pepper DH line - initial PVY clone combinations used for the experimental evolution. The pepper lines HD219, HD2256, HD2321, HD2349 and HD2344 have only been inoculated with the PVY SON41-119N clone, while HD2173 has also been separately inoculated with the SON41-115K and SON41-101G clones. The 8 virus-host combinations are plotted according to three traits: (i) the selection coefficient (s) and (ii) the effective population size (N_e) characterizing the evolution of an artificial PVY population (Rousseau et al. 2017), and (iii) the initial virus accumulation (VA) of each clone.

The PVY SON41-119N cDNA clone has started the experimental evolution for all DH lines, whereas the SON41-101G and -115K clones have been passaged only on DH line HD2173 (Figure 2). After each passage, the VPg cistron of the PVY populations, where adaptive mutations are likely to occur, has been sequenced and single nucleotide polymorphisms (SNPs) have been detected by comparison with the sequence of the starting clones.

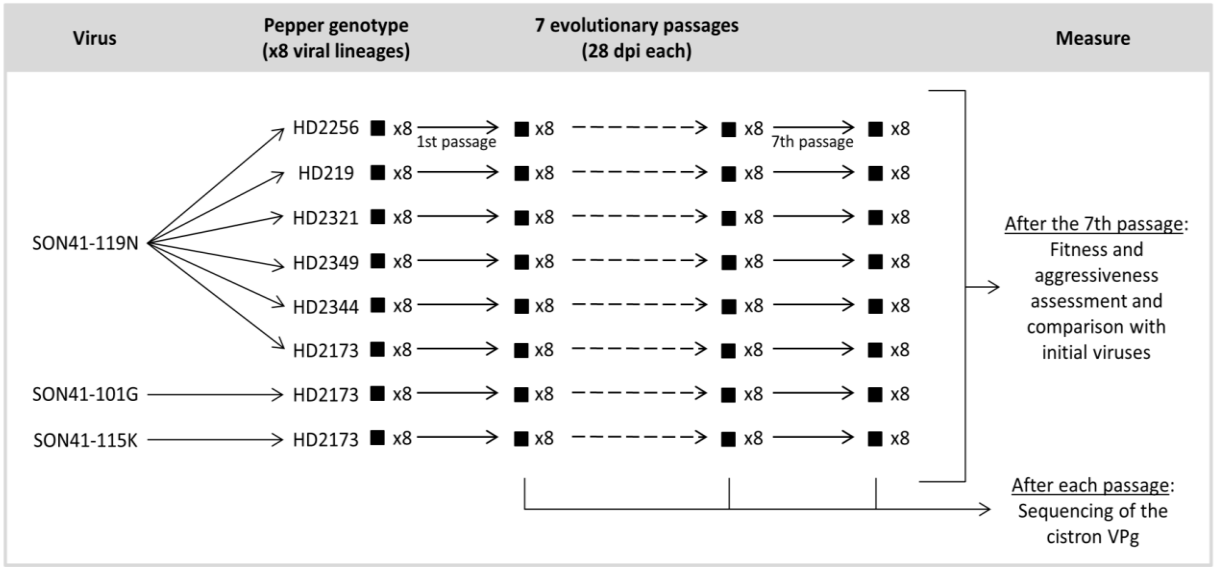


Figure 2: Experimental design of the study. Six doubled-haploid (DH) pepper lines have been inoculated with the SON41-119N clone of PVY. The HD2173 line has also been inoculated with the SON41-101G and the SON41-115K clones separately. For each virus-host combination, 8 evolutionary lineages of PVY have been maintained during 7 serial passages, the passages being performed every 28 days. The VPg coding region of all the evolutionary lineages was sequenced by the Sanger method after each passage. Eventually, the fitness and aggressiveness of the final lineages have been assessed.

Table 1 summarizes the evolutionary trajectories of the 64 PVY lineages, named L1 to L64, during the infection cycles and in each of the DH line – initial PVY clone combinations. At the end of the experiment, 9 lineages went to extinction, as no PVY was detected by RT-PCR and none of the 3 inoculated plants corresponding to the next cycle were infected. Three lineages of 8 became extinct in HD2256 and 6 lineages of 8 in HD2321. Then, 33 lineages did not show any mutation in the VPg cistron. Most of them correspond to pepper lines HD2256, HD219 and HD2173 with SON41-119N, plus SON41-115K for the latter. Finally, 25 lineages have shown at least one SNP. A total of 30 mutation events occurred in those lineages. Among these mutations, 27 were nonsynonymous, whereas only 3 were synonymous. All the SNPs were either fixed in the population as soon as their first detection or became fixed during the cycle following their first detection, with only two exceptions. These exceptions were the mutation 139V in HD219 (lineage L9), which has arisen during the 2nd cycle but was not fixed in the population until the 6th cycle, and the mutation 115K in HD2321 (lineage L22), which has arisen during the 3rd cycle but has never been fixed because the lineage went to extinction during the 4th cycle. Considering only nonsynonymous mutations, 10 different mutants were detected among the lineages, including 7 single mutants and 3 double mutants.

A total of 8 different mutations were observed (Table S2). Some of the nonsynonymous mutations have arisen several times independently (*i.e.* parallel mutations) among lineages.

Table 1: Appearance of amino acid substitutions for the 64 evolutionary lineages over 7 infection cycles. The selection coefficient (s), the effective population size (N_e) and the virus accumulation (VA) characterizing each DH line – PVY clone combination are expressed as a fraction of their maximum value. Each row represents a lineage, named from L1 to L64. Amino acid substitutions are represented by dots plotted according to their order or appearance and the colors associated to the dot distinguished the different substitutions. A grey square indicates the extinction of an evolutionary lineage, that is, the absence of PVY detection by RT-PCR performed with this lineage and the absence of symptoms on the three plants of the next passage inoculated with this lineage. The substitution in bracket was not fixed in the lineage.

DH line	DH line features	Initial PVY clone	Lineage	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Passage 6	Passage 7	Total
HD2256	s N_e VA 0–1	SON41 119N	L1–L8								
HD219	s N_e VA 0–1	SON41 119N	L9–L16								
HD2321	s N_e VA 0–1	SON41 119N	L17–L24			(●)					(●)
HD2349	s N_e VA 0–1	SON41 119N	L25–L32								
HD2344	s N_e VA 0–1	SON41 119N	L33–L40								
HD2173	s N_e VA 0–1	SON41 119N	L41–L48								
HD2173	s N_e VA 0–1	SON41 101G	L49–L56								
HD2173	s N_e VA 0–1	SON41 115K	L57–L64								

Amino acid substitutions: 101G 102K 109V 115K 115M 119N 139V

Regarding the 48 lineages deriving from the SON41-119N clone, the most frequent nonsynonymous mutation was 115M, which changes the threonine at position 115 of the VPg into a methionine. This substitution has occurred in 8 independent evolutionary lineages and 4 different DH lines. The second most frequent nonsynonymous mutation was 115K, which occurred at the same position and changes the threonine into a lysine. This substitution was detected in 5 other independent evolutionary lineages and 3 different DH lines. Parallel mutations also occurred among the 8 lineages coming from the SON41-101G clone. Indeed, 6 independent lineages showed the 119N mutation, which changes the aspartic acid at position 119 of the VPg by an asparagine. Finally, no substitutions occurred among the 8 lineages derived from the SON41-115K clone.

2.2 Significant increase in virus accumulation but little change in aggressiveness after experimental evolution

After 7 passages, the accumulation of the remaining virus populations was assessed on the DH lines they have been passaged on and compared to the initial PVY clones. On average, the mean virus accumulation of the final lineages has increased significantly in 4 DH lines: HD2321, HD2349 and HD2344 inoculated with SON41-119N, as well as HD2173 inoculated with SON41-101G (Kruskal-Wallis test, $p < 0.001$, Figure 3).

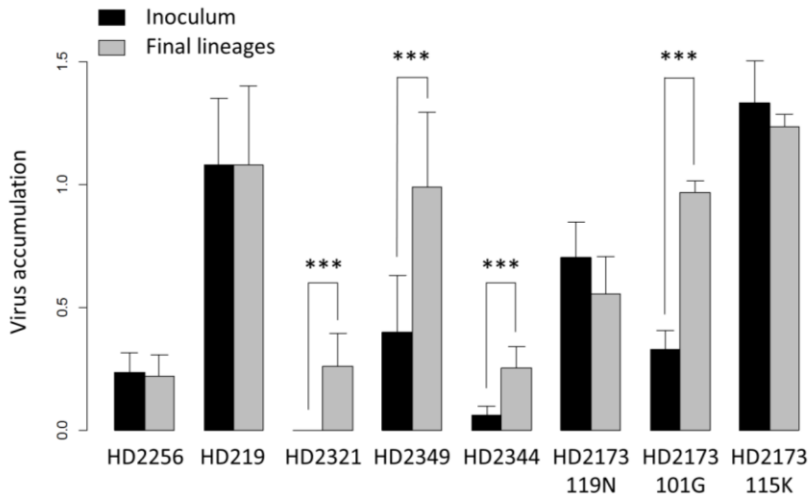
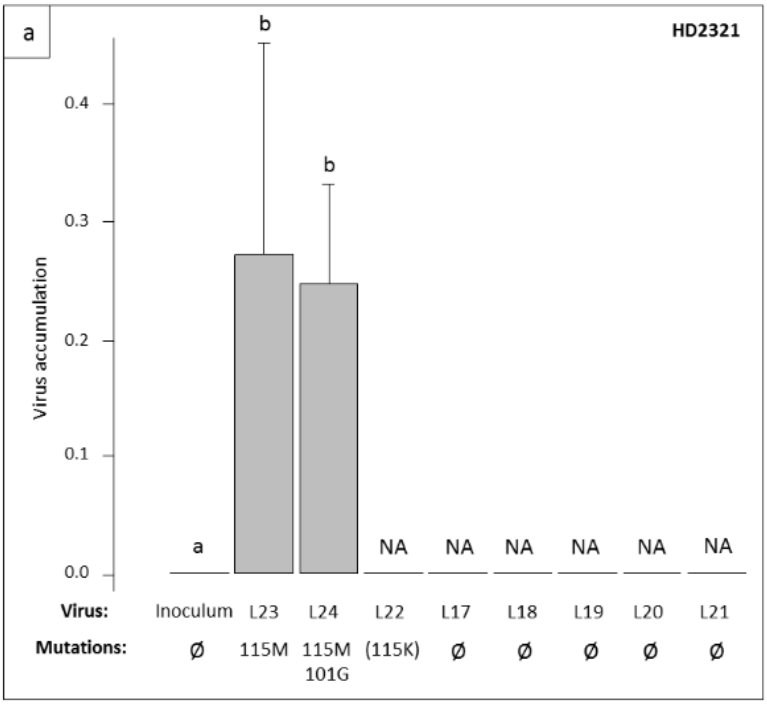
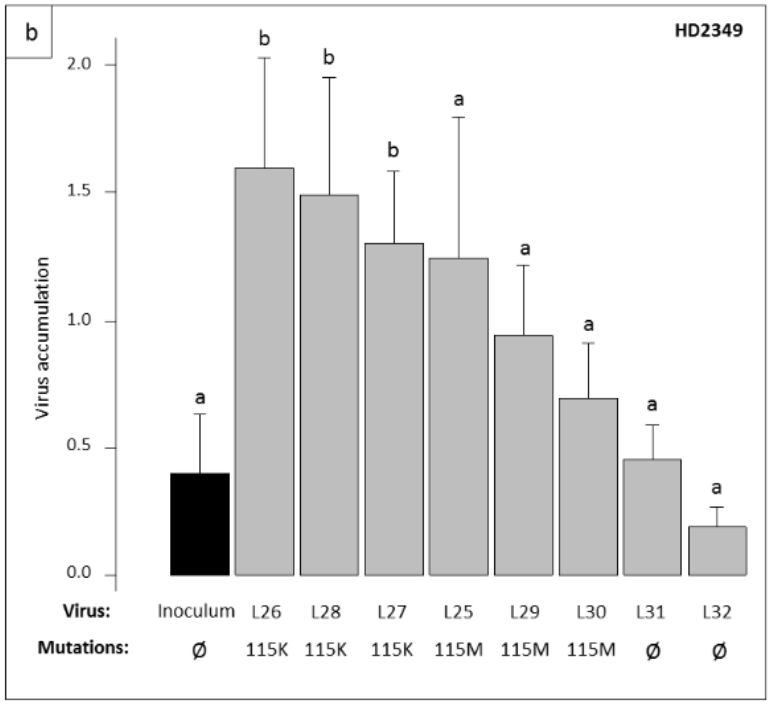


Figure 3: Virus accumulation of the initial inoculum (black) and mean virus accumulation of the final evolutionary lineages (grey) into the six DH lines. The SON41 viral clone used to start the infection is indicated for the HD2173 genotype. For each pepper genotype, stars represent significant differences between the inoculum and the mean virus accumulation of the lineages (Wilcoxon test, $p < 0.001$). Error bars represent standard error of the mean.

More specifically, the virus accumulation was significantly higher than the initial populations for the 2 remaining lineages in HD2321, for 3 of 8 lineages in HD2349, 5 of 8 lineages in HD2344 and all lineages in HD2173 (Dunnett test, Figure 4). This higher virus accumulation was almost always associated with the fixation of one or two nonsynonymous mutations in the VPg cistron of the related lineages. Fixation of 115K was associated with the highest increases in accumulation. Fixation of mutation 115M was also associated with significant increases in accumulation. Fixation of 119N in lineages that have been passaged on HD2173 inoculated with SON41-101G was also always associated with significant increases in accumulation.





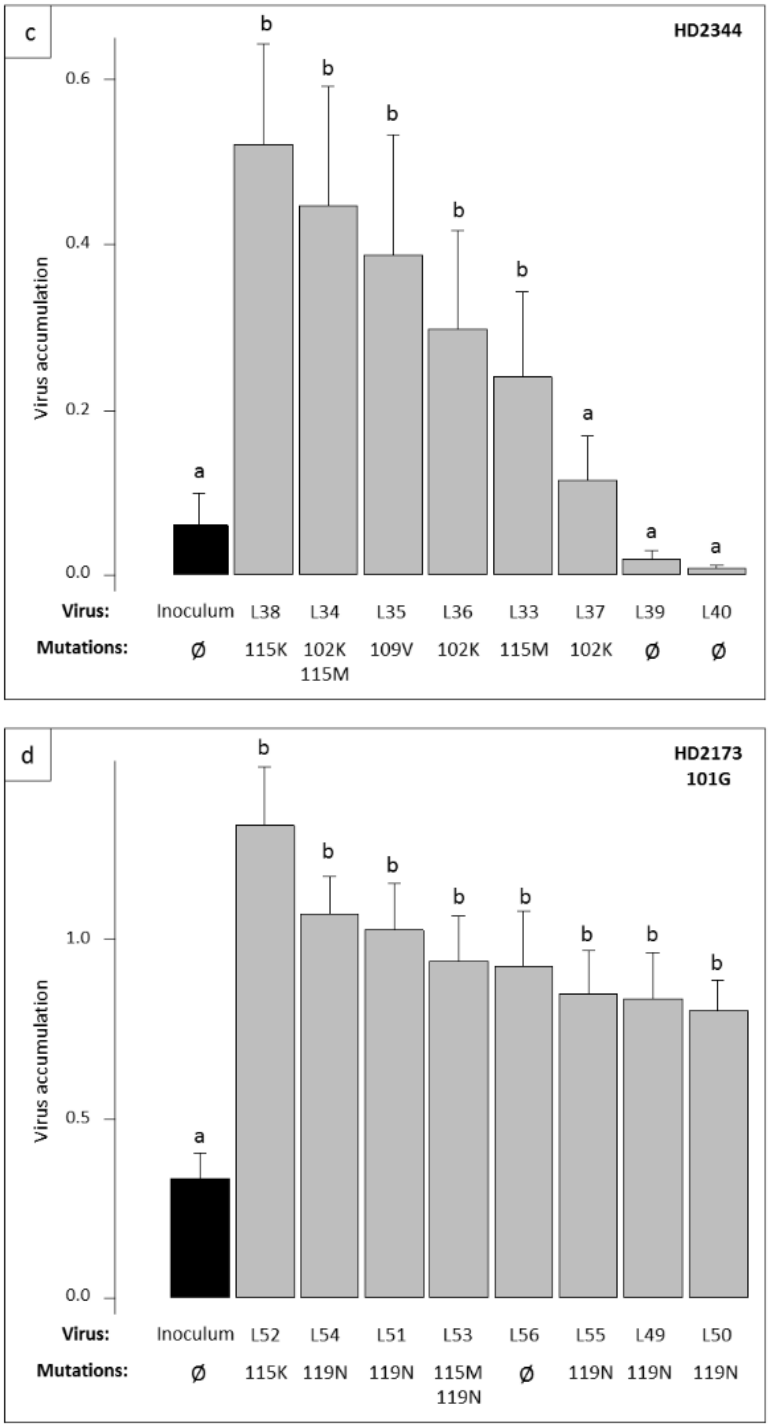


Figure 4: Virus accumulation of the initial inoculum (black) and the 8 final evolutionary lineages (grey) into four DH lines: HD2321 (a), HD2349 (b), HD2344 (c) and HD2173 inoculated with SON41-101G (d). The non-synonymous mutations in the VPg cistron of each evolutionary lineage or the absence of mutation (Ø) are indicated. The substitution in bracket was not fixed in the lineage. The letters a and b represent the different groups obtained after the comparison of each lineage to the inoculum using Dunnett test ($p < 0.05$). NA: not measured (extinct lineage). Error bars indicate standard error of the mean.

Further, in HD2349 and HD2344, the two lineages with the lowest virus accumulation were the only two lineages without any mutation. Note that the appearance of mutation 115K has

not prevented one lineage, L22, to become extinct in HD2321. We have also found one lineage, L56, showing an increase in fitness in HD2173 without having any fixed or detectable mutation in the VPg cistron. Regarding the least frequent mutants, the double mutants 115M-101G and 102K-115M were all associated with significant increases in accumulation. Mutation 102K was also associated with an increase in fitness, which was statistically significant in L36 but not in L37. Finally, mutation 109V was associated with a significant fitness gain in L35 while mutation 139V did not change the virus accumulation in L9.

The aggressiveness of final and initial PVY populations was also measured at the end of the experiment. Aggressiveness is defined here as the quantity of plant damages caused by viruses. The height and the fresh weight of plants infected with the different PVY populations or mock-inoculated have been compared 30 days post inoculation (dpi). Regarding the mean aggressiveness of lineages corresponding to the same DH line-PVY clone combination, there was no significant difference between the height of the plants infected with the initial and the final PVY populations (Figure S1A). For the plant fresh weight, only the HD219 line has shown a significantly higher plant fresh weight when inoculated with the final PVY populations compared to the initial one (Kruskal-Wallis test, $p < 0.05$, Figure S1B). Regarding the aggressiveness of each PVY population independently, only 2 lineages among the 55 remaining have shown significant differences compared to the starting clone. On HD2344, the lineage L38 has decreased significantly both the height and the weight of the plants compared to the initial PVY clone (Dunnett test, $p < 0.01$, Figure S2). This lineage was the only one carrying the mutation 115K in HD2344 and was also the lineage showing the highest increase of accumulation (Figure 4). Conversely, on HD219, the lineage L9 was significantly less aggressive than the initial clone, occasioning a lower decrease of plants height and weight (Dunnett test, $p < 0.05$, Figure S2). It was also the only lineage carrying the mutation 139V (Table 1). No change in virus accumulation was associated with this lineage.

2.3 Validation of the impact of the two most frequent *de novo* mutations on virus accumulation

Our goal was first to validate the role of the two most frequent *de novo* mutations, 115M and 115K, in the gain of fitness observed for the lineages carrying these mutations. Then, we wanted to test if these mutations also conferred a fitness gain in the DH lines where they were not observed. Competition experiments were performed to assess the relative fitness of 3 PVY variants: the initial SON41-119N clone and the SON41-115K-119N and SON41-115M-119N mutants. First, a 2:1 ratio mixture of SON41-119N and one of the two mutants (SON41-115K-119N or -115M-119N) was inoculated in 4 DH lines: HD219, HD2321, HD2344 and HD2349. Thirty days post inoculation, the relative proportions of each variant were estimated by sequencing. For the SON41-119N *vs.* -115K-119N competition, the SON41-115K-119N mutant was either predominant or alone in the plants for all the DH lines. The final viral population composition was always significantly different from the inoculum composition, in which SON41-119N was predominant (Fisher exact test, $p < 0.0025$, Figure S3A). Similar results were obtained for the SON41-119N *vs.* -115M-119N competition, with the proportion of SON41-115M-119N becoming significantly higher in the plants than in the inoculum for all the DH lines (Fisher exact test, $p < 0.0025$, Figure S3B). Thus, SON41-115K-119N and -115M-119N mutants appeared to be more competitive than 119N in these 4 DH lines. Then, a 2:1 ratio mixture of SON41-115M-119N and -115K-119N, respectively, was inoculated in the same DH lines. At 30 dpi, the SON41-115K-119N mutant was predominant or alone in nearly all the plants. Again, all the DH lines showed a viral composition significantly different from the inoculum one, where SON41-115M-119N predominated (Fisher exact test, $p < 0.0025$, Figure S3C). The SON41-115K-119N mutant was therefore more competitive than the SON41-115M-119N mutant in the 4 DH lines analyzed.

2.4 Contrasted effects of selection, genetic drift and initial accumulation on virus evolution

At the end of the experimental evolution, two variables representing virus evolution have been assessed for each PVY lineage: (i) the change in virus accumulation between the initial and final PVY populations and (ii) the occurrence of fixed nonsynonymous mutations. The speed of fixation of nonsynonymous mutations has also been studied and has provided similar results as the occurrence of fixed nonsynonymous mutations (data not shown). We have first

studied the effect of the pepper DH lines used to perform the experimental evolution on these two variables. The pepper genotype had a significant effect on the change in virus accumulation (one-way ANOVA, $F = 6.86$, $p < 0.001$) and on the occurrence of mutations (Pearson's chi-squared test, $\chi^2 = 16.71$, $p < 0.01$).

To investigate more deeply the effect of pepper genotypes on virus evolution, we performed generalized linear models (GLM). Our goal was to explain virus evolution according to three DH line features: the level of selection imposed by the line (represented by the selection coefficient s), the level of genetic drift imposed by the line (represented by the effective population size N_e) and the initial virus accumulation of the PVY clone in the line (VA). We carried out stepwise selection and ended up with the variables and their pairwise interactions that best explain the dependent variables (Table 2).

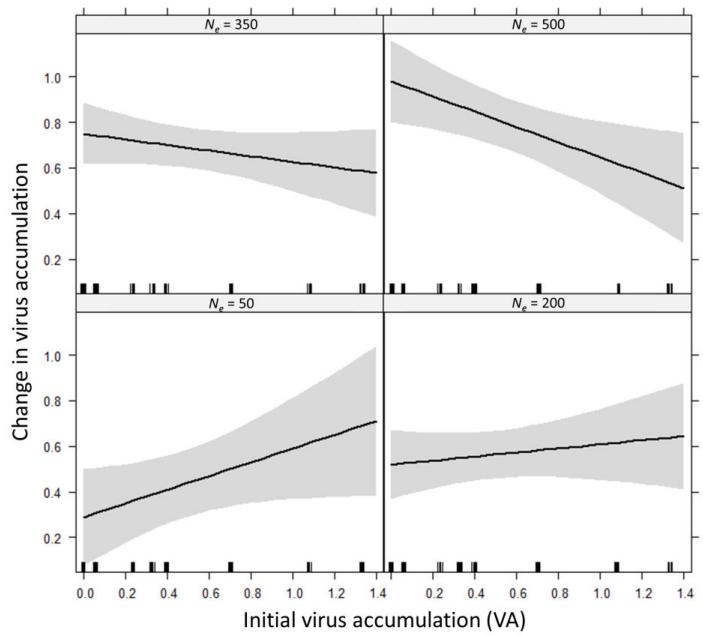
Table 2: GLM analyses of two variables linked to the virus ability to adapt. The explanatory variables are the selection coefficient (s) and the effective population size (N_e) characterizing the evolution of an artificial PVY population in each DH line, and the initial virus accumulation (VA) characterizing each DH line – PVY clone combination used in the experimental evolution. Their pairwise interactions were also included in the models. Only the variables kept after performing a stepwise selection procedure and the selection of the best-fitting models using the Akaike's information criterion (AIC) are shown. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

Virus features	Explanatory variables	Estimate	Std. Error	Wald test	*p-value*
Change in virus accumulation	Intercept	0.212	0.119	1.773	0.081
	N_e	0.001	3.21e-04	4.792	1.130e-05 ***
	VA	0.372	0.179	2.079	0.042 *
	VA × N_e	-0.001	4.63e-04	-3.05	0.003 **
Mutation occurrence	Intercept	-433.239	206.325	-2.1	0.036 *
	N_e	0.903	0.43	2.105	0.035 *
	s	537.551	256.7444	2.094	0.036 *
	VA	326.623	156.28	2.09	0.036 *
	VA × N_e	-0.683	0.324	-2.111	0.035 *
	s × N_e	-1.13	0.541	-2.09	0.037 *

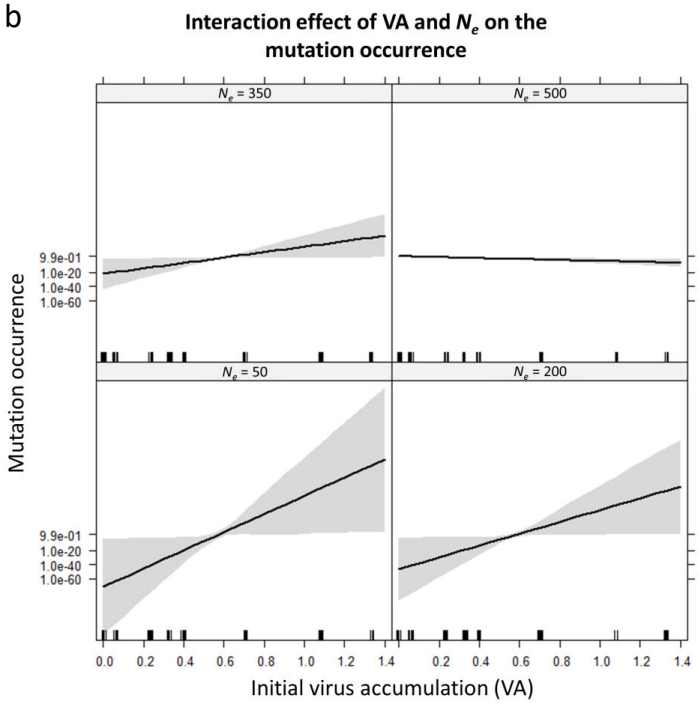
The explanatory variables VA and N_e only had a significant positive effect on the change in virus accumulation. The change in virus accumulation was also significantly affected by VA × N_e. For low values of N_e (50 and 200), the virus accumulation increased with VA, while

for higher values of N_e (350 and 500), the virus accumulation decreased with VA (Figure 5a). For mutation occurrence, N_e, s and VA had a significant positive effect. The interaction terms VA × N_e and s × N_e also showed a significant effect. For low values of N_e (50, 200 and 350), the mutation occurrence increased with VA or s, while for higher values of N_e (500), the mutation occurrence slightly decreased or stayed constant with VA or s (Figures 5b and c).

a



b



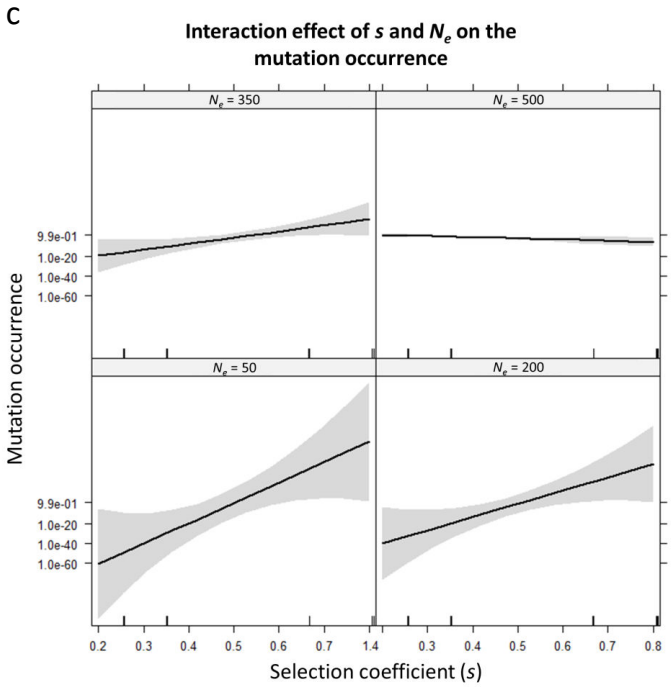


Figure 5: Predicted values from the GLMs for the change in virus accumulation (a) or for mutation occurrence (in log odds, b and c) during the experimental evolution according to VA, N_e and/or s. The grey bands give the 95% confidence limits around the fitted effects.

3 Discussion

Using experimental evolution, we have studied the evolution of 64 independent PVY lineages on 6 pepper DH lines. These lines have been chosen for the different levels of (i) selection (s) and (ii) genetic drift (N_e) that they impose during infection and for (iii) the different initial levels of accumulation of PVY (VA) (Figure 1, Table S1). The goal of this experimental evolution was to answer two major questions: (i) can the virus take divergent evolutionary trajectories when it evolves in different plant lines of the same species and, in particular, can virus adaptation be avoided? (ii) If different evolutionary trajectories occur, can we explain these differences by taking into account the evolutionary forces imposed by the host on virus populations?

3.1 Closely-related pepper lines impose divergent evolutionary trajectories to PVY

PVY lineages have evolved on closely-related DH lines sharing, on average, 50% of the polymorphic sites that distinguish the 2 parental lines. Our results demonstrate that PVY takes

divergent evolutionary trajectories when passaged on those pepper genotypes coming from the same DH population. First, some PVY lineages were not able to adapt to some pepper genotypes. Indeed, a total of 9 lineages went to extinction in the DH lines HD2256 and HD2321. Secondly, 33 lineages did not adapt. They did not show any increase in fitness or mutation in the central part of the VPg (Table 1). Thirdly, we also found multiple evidence of PVY adaptation to certain DH lines. Eighteen lineages evolved in 4 DH lines showed a significant increase in virus accumulation compared to the initial PVY clone (Figure 4). Since virus accumulation is a proxy of virus fitness, we can conclude that some PVY lineages have increased their fitness during the experimental evolution. Moreover, those increases in fitness were always associated with the fixation of one or two mutations in the VPg, except for lineage L56. For this specific lineage, mutations outside the VPg cistron might have played a role in this increase in fitness.

Most of the VPg mutations associated with fitness gains in the PVY populations have been previously shown to determine adaptation to the *pvr2*^{3} resistance gene, present in all DH lines in this study, by reverse genetics using PVY cDNA clones (Ayme et al. 2006; Ayme et al. 2007). This is the case of mutations 101G, 115K and 119N. The role of mutation 115M that arose frequently in our experimental evolution was studied by reverse genetics in the present study (Figure S3). In the context of SON41-119N, the mutation 115M conferred a gain of fitness in the 4 DH lines studied. The reverse genetics study showed also that mutations 115M and 115K conferred a large fitness gain in DH lines HD219 and HD2321. This seems to contradict the fact that none of the PVY lineages evolved in HD219 showed a gain of fitness or fixed any adaptive mutation in the central part of the VPg. Similarly, in HD2321, only 2 lineages of 8 showed PVY adaptation linked to mutation 115M and/or 101G. This suggests that additional forces have counteracted the appearance and/or emergence of adaptive mutations in these lines. We have proved that the pepper genotype had a significant effect on both the occurrence of fixed nonsynonymous mutations and the change in virus accumulation after 7 passages. We can therefore assume that some plant traits are behind the divergent evolutionary trajectories observed.

3.2 Choosing the plant traits most effective to avoid PVY adaptation

3.2.1 Relevance and precision of estimation of plant traits used to explain PVY evolutionary trajectories

We proposed that the different evolutionary trajectories observed were the result of the action of plant traits linked to the evolutionary forces exerted on PVY populations. The DH lines were chosen on the basis of three traits: the selection imposed by the plant genotype on a virus population (s), the genetic drift imposed by the plant genotype on that virus population (N_e) and the initial level of accumulation of PVY in the plant genotype (VA). Both s and N_e have been assessed using data from a previous experiment where an equimolar mixture of 5 PVY variants has been inoculated in the same DH lines as the one used in this study (Rousseau et al. 2017, in press). The frequencies of the 5 variants have been measured at different time points and a Wright-Fisher model has been developed to infer N_e and the growth rates of each PVY variant in each DH line. The selection coefficient s has been calculated as the difference between the growth rates of the fittest and the weakest variants in each DH line. This trait is therefore dependent on the PVY population used to perform the experiment and the relevance of its use for our study could be questioned. The 5 PVY variants used to measure s included the 3 variants that have begun the experimental evolution: SON41-101G, -115K and -119N. The 2 other variants used to measure s were SON41-101G-115K and -115K-119N. Taken together, the 3 mutations 101G, 115K and 119N represent 48% (13/27) of the *de novo* mutations observed in the experimental evolution. Since the viral populations of our experimental evolution and the one used to measure s were composed of similar variants, we can reasonably assume that s is a relevant trait to consider. Rousseau et al. (2017) also shown that selection was a highly heritable trait ($h^2 = 0.94$, Table S1) among the DH lines, thereby strengthening its reliability. However, several mutations that have been fixed during the experimental evolution have not been taken into account in the experiment designed to estimate s and N_e, in particular mutation 115M. Moreover, the composition of the PVY population (*i.e.* number and nature of the variants) and the frequencies of the variants were quite different in the experimental evolution, where *de novo* mutations have emerged, and in the previous experiment, where evolution of the standing variation was considered. Consequently, s may not be well estimated and its representativeness in our experimental context can be questioned. Regarding N_e, it was estimated from the same previous experimental data using the same DH lines and the PVY population composed of 5 variants.

The heritability was quite high for this trait (h^2 = 0.63). We can assume that its stochastic nature makes it less sensitive to the composition of the viral population. Accordingly, we have shown that N_e estimates were quite independent on the PVY population used (Rousseau et al. 2017, in press). Finally, the initial virus accumulation (VA) of the three initial PVY clones has been assessed with a DAS-ELISA. It represents the initial level of adaptation of each virus to the DH lines, an important parameter to calculate the fitness gain after 7 passages. It is also a very reliable trait with a high heritability (h^2 = 0.85, Table S1). We can conclude that N_e, s and VA were estimated with a high degree of precision based on heritability values on the plant side and that N_e and VA estimates are representative of the context of the experimental evolution. The representativeness of s estimates is however more questionable, which should be taken into account in the interpretation of our results.

3.2.2 Effects of plant traits on virus evolution

GLM were performed to study the effect of plant traits on virus evolution. We wanted to explain two variables linked to virus evolution: the change in virus accumulation after 7 passages and the occurrence of fixed nonsynonymous mutations in the VPg cistron. The change in virus accumulation is a direct proof and quantification of the virus adaptation during the experimental evolution, whereas the occurrence of mutations is rather indirect evidence and not a true quantification of that adaptation. The change in virus accumulation is significantly impacted by N_e and VA only (Table 2). The interaction term VA × N_e has also a significant effect: when N_e is low, the virus fitness gain increases with VA, whereas when N_e is high, the fitness gain decreases with VA (Figure 5a).

We will first discuss the case where N_e is low (N_e < 200). We can reasonably assume that the mutation occurrence probability increases when the initial population size is large (*i.e.* VA is high) because more replication events are likely to occur to reach a large population size, leading to a higher number of incorrect nucleotide incorporations. Consequently, the probability that adaptive *de novo* mutations arise in the population will be higher when VA is high than when VA is low. Since N_e is small, genetic drift will be strong and some mutations will randomly be eliminated independently of their fitness effects. Adaptive mutations are scarcer in populations where VA is low, therefore these populations are likely to lose more adaptive mutations than populations where VA is high. In other words, a cumulative effect of the low N_e and low VA to avoid virus adaptation will be observed. This case is well illustrated

by the DH lines HD2256, HD219 and HD2321. In the three lines, the genetic drift was strong (N_e < 130) but the lines differed in their value of VA (Figure 1, Table S1). In HD219, VA was high, which means that the initial clone was already well-adapted. Therefore, the viral lineages were all able to maintain during the experimental evolution and to keep a high accumulation level after 7 months (Figure 3). In HD2256 and HD2321, VA was low, which means that the initial clone was poorly adapted to these two lines. In this situation, the effect of genetic drift has probably avoided the adaptation of some lineages, since the final accumulation level was still low and 9 extinction events have occurred (Figure 3, Table 1). This high number of PVY extinctions could be explained by the action of Muller's ratchet. This process occurs in finite populations of small size like the initial PVY clone in HD2256 and HD2321. Theory predicts that the population will accumulate slightly deleterious mutations over time. Individuals without deleterious mutations will become rare and will be definitively lost by genetic drift. If the deleterious mutations continue to accumulate, it can lead to the extinction of the population (Chao 1990). Since we have only sequenced the VPg cistron, we cannot exclude the possibility that deleterious mutations have occurred in other parts of the PVY genome and have conducted to the extinction of the lineages. Interestingly, the mutation 115K has arisen but was not fixed in the L22 lineage in HD2321. We have proved that this mutation confers a selective advantage to the virus in this DH line (Figure S3A). However, the lineage went to extinction in the next cycle, probably because the advantageous mutation has been lost by genetic drift.

We will now discuss the case where N_e is high (N_e > 400) and for which the fitness gain decreases when VA increases (Figure 5a). Since genetic drift is weak, we expect that the *de novo* adaptive mutations in the PVY populations will more rarely be lost after their appearance and that the fixation of adaptive mutations will mostly depend on their fitness effects. In this situation, some predictions can be made on the evolution of the viral lineages. Fisher's geometric model of adaptation describes a multidimensional phenotypic space where each axis corresponds to a phenotypic trait and where only one single fitness optimum can be reached. A genotype is represented by a point in this space and moves across the fitness landscape according to the effects of *de novo* mutations. The model predicts that the probability of increasing fitness will be proportional to the distance from the optimum. It means that when a genotype is close to the fitness optimum, the adaptive mutations will have a smaller effect on its fitness (Tenaillon 2014). Several studies have provided experimental evidences of this prediction. For instance, in an experimental evolution of *Escherichia coli*,

Barrick et al. (2010) have shown that mutants with a low fitness have fixed beneficial mutations with a larger effect than mutant with higher initial fitness. The ratio between the number of beneficial and deleterious mutations could also be involved in the higher increase in fitness of the populations with a low VA. Using mutation accumulation experiment with the bacteriophage ϕX174, Silander et al. (2007) have compared the number of *de novo* mutations and their effect on high-fitness and low-fitness lines. They found no beneficial mutation in high-fitness lines, whereas the rate of beneficial mutations was significantly higher in low-fitness lines. In our experiment, the evolutionary trajectories in DH lines HD2349, HD2344 and HD2173 are also in agreement with those findings. In the three lines, the genetic drift was low (N_e > 450) but the lines differed in their value of VA (Figure 1, Table S1). In HD2349, HD2344 and HD2173 inoculated with SON41-101G, VA was low and 19 lineages among 24 were able to increase their fitness. In HD2173 inoculated with SON41-119N and -115K, VA was high and no increase in fitness was observed.

The second variable linked to virus evolution was the mutation occurrence. We have shown that the occurrence of nonsynonymous mutations is positively impacted by s, N_e and VA (Table 2). The interaction terms VA × N_e and s × N_e have also a significant effect: when N_e is low, the mutation occurrence increases with VA or s, whereas when N_e is high, the mutation occurrence slightly decreases or stays constant with VA or s (Figure 5b and c). Since the mutations are mostly adaptive and linked to an increase in fitness, the interpretation of the VA × N_e interaction is the same as above for the increase in virus accumulation. The s × N_e interaction shows that combining a strong genetic drift and a low selective pressure avoids the occurrence of mutation, while increasing one or the two parameters leads to an increase in the mutation occurrence. That is what we observe in HD2256 and HD219, where both s and N_e are low. Only one nonsynonymous mutation (139V), probably not adaptive, has occurred in the lineages that have evolved in these lines. At the opposite, in HD2344, HD2349 and HD2173 with SON41-101G, both genetic drift and selection are relaxed and VA is low enough to allow adaptation. Therefore, the majority of the lineages have fixed adaptive mutations.

We can conclude that the different evolutionary trajectories taken by the PVY lineages are very well explained by the genetic drift imposed by the host, by the initial level of virus accumulation in the host and by the interaction between those two traits. In contrast, the selection pressures exerted by the host had a significant effect only on VPg mutation

occurrence. We cannot exclude that the lower importance attributed to *s*, compared to N_e and VA, in explaining the PVY evolutionary trajectories is the result of an imperfect estimation of *s*, that does not represent adequately the context of the experimental evolution. It is possible that other, more appropriate *s* estimates, would significantly explain the PVY evolutionary trajectories.

3.2.3 Agronomical perspectives

Several authors have proposed to fight against the emergence of virulent pathogen variants by taking advantage of evolutionary forces exerted by the host like genetic drift (Abel et al. 2015), selection (Lê Van et al. 2013) or the number of mutations required to acquire the virulence (Harrison 2002). The objective was to avoid the pathogen to better adapt to its resistant host or to antimicrobial drugs like antibiotics. However, to date, no study has shown the efficiency of this approach experimentally. Multiple studies have demonstrated that a strong genetic drift could cause a loss in fitness because of the Muller's ratchet process (Chao 1990; Duarte et al. 1992; Escarmís et al. 1996; de la Iglesia and Elena 2007). Nevertheless, all these studies used mutation accumulation experiments, where serial bottlenecks are not imposed by a host but by the experimenter to increase genetic drift and cause the fixation of deleterious mutations. Using mosquito-borne viruses, Grubaugh et al. (2016) have proved that virus populations undergo repetitive bottlenecks during mosquito infection in different mosquito species which leads to a loss of fitness, probably due to the Muller's ratchet. In this experiment, however, no different host genotypes belonging to the same species were used. Moreover, the pathogen populations were not exposed to a strong selective pressure like the major resistance gene in our study. An experimental proof of the manipulation of evolutionary forces exerted by the host to avoid the adaptation of the pathogen in a resistant host-pathogen system was therefore needed.

In a previous study (Rousseau et al. 2017), we have shown that N_e values imposed by a pepper host to a viral population were highly variable among pepper lines, even if these lines were genetically closely related. This trait was also highly heritable among the lines. Those high variability and heritability are important prerequisite for the use of N_e in breeding. In this study, we demonstrated that the use of evolutionary forces to avoid viral adaptation is efficient in the presence of a factor exerting a strong selective pressure (*i.e.* the *pvr2*^{3} major resistance gene). Therefore, the use of genetic drift was sufficient in this context to avoid the

selection of virulent variants. Our study also demonstrated that the experimental estimation of N_e was sufficiently accurate and representative of the viral population to allow the anticipation of the future evolution of the virus. Finally, to avoid virus adaptation, it is important to take simultaneously into consideration N_e and VA because of the strong impact of their interaction on the evolutionary trajectories. Indeed, we proved that it is possible to combine a high level of initial resistance (*i.e.* a low VA) and to avoid the virus adaptation to this resistance on the long term by choosing plants that impose a low N_e to virus populations.

4 Materials and methods

4.1 Virus and plant material

In this study, we used three different PVY infectious clones derived from the SON41p isolate. They were named SON41-101G, -115K and -119N according to the amino acid substitutions compared to SON41p and their positions in the VPg (viral protein genome-linked). These three mutations were shown to be responsible for PVY adaptation to the pepper resistance allele *pvr2*^{3} (Ayme et al. 2006). Moreover, they show a low, high and medium fitness, respectively, in *pvr2*^{3}-carrying pepper plants (Rousseau et al. 2017, in press).

Six doubled-haploid (DH) lines of *C. annuum* were used: HD219, HD2173, HD2256, HD2321, HD2344 and HD2349. They all derived from a cross between Yolo Wonder, an inbred line susceptible to PVY isolates, and Perennial, an inbred line carrying the PVY resistance allele *pvr2*^{3}. All DH lines carried *pvr2*^{3} and differed in their genetic background. They have been chosen based on the different levels of selection and genetic drift that they imposed on PVY populations (Figure 1, Table S1). The initial fitness levels of PVY variants were also contrasted between the DH lines. These selective, genetic drift and fitness levels were quantified thanks to three measures: the selection coefficient (s) and the effective population size (N_e) characterizing an artificial PVY population, and the initial virus accumulation (VA). Both s and N_e have been estimated by Rousseau *et al.* (2017). They have inoculated the first leaf of 15 pepper DH lines with an equimolar mixture of 5 PVY variants, including the 3 variants used in our study SON41-101G, -115K and -119N, as well as two double mutants named SON41-101G-115K and -115K-119N. At different time points after inoculation until 34 dpi, the virus populations were sampled. From these samples, RNA extraction, reverse transcription-polymerase chain reaction (RT-PCR) and next generation

sequencing (Miseq Illumina) were performed allowing to estimate the frequencies of the PVY variants in the plants. Thanks to these measures, the effective population size N_e and the growth rate of the PVY variants in each DH line from inoculation to 34 dpi have been estimated with a mechanistic-statistical model. Then, the selection coefficient *s* has been calculated as the difference between the growth rates of the fittest and the weakest variants. Note that *s* is strongly correlated to the variance of the growth rates of the 5 variants (r = 0.999 for the 15 DH lines). The broad-sense heritability (h^2), defined as the part of the total variance of the trait due to genetic variance, was high both for *s* and N_e (h^2=0.94 and h^2=0.63, respectively).

The initial virus accumulation (VA) was assessed in all DH lines for SON41-119N and only in HD2173 for SON41-101G and -115K during this study. The 3 PVY clones were first propagated in *Nicotiana tabacum* cv. Xanthi plants. Pepper plants were mechanically inoculated on their two cotyledons, with 20 plants per DH line. At 30 dpi, a quantitative DAS-ELISA (Double Antibody Sandwich - Enzyme Linked Immunosorbent Assay) was performed as described by Ayme et al. (2006). The mean virus concentration per DH line was assessed using serial dilutions of the infected plants and calculated relatively to a common control sample added in each ELISA plate. The experiment was performed in a climate-controlled growth room (20–22°C, 12-h light/day).

4.2 Experimental evolution

The three cDNA clones SON41-101G, -115K and -119N were first multiplied in Xanthi plants. To ensure that the initial concentration was the same between the three clones, virus accumulation of each clone was assessed by quantitative DAS-ELISA two weeks after inoculation and adjusted when necessary. Sixteen plants per DH line were then mechanically inoculated with SON41-119N on their two cotyledons in order to initiate 8 independent evolutionary lineages for each DH line. HD2173 was also inoculated with SON41-101G and -115K (Figure 2). Twenty-eight days post inoculation, 8 plants of the 16 were randomly chosen for each DH line-PVY clone combination to initiate the 64 PVY lineages. For each of these plants, 1 gram from 3 apical leaves showing mosaic symptoms was ground in a phosphate buffer (0.03 M Na_2HP0_4, 0.2% sodium diethyldithiocarbamate, 4 mL buffer per gram of leaves) and used to inoculate 3 plants of the same genotype per evolutionary lineage. Then, one of the three plants was randomly selected to inoculate the three plants of the next

passage. A total of 7 passages were performed similarly with a time interval of 28 days between each passage. All plants were kept in the same climate-controlled room during the experiment as described above.

4.3 Virus sequencing

At the end of each cycle, a sample of 200 µl was removed from each of the 64 inocula and kept at -20°C. From these samples, we performed RNA extraction using Trizol and RT-PCR of the VPg cistron using specific primers (Forward primer: 5'-GACCTTAAGCTGAAGGGAGTTTGGAAGAAGTCGC-3'; Reverse primer: 5'-ATTTGCTATTATGTAAGCCCC-3'). RT-PCR products were sent to Genoscreen (Lille, France) for sequencing. Sequences were visualized with the software ChromasPro v1.7.6 (www.technelysium.com.au/chromas.html) and aligned with the sequences of the PVY clones used to start the experimental evolution to detect mutations.

4.4 Measures of aggressiveness and virus accumulation

After the 7th infection cycle, all the evolutionary lineages that did not become extinct and the three initial PVY variants were mechanically inoculated on the two cotyledons of the DH line on which they have evolved, with 20 plants per PVY population. Six mock-inoculated plants were also used as controls for each DH line. At 30 dpi, the aggressiveness of the PVY populations was estimated by the impact of the virus on the height and the fresh weight of the plants. Fresh and dry weights of pepper plants belonging to the same progeny were shown previously to be highly correlated ($r = 0.957$, $p < 0.0001$) (Montarry et al. 2012). The plants were first cut at the cotyledon node and immediately measured and weighted. Then, 1 g from 3 apical leaves showing symptoms was sampled from each plant and a quantitative DAS-ELISA was performed to assess the final concentration of the lineages. The level of adaptation to the plant was then obtained by comparing the concentrations of the final PVY populations to those of the initial PVY populations. The experiment was performed in a climate-controlled room as previously described.

4.5 Measures of virus competitiveness

Competition experiments have been performed between the most frequent mutants of SON41-119N observed during the experimental evolution and the initial virus to assess their relative fitness. We used the SON41-119N clone and two infectious clones derived from the SON41p isolate and carrying either the 115K-119N or the 115M-119N amino acid substitutions in the VPg. The three clones were multiplied in Xanthi plants. Virus accumulation of each clone was then assessed by quantitative DAS-ELISA two weeks after inoculation to prepare inocula with the desired stoichiometry. Three competition experiments have been performed in 4 DH lines: HD219, HD2321, HD2344 and HD2349. For each DH line, 10 plants were inoculated with a 2:1 ratio mixture, based on quantitative DAS-ELISA, of the PVY mutants (i) SON41-119N and SON41-115K-119N, (ii) SON41-119N and -115M-119N or (iii) SON41-115M-119N and -115K-119N. In each case, the mutant expected to have the lowest competitiveness based on experimental evolution results (Figure 4) was the major component. Thirty days after inoculation, RNA extraction, RT-PCR of the VPg cistron and sequencing were realized as described above. To estimate the relative proportion of the mutants at the end of the competition experiment, the height of peaks at the codon position 115 were measured in the chromatograms. Although this method was shown to provide precise quantitative estimates of the relative proportions of the 2 virus variants (Fabre et al. 2014), we took into account only strong trends of virus frequency evolution. For this, plants were grouped in 4 categories where (i) variant 1 or (ii) variant 2 only could be detected on chromatograms or where (iii) variant 1 or (iv) variant 2 predominated in the population but the two variants could be detected on chromatograms.

4.6 Statistical analyses

All statistical analyses were performed using the R software (http://www.r-project.org/). Wilcoxon test was used to compare virus accumulation mean values between initial and final viral populations. Kruskal–Wallis test followed by the Nemenyi *post hoc* test were used to compare aggressiveness mean values between mock-inoculated plants and plants inoculated with either the initial PVY clones or the final PVY populations. Dunnett test was used to compare the virus accumulation as well as the aggressiveness between every final PVY population and the initial PVY populations. We used general linear models (GLM) to study the effects of N_e, s and VA on two response variables linked to the virus adaptation during

experimental evolution. The first response variable was the change in virus accumulation, a continuous variable calculated as the ratio of the concentration of the final PVY populations and the sum of the concentrations of the starting and the final PVY populations, assessed with quantitative DAS-ELISA. A value of 0 was assigned to the extinct lineages. To approximate normality, Yeo-Johnson transformation was applied to this variable. The second variable ("mutation occurrence") was a binary variable representing the occurrence (1) or absence (0) of fixation of at least one nonsynonymous mutation in a PVY evolutionary lineage during the cycles. We used a Gaussian distribution with an identity link function for the change in virus accumulation and a binomial distribution with a logit link function for the mutation occurrence. All the individual variables and their pairwise interactions were included in the full models and backward stepwise selection was performed to eliminate variables. The Akaike's information criterion (AIC) was used to select the best-fitting models (Akaike 1974). The GLM analyses were done with the R software using the glm function and the packages 'car' and 'effects'.

Funding information

L. Tamisier's PhD was supported by the BAP (Biologie et Amélioration des Plantes) department and SMaCH (Sustainable Management of Crop Health) metaprogramme of INRA and by the Région Provence-Alpes-Côte d'Azur (PACA). The experimental work was supported by the SMaCH metaprogramme.

Acknowledgments

We thank the greenhouse staff for support in plant experimentation. We thank HM Clause for its support.

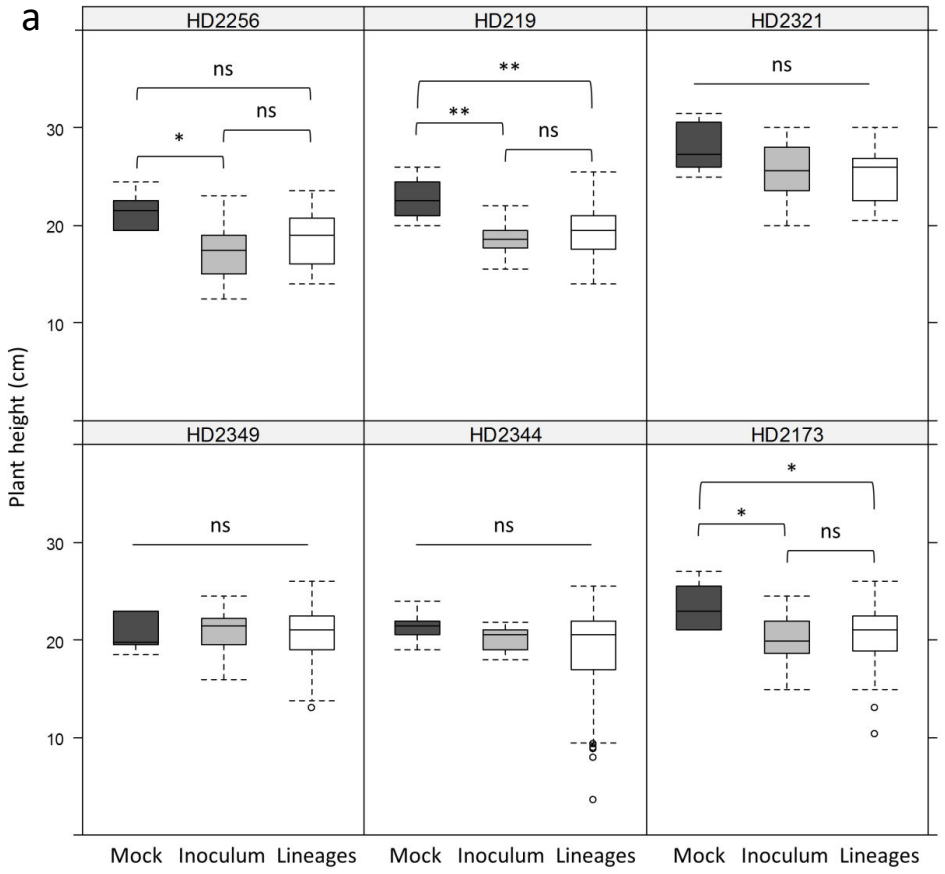
5 Supplementary material

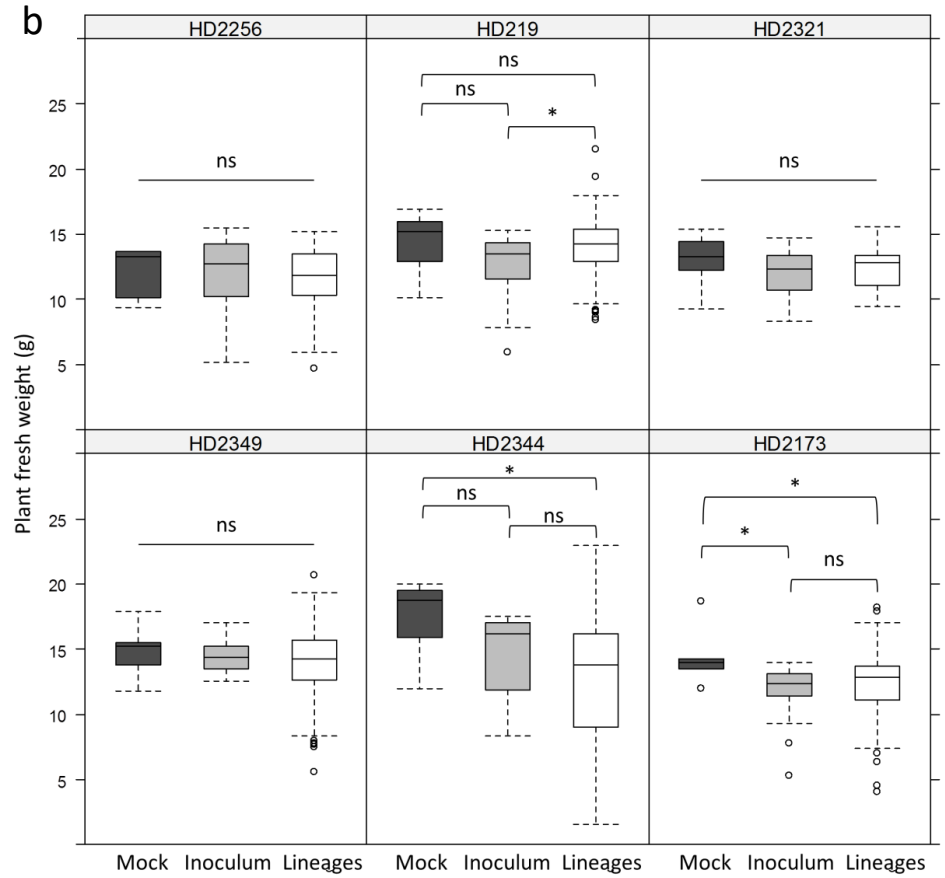
Table S1: Traits characterizing each pepper DH line - initial PVY clone combination used for the experimental evolution. When possible, the broad-sense heritability (h^2) of the trait has been calculated.

DH lines	*s* ($h^2 = 0.94$)	N_e ($h^2 = 0.63$)	VA 119N ($h^2 = 0.85$)	VA 101G	VA 115K
HD2256	0.26	34.86	0.23	-	-
HD219	0.15	46.93	1.08	-	-
HD2321	0.81	120.76	1.47×10^{-5}	-	-
HD2349	0.81	462	0.4	-	-
HD2344	0.67	516.03	0.06	-	-
HD2173	0.35	493.11	0.7	0.33	1.33

Table S2: Amino acid substitutions in the VPg of 64 evolutionary lineages derived from either SON41-119N, SON41-101G or SON41-115K after 7 passages on pepper DH lines.

Source inoculum	Number of lineages	Amino acid substitutions in the VPg																									
		101									110											120				139		
SON41 119N	31	S	E	V	R	R	K	M	V	E	D	D	E	I	E	T	Q	A	L	N	S	H	...	K	I	D
	2	-	K	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-
	1	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	K	-	-	-	-	-	-		-	-	-
	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-	-	-		-	-	-
	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	V	-
	1	G	-	-	-	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-	-	-		-	-	-
	1	-	K	-	-	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-	-	-		-	-	-
		101									110											120				139		
SON41 101G	1	G	E	V	R	R	K	M	V	E	D	D	E	I	E	T	Q	A	L	D	S	H	...	K	I	D
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-		-	-	-
	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	K	-	-	-	-	-	-		-	-	-
	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	M	-	-	-	N	-	-		-	-	-
		101									110											120				139		
SON41 115K	8	S	E	V	R	R	K	M	V	E	D	D	E	I	E	K	Q	A	L	D	S	H	...	K	I	D
		No substitutions																									


a
HD2256
HD219
HD2321
HD2349
HD2344
HD2173
Plant height (cm)
ns
*
**
Mock
Inoculum
Lineages


b
HD2256
HD219
HD2321
HD2349
HD2344
HD2173
Plant fresh weight (g)
ns
*
Mock
Inoculum
Lineages

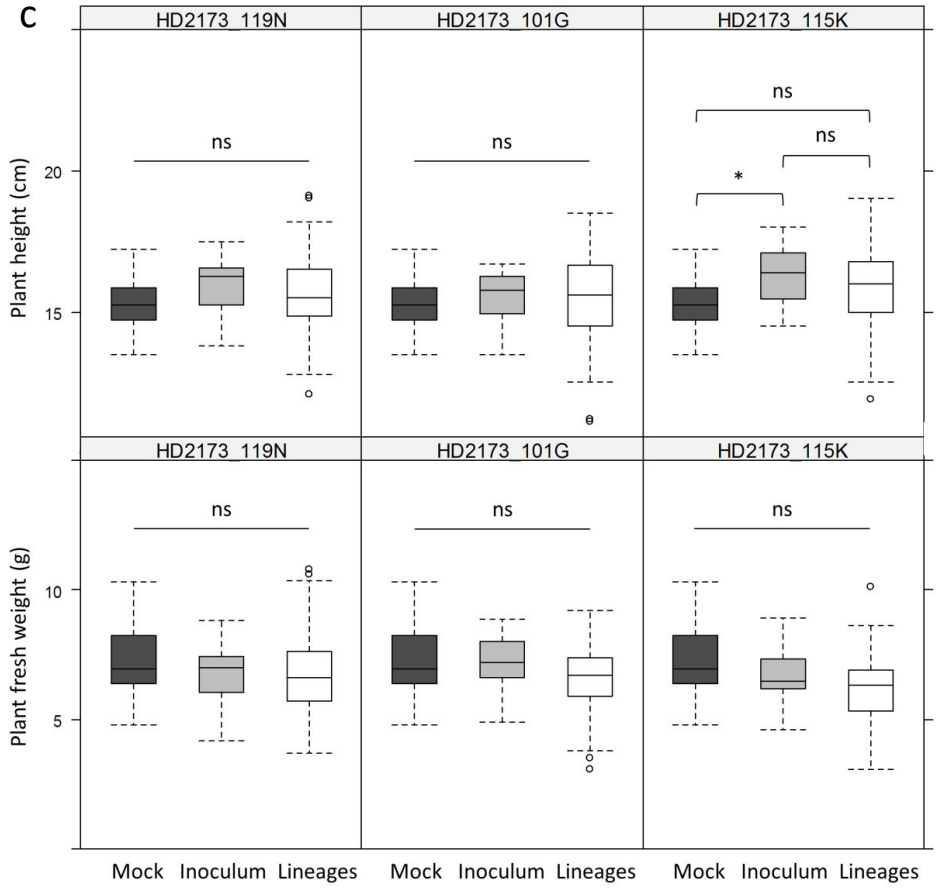


Figure S1: Aggressiveness of the final evolutionary lineages that have evolved on 6 pepper DH lines. Six plants per DH line have been mock-inoculated (dark gray), while 20 plants per DH line have been inoculated either with the initial infectious cDNA clone SON41-119N (light grey) or with each of the remaining evolutionary lineages that have been passaged on these lines (white). Boxplots of plants height (a) and fresh weight (b) at 30 dpi are represented. The same experiment has been performed independently with the 3 initial cDNA clones SON41-119N, -101G and -115K as well as the resulting lineages passaged on HD2173 (c). Significance levels were obtained with the Kruskal–Wallis test followed by the Nemenyi *post hoc* test (ns: not significant, * $p < 0.05$, ** $p < 0.01$).

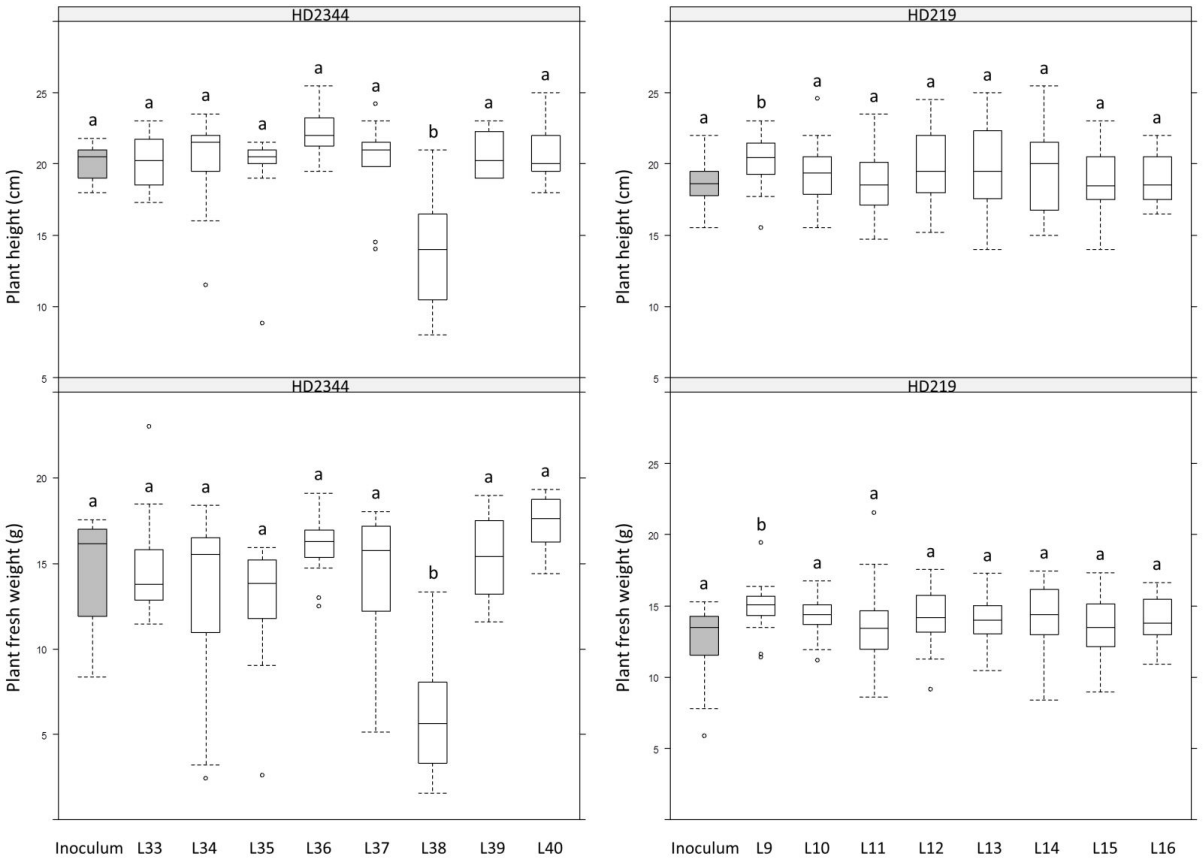
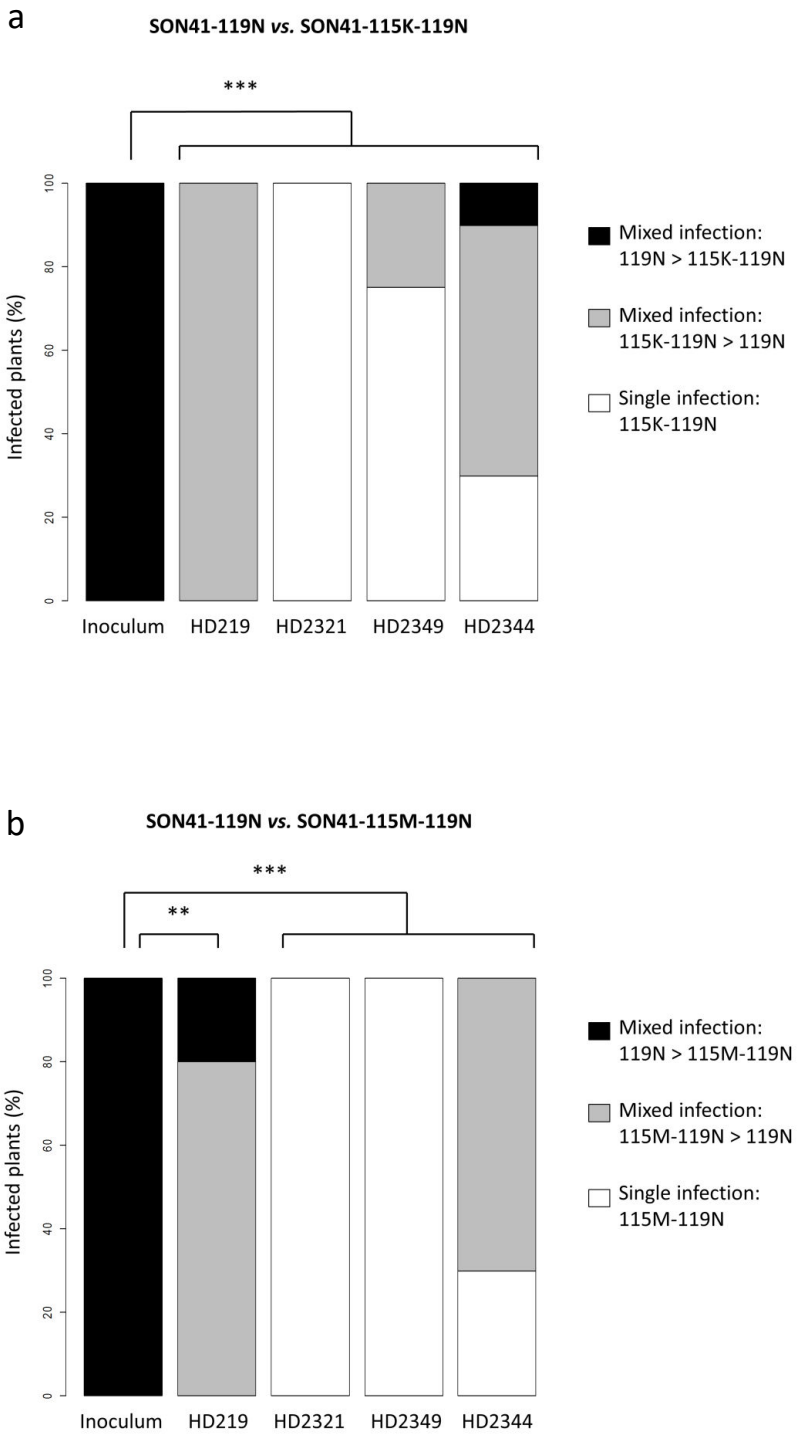


Figure S2: Aggressiveness of the 8 evolutionary lineages passaged on the pepper lines HD2344 and HD219. The initial infectious cDNA clone SON41-119N (light grey) and each of the lineages (white) have been inoculated on 20 plants. Boxplots of plants height and fresh weight at 30 dpi are represented. The letters a and b indicated the different groups obtained after the comparison of each lineage to the inoculum using Dunnett test ($p < 0.05$).


a
SON41-119N vs. SON41-115K-119N
***
Infected plants (%)
0
20
40
60
80
100
Inoculum
HD219
HD2321
HD2349
HD2344
Mixed infection: 119N > 115K-119N
Mixed infection: 115K-119N > 119N
Single infection: 115K-119N
b
SON41-119N vs. SON41-115M-119N
***
**
Infected plants (%)
0
20
40
60
80
100
Inoculum
HD219
HD2321
HD2349
HD2344
Mixed infection: 119N > 115M-119N
Mixed infection: 115M-119N > 119N
Single infection: 115M-119N

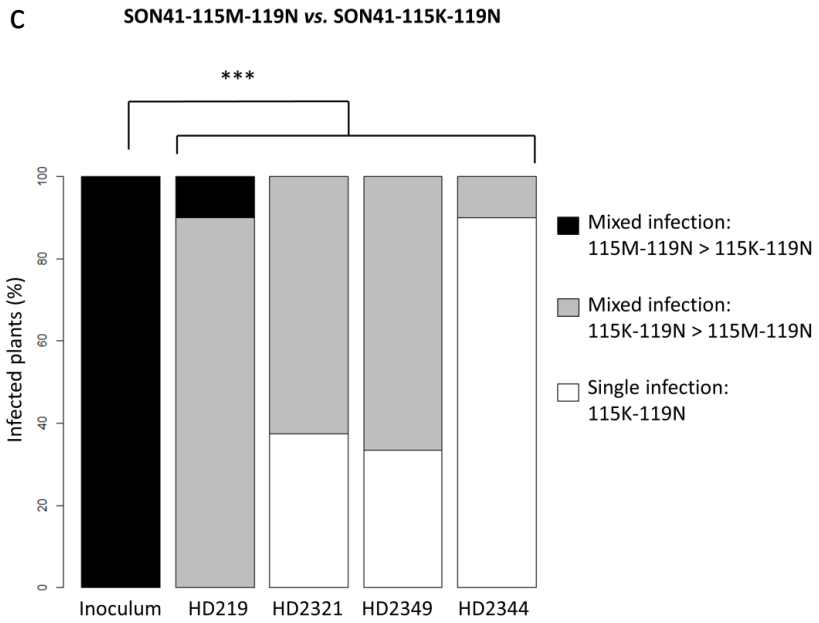


Figure S3: Competition experiments between SON41-119N *vs.* SON41-115K-119N (a), SON41-119N *vs.* SON41-115M-119N (b) and SON41-115M-119N *vs.* -115K-119N (c). The competitions have been performed on 4 DH lines: HD219, HD2321, HD2349 and HD2344. The inoculum was composed of a 2:1 ratio mixture of the two viruses in competition. The first bar represents the inoculum composition and each other bar represents the composition of the virus populations at 30 dpi among 10 plants. Three groups of plants were distinguished for each DH line – competition combination: (i) the first or (ii) the second virus predominated in mixed infection or (iii) the second virus only was detected (single infection). There was no plant where the first virus, in majority in the inoculum, was in single infection. Fisher's exact test was used to compare plant distribution among the two categories: (i) virus 1 or (ii) virus 2 predominated in mixed infection or was only detectable (single infection) to the distribution expected from the inoculum composition, *i.e.* the predominant virus in the inoculum was also predominant in the infected plants. Bonferroni correction was used to correct for multiple testing (** $p < 0.01$, *** $p < 0.001$).

6 Complément d'information à l'article

Dans cette étude, trois variables ont été utilisées pour expliquer les trajectoires évolutives des populations de PVY : la dérive génétique induite par la plante durant les 34 jours suivant l'inoculation (N_e), la sélection (s) et l'accumulation virale (VA). La dérive génétique à l'inoculation (N_e^{inoc}) avait également été mesurée pour ces lignées de piment par la même méthode que celle présentée dans les chapitres précédents. Cependant, cette variable n'a pas pu être incluse dans le modèle linéaire généralisé (GLM) utilisé car une forte colinéarité entre la sélection et la dérive génétique à l'inoculation a été détectée, avec un coefficient de corrélation de Pearson de 0,82 ($p < 0,001$) et un facteur d'inflation de la variance (VIF) de 2090 et 391 pour chacune des deux variables. La mesure du VIF est une technique classique de détection de la colinéarité qui se base sur la proportion de variance partagée par la variable considérée avec les autres variables explicatives. Généralement, un VIF supérieur à 10 suggère une forte colinéarité avec d'autres variables (Quinn and Keough 2002). La présence de colinéarité peut perturber les estimations des paramètres du modèle, modifiant notamment les signes et les valeurs des coefficients, et doit donc être évitée. Dans notre cas, la colinéarité observée ne doit pas être interprétée comme un potentiel lien biologique entre la sélection et la dérive génétique à l'inoculation. En effet, la liaison détectée est avant tout attribuable au faible effectif considéré (6 lignées HD de piment). La colinéarité n'est d'ailleurs plus présente lorsque l'on étudie les relations entre ces deux variables sur 89 lignées HD, comme présenté dans le chapitre 3.

Nous avons choisi d'exclure la dérive génétique à l'inoculation et de conserver la sélection dans l'analyse car l'impact de la dérive était déjà en partie mesuré par la dérive génétique globale sur 34 jours. Il nous semblait donc plus pertinent d'inclure l'effet de la seconde force évolutive qu'est la sélection dans le modèle. Néanmoins, l'analyse a également été réalisée en conservant la dérive génétique à l'inoculation et en éliminant la sélection du modèle. Les résultats obtenus sont similaires à ceux du premier modèle décrit dans l'article et sont présentés dans le tableau ci-dessous.

Tableau supplémentaire 1 : Résultats des GLM pour les deux variables liées à la capacité d'adaptation du virus. Les variables explicatives sont la taille efficace de population à l'inoculation (N_e^{inoc}), la taille efficace de population au cours de l'infection (N_e) et l'accumulation virale (VA) qui caractérisent chaque couple de clone infectieux de PVY/lignée HD de piment utilisés durant l'expérience d'évolution. Les interactions deux à deux ont été incluses dans le modèle. Seules les variables conservées après avoir effectué de la sélection de modèle et avoir retenu le modèle final en fonction du critère d'information d'Akaike (AIC) sont présentées. Significativité : * $p < 0,05$: ** $p < 0,01$ et *** $p < 0,001$.

Variable d'adaptation du virus	Variable explicative	Estimation	Erreur standard	Test de Wald	Significativité
Changement d'accumulation	Ordonnée à l'origine	0,212	0,119	1,773	0,081
	N_e	0,001	3,21e-04	4,792	1,130e-05 ***
	VA	0,372	0,179	2,079	0,042 *
	VA x N_e	-0,001	4,63e-04	-3,05	0,003 **
Fixation de mutation	Ordonnée à l'origine	-22,493	10,054	-2,237	0,025 *
	N_e	0,059	0,025	2,33	0,020 *
	N_e^{inoc}	1,609	0,756	2,129	0,033 *
	VA	9,751	5,096	1,913	0,056
	VA x N_e	-0,041	0,017	-2,333	0,020 *
	N_e^{inoc} x N_e	-0,003	0,002	-2,123	0,034 *

References

Abel, S., Abel zur Wiesch, P., Davis, B. M. and Waldor, M. K. (2015). Analysis of Bottlenecks in Experimental Models of Infection. *PLoS Pathog.* **11**, e1004823.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723.

Ayme, V., Souche, S., Caranta, C., Jacquemond, M., Chadœuf, J., Palloix, A. and Moury, B. (2006). Different mutations in the genome-linked protein VPg of *Potato virus Y* confer virulence on the *pvr2*^{3} resistance in pepper. *Mol. Plant. Microbe Interact.* **19**, 557–563.

Ayme, V., Petit-Pierre, J., Souche, S., Palloix, A. and Moury, B. (2007). Molecular dissection of the *Potato virus Y* VPg virulence factor reveals complex adaptations to the *pvr2* resistance allelic series in pepper. *J. Gen. Virol.* **88**, 1594–1601.

Barrick, J. E., Kauth, M. R., Strelioff, C. C. and Lenski, R. E. (2010). *Escherichia coli* rpoB mutants have increased evolvability in proportion to their fitness defects. *Mol. Biol. Evol.* **27**, 1338–1347.

Bedhomme, S., Lafforgue, G. and Elena, S. F. (2012). Multihost Experimental Evolution of a Plant RNA Virus Reveals Local Adaptation and Host-Specific Mutations. *Mol. Biol. Evol.* **29**, 1481–1492.

Betancourt, M., Fereres, A., Fraile, A. and Garcia-Arenal, F. (2008). Estimation of the Effective Number of Founders That Initiate an Infection after Aphid Transmission of a Multipartite Plant Virus. *J. Virol.* **82**, 12416–12421.

Chao, L. (1990). Fitness of RNA virus decreased by Muller's rachet. *Nature* **348**, 454–455.

de la Iglesia, F. and Elena, S. F. (2007). Fitness Declines in *Tobacco Etch Virus* upon Serial Bottleneck Transfers. *J. Virol.* **81**, 4941–4947.

Duarte, E., Clarke, D., Moya, A., Domingo, E. and Holland, J. (1992). Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proc. Natl. Acad. Sci. USA.* **89**, 6015–6019.

Elena, S. F. and Sanjuán, R. (2007). Virus Evolution: Insights from an Experimental Approach. *Annu. Rev. Ecol. Evol. Syst.* **38**, 27–52.

Escarmís, C., Dávila, M., Charpentier, N., Bracho, A., Moya, A. and Domingo, E. (1996). Genetic lesions associated with Muller's ratchet in an RNA virus. *J. Mol. Biol.* **264**, 255–267.

Fabre, F., Moury, B., Johansen, E. I., Simon, V., Jacquemond, M. and Senoussi, R. (2014). Narrow bottlenecks affect *Pea seedborne mosaic virus* populations during vertical seed transmission but not during leaf colonization. *PLoS Pathog.* **10**, e1003833.

French, R. and Stenger, D. C. (2003). Evolution of *Wheat streak mosaic virus*: dynamics of population growth within plants may explain limited variation. *Annu. Rev. Phytopathol.* **41**, 199–214.

Gibbs, A., Gibbs, M., Ohshima, K. and Garcia-Arenal, F. (2008). More about plant virus evolution; past, present and future. In *Origin and Evolution of Viruses,* 229–50.

Grubaugh, N. D., Weger-Lucarelli, J., Murrieta, R. A., Fauver, J. R., Garcia-Luna, S. M., Prasad, A. N., Black, W. C. and Ebel, G. D. (2016). Genetic drift during systemic arbovirus infection of mosquito vectors leads to decreased relative fitness during host switching. *Cell Host Microbe* **19**, 481–492.

Guidot, A., Jiang, W., Ferdy, J.-B., Thébaud, C., Barberis, P., Gouzy, J. and Genin, S. (2014). Multihost experimental evolution of the pathogen *Ralstonia solanacearum* unveils genes involved in adaptation to plants. *Mol. Biol. Evol.* **31**, 2913–2928.

Harrison, B. D. (2002). Virus variation in relation to resistance-breaking in plants. *Euphytica* **124**, 181–192.

Hillung, J., Cuevas, J. M., Valverde, S. and Elena, S. F. (2014). Experimental evolution of an emerging plant virus in host genotypes that differ in their susceptibility to infection. *Evolution* **68**, 2467–2480.

Lê Van, A., Caffier, V., Lasserre-Zuber, P., Chauveau, A., Brunel, D., Le Cam, B. and Durel, C.-E. (2013). Differential selection pressures exerted by host resistance quantitative trait loci on a pathogen population: a case study in an apple × *Venturia inaequalis* pathosystem. *New Phytol.* **197**, 899–908.

Montarry, J., Cartier, E., Jacquemond, M., Palloix, A. and Moury, B. (2012). Virus adaptation to quantitative plant resistance: erosion or breakdown? *J. Evol. Biol.* **25**, 2242–2252.

Moury, B., Fabre, F. and Senoussi, R. (2007). Estimation of the number of virus particles transmitted by an insect vector. *Proc. Natl. Acad. Sci. USA.* **104**, 17891–17896.

Quinn, G. P. and Keough, M. J. (2002). Experimental design and data analysis for biologists. Cambridge University Press.

Rouzine, I. M., Rodrigo, A. and Coffin, J. M. (2001). Transition between Stochastic Evolution and Deterministic Evolution in the Presence of Selection: General Theory and Application to Virology. *Microbiol. Mol. Biol. Rev.* **65**, 151–185.

Sacristán, S., Malpica, J. M., Fraile, A. and Garcia-Arenal, F. (2003). Estimation of Population Bottlenecks during Systemic Movement of *Tobacco mosaic virus* in Tobacco Plants. *J. Virol.* **77**, 9906–9911.

Sanjuan, R., Nebot, M. R., Chirico, N., Mansky, L. M. and Belshaw, R. (2010). Viral mutation rates. *J. Virol.* **84**, 9733–9748.

Silander, O. K., Tenaillon, O. and Chao, L. (2007). Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. *PLoS Biol.* **5**, e94.

Smith, D. R., Adams, A. P., Kenney, J. L., Wang, E. and Weaver, S. C. (2008). *Venezuelan equine encephalitis virus* in the mosquito vector *Aedes taeniorhynchus*: infection initiated by a small number of susceptible epithelial cells and a population bottleneck. *Virology* **372**, 176–186.

Tenaillon, O. (2014). The utility of Fisher's geometric model in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* **45**, 179–201.

Tromas, N., Zwart, M. P., Lafforgue, G. and Elena, S. F. (2014). Within-Host Spatiotemporal Dynamics of Plant Virus Infection at the Cellular Level. *PLoS Genet.* **10**, e1004186.

Zwart, M. P. and Elena, S. F. (2015). Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annu. Rev. Virol.* **2**, 161–179.

Zwart, M. P., Daròs, J.-A. and Elena, S. F. (2011). One Is Enough: In Vivo Effective Population Size Is Dose-Dependent for a Plant RNA Virus. *PLoS Pathog.* **7**, e1002122.

Chapitre 5

Discussion et perspectives

L'objectif de ce travail de thèse était de comprendre par quels mécanismes évolutifs le fonds génétique de la plante pouvait contraindre l'évolution des populations virales et protéger un gène majeur de résistance du contournement. Pour répondre à cet objectif, la cartographie des régions génomiques du piment induisant des effets de dérive génétique sur les populations de PVY lors de l'étape d'inoculation a été réalisée. Ces régions génomiques ont été identifiées à la fois par cartographie de QTL chez une population de piments issue d'un croisement biparental et par génétique d'association dans une collection de piments représentative de la diversité génétique de l'espèce. Par la suite, l'impact des forces évolutives induites par la plante sur l'évolution des populations virales a été étudié en effectuant de l'évolution expérimentale ainsi qu'une analyse statistique globale de plusieurs jeux de données.

1 Synthèse des principaux résultats

Dans le chapitre 1, les QTL contrôlant la taille efficace de population (N_e) du PVY et du CMV lors de l'étape d'inoculation du virus dans la plante ont été cartographiés chez une descendance haploïde doublée (HD) de piment. Deux QTL communs expliquant en moyenne 33 et 24 % de la variation phénotypique observée ont été identifiés sur les chromosomes 7 et 12. Une interaction épistatique a également été révélée entre ces deux QTL. Par ailleurs, un QTL spécifique de chaque virus a été identifié sur le chromosome 6, expliquant 6 % de la variation du trait pour le PVY et 11 % pour le CMV. Ces résultats ont permis de démontrer que des mécanismes généraux mais également spécifiques de chaque virus étaient impliqués dans le contrôle de N_e à l'inoculation. Les N_e estimés pour un troisième virus, le ToMV, sur un sous-ensemble de lignées se sont révélés être positivement corrélés avec les N_e du PVY et non corrélés avec ceux du CMV, confirmant la présence de mécanismes généraux et spécifiques de chaque espèce virale. Nous avons émis l'hypothèse que les mécanismes généraux agiraient durant les toutes premières étapes de l'infection, avant l'établissement d'interactions moléculaires entre les virus et les composants intracellulaires de la plante. Ils pourraient par exemple agir sur l'épaisseur de la feuille ou sa capacité à résister aux blessures, limitant ainsi l'entrée des particules virales dans la plante. Les mécanismes spécifiques de chaque virus se mettraient en place légèrement plus tard au cours de l'inoculation et pourraient jouer sur la probabilité d'un virus à parvenir à initier le foyer d'infection une fois les premières cellules de la feuille infectées. Enfin, les résultats ont été comparés avec ceux précédemment obtenus par Quenouille et al. (2014) sur la même population de piments. Ils avaient notamment identifié VA-6, un QTL localisé sur le chromosome 6 et affectant l'accumulation virale. L'intervalle de confiance de VA-6 inclut une partie de l'intervalle de confiance du QTL spécifique du PVY,

nommé PVY-6, même si les deux QTL n'étaient pas exactement à la même position. Par ailleurs, VA-6 est en interaction avec un QTL qui contrôle la capacité du virus à contourner le gène majeur *pvr2*^{3} efficace contre le PVY. Nous avons supposé que VA-6 et PVY-6 sont liés, ou appartiennent au même locus, et donc que PVY-6 contribue à augmenter la durabilité de la résistance. Cette hypothèse pourrait être valide car l'allèle qui diminue N_e à l'inoculation diminue également l'accumulation virale et augmente la durabilité du gène majeur. En réduisant N_e, PVY-6 pourrait notamment aider à éliminer des variants déjà adaptés au gène majeur durant l'étape d'inoculation, augmentant ainsi la durabilité de la résistance. Néanmoins, de plus amples analyses sont nécessaires pour valider cette hypothèse, un manque de marqueurs ne permettant pas une localisation précise de PVY-6 dans cette région génomique.

Dans le chapitre 2, les régions génomiques contrôlant la taille efficace de population (N_e) lors de l'inoculation ainsi que l'accumulation virale (VA) du PVY dans la plante ont été identifiées par génétique d'association dans une core-collection de 276 accessions de piments représentative de la richesse allélique de la collection entière. Le génotypage par séquençage (GBS) de cette core-collection a permis d'obtenir 10308 marqueurs SNP répartis sur les 12 chromosomes du piment. La génétique d'association a mis en évidence 6 SNP appartenant à 3 régions génomiques sur les chromosomes 4, 6 et 12 en association avec au moins un des deux traits étudiés. Les marqueurs identifiés sur le chromosome 4 co-localisent avec le gène majeur *pvr2*. Pour le N_e à l'inoculation, les régions détectées sur les chromosomes 6 et 12 sont éloignées de 10 à 20 Mb des QTL précédemment identifiés, laissant supposer que plusieurs loci sont impliqués dans le contrôle de ce trait ou que les QTL précédemment identifiés n'étaient pas cartographiés précisément par manque de marqueurs. Les résultats révèlent également l'existence d'une région génomique sur le chromosome 6 impliquée à la fois dans le contrôle du N_e des populations virales à l'inoculation et de VA. Cette région co-localise avec le QTL d'accumulation virale VA-6 détecté par cartographie de QTL sur ce chromosome et qui est en interaction épistatique avec un QTL lié à la durabilité du gène majeur. De plus, pour les deux SNP en association sur le chromosome 6, l'allèle favorisant la diminution du N_e favorise également la réduction de VA. Cette étude permet donc de confirmer qu'un facteur génétique commun contrôle la dérive génétique à l'inoculation et l'accumulation virale dans la plante entière.

Dans le chapitre 3, nous avons étudié les mécanismes mis en place par le fonds génétique de la plante pour orienter l'évolution des populations virales et protéger le gène majeur du contournement. L'impact de plusieurs forces évolutives (la dérive génétique à l'inoculation, la

dérive génétique globale lors de l'infection des plantes et la sélection) ainsi que l'impact de l'accumulation virale (VA) sur la durabilité du gène majeur *pvr2*^{3} ont été étudiés via des mesures réalisées avec différents clones infectieux de PVY sur les lignées HD porteuses de *pvr2*^{3}. La durabilité avait été mesurée par Quenouille et al. (2014) en inoculant les lignées HD avec la chimère CI de PVY, qui ne contourne pas initialement *pvr2*^{3}, et en mesurant le pourcentage de plantes infectées, et donc contournées, pour chaque lignée en fin d'expérience (38 jours après inoculation). L'accumulation virale avait également été mesurée par Quenouille et al. (2014) en inoculant les lignées HD avec un variant de la chimère CI contournant *pvr2*^{3} et en effectuant un ELISA quantitatif. La dérive génétique à l'inoculation avait été évaluée dans le chapitre 1 en inoculant les lignées HD avec un clone infectieux de PVY contournant *pvr2*^{3}. Ce clone exprimait la GFP (green fluorescent protein) ce qui a permis de dénombrer les foyers d'infection sur les cotylédons inoculés. Enfin, la dérive génétique globale et la sélection induites par le fonds génétique durant les 21 jours suivant l'inoculation ont été mesurées en inoculant les lignées HD avec un mélange équimolaire de 5 variants de PVY (nommés G, N, K, KN et GK en fonction de la (ou des) mutation(s) leurs conférant la capacité à contourner *pvr2*^{3}). Le modèle de Rousseau et al. (2017) a ensuite permis d'évaluer l'impact de la dérive génétique et de la sélection imposées par chacune des lignées. Un modèle linéaire généralisé a par la suite permis d'identifier quelles variables affectaient significativement la fréquence de contournement du gène majeur. Les quatre variables explicatives (la dérive génétique à l'inoculation, l'accumulation virale, la dérive génétique globale et la sélection) et leur interactions deux à deux se sont révélées avoir un effet significatif sur la durabilité du gène majeur. Une augmentation de VA augmente les risques de contournement, une plus grande taille de population augmentant la probabilité d'apparition de mutations permettant de contourner le gène majeur. Une augmentation du N_e à l'inoculation ou du N_e global diminue également la durabilité du gène majeur, la dérive génétique ayant certainement un effet protecteur en éliminant les variants contournants préexistants ou en favorisant leur extinction suite à leur apparition par mutation dans la plante inoculée. Enfin, un fort coefficient de sélection entraîne un plus fort risque de contournement en cas de forts VA et N_e. Quand la taille de population est grande et la dérive génétique faible, les variants contournants ont une plus grande probabilité d'apparaître et l'effet de la sélection sera supérieur à celui de la dérive génétique, entraînant rapidement l'augmentation en fréquence de ces variants. Un résultat étonnant de cette analyse est qu'un fort coefficient de sélection diminuerait les risques de contournement en cas de faibles VA et N_e. Cependant, ce résultat est la conséquence de notre mesure de la sélection, et en réalité un faible coefficient de sélection entre variants viraux

davantage représentatifs de nos conditions expérimentales semble bien être à l'origine d'une diminution de la fréquence de contournement (voir partie 2.2.2 de la discussion pour plus de détails).

Dans le chapitre 4, nous avons étudié les mécanismes mis en place par le fonds génétique de la plante pour orienter l'évolution des populations virales partiellement adaptées au gène majeur. Pour ce faire, de l'évolution expérimentale de PVY a été réalisée durant 7 mois en effectuant des passages en série des populations virales sur des lignées HD imposant différents degrés de dérive génétique, de sélection et d'accumulation virale. Dans cette étude, les lignées HD étaient toutes porteuses du gène majeur *pvr2*^{3}. De ce fait, les populations de PVY ayant initié l'expérience étaient toutes porteuses d'une mutation leur permettant de contourner le gène majeur et de se maintenir dans la plante. Au cours de l'expérience d'évolution, des trajectoires évolutives contrastées ont été mises en évidence en fonction des lignées de piment. Ainsi, certaines populations de PVY se sont éteintes, d'autres n'ont montré aucun signe d'adaptation et certaines ont fixé des mutations non synonymes dans la partie centrale de la VPg. Pour la majorité des populations, la fixation d'une mutation était corrélée à un gain d'accumulation virale, et donc de valeur sélective, dans la lignée de piment dans laquelle elles avaient évolué. Des mutations parallèles ont également été observées chez des populations virales évoluant dans les mêmes lignées, les mutations 115K et 115M étant les plus répandues. Nous avons par la suite démontré que ces mutations étaient adaptatives pour le virus par génétique inverse. Ces différentes trajectoires évolutives ont pu être particulièrement bien expliquées par les forces évolutives imposées par les lignées de piment. La taille efficace de la population au cours de l'infection, l'accumulation virale et l'interaction de ces deux variables se sont révélées jouer un rôle majeur dans l'évolution des populations virales. Plus précisément, lorsque N_e est faible (*i.e.* la dérive génétique est forte), la capacité d'adaptation de la population virale augmente avec VA. Le nombre de mutations avantageuses disponibles est sûrement plus élevé quand VA est fort, alors que lorsque VA est faible, ces mutations sont peu nombreuses et potentiellement éliminées par effet de dérive génétique. Inversement, lorsque N_e est fort (*i.e.* la dérive génétique est faible), la capacité d'adaptation de la population diminue avec VA. Ce résultat est en accord avec le modèle géométrique de Fisher qui postule que lorsque le virus est proche de son optimum de valeur sélective (VA est fort), le nombre et l'effet des mutations avantageuses disponibles sont plus faibles que lorsque le virus est loin du pic de valeur sélective. En conclusion, cette étude démontre que combiner dans un même génotype une forte réduction du N_e à l'échelle de la plante et une forte réduction de l'accumulation virale permet d'empêcher

l'adaptation du virus et entraîne même des extinctions de populations, potentiellement par des effets de cliquet de Muller.

2 Comparaison des approches employées

2.1 Cartographie de QTL en population biparentale et GWAS

Dans ce manuscrit, deux méthodes de génétique d'association ont été employées : de la cartographie de QTL chez une population haploïde doublée issue d'un croisement biparental (chapitre 1) et de la détection de QTL sur une core-collection représentative de la diversité génétique de la collection de piment (chapitre 2). Ces deux approches ont été utilisées car elles sont assez complémentaires au regard des avantages et des limites de chacune (Tableau 1).

Tableau 1 : Avantages et limites de trois types de populations utilisées en génétique d'association. Modifié d'après Pascual et al. (2015).

Population	Avantages	Limites
Population biparentale (chapitre 1)	Construction aisée de la population Permet de cartographier des allèles rares Facile à analyser	Limité à deux allèles (espèces diploïdes) Peu d'événements de recombinaison QTL cartographiés sur de grands intervalles
Collection diversifiée (chapitre 2)	Collections déjà existantes Forte diversité génétique Evénements de recombinaison naturels Cartographie précise quand le déséquilibre de liaison est faible	Requiert un grand nombre de marqueurs Structure de la population Ne détecte pas les allèles rares Cartographie grossière quand le déséquilibre de liaison est grand
Population multiparentale (ex : MAGIC)	Plusieurs allèles et QTL ségrègent Cartographie fine Utile pour des approches gènes candidats Pas de structure de la population	Populations longues à établir Nécessite davantage de marqueurs et d'individus que les populations biparentales

La population HD nous a permis d'identifier des régions génomiques contrôlant le N_e à l'inoculation. Les QTL étaient cependant localisés sur de larges intervalles chromosomiques qui représentent plusieurs Mb en position physique. La core-collection a permis de mieux cartographier certains QTL et notamment d'identifier une co-localisation entre les QTL

contrôlant le N_e à l'inoculation et l'accumulation virale VA. Certains QTL n'ont cependant pas été retrouvés, potentiellement car ils étaient trop rares dans notre collection comme le QTL d'accumulation virale localisé sur le chromosome 6. Par ailleurs, des incertitudes subsistent quant à la localisation des QTL contrôlant le N_e à l'inoculation sur les chromosomes 6 et 12. L'utilisation de nouvelles populations plus adaptées à la cartographie fine pourrait être envisagée comme les populations MAGIC (multiparent advanced generation inter-cross) ou NILs (near-isogenic lines). Néanmoins, la création de telles populations est extrêmement longue et n'est probablement pas la solution la plus adéquate pour la poursuite de ces analyses. Des approches permettant l'ajout de marqueurs dans ces régions spécifiques chez les populations biparentales ou dans la core-collection permettraient potentiellement de mieux localiser ces QTL et de conclure quant à la présence d'une ou plusieurs régions génomiques contrôlant ce trait.

2.2 Evolution expérimentale par passages successifs et mesure de la durabilité du gène majeur *pvr2*^{3}

Le second objectif de la thèse était d'étudier l'impact des forces évolutives induites par la plante sur l'évolution des populations virales, l'ambition sur le long terme étant de déployer des génotypes d'hôtes contraignant l'évolution des virus afin de protéger la résistance conférée par le gène majeur. Deux stratégies ont été mises en place pour répondre à cet objectif. D'une part, la fréquence de contournement du gène majeur *pvr2*^{3} a été mesurée (Quenouille et al. 2014) et l'impact des forces évolutives induites par la plante sur la capacité du virus à contourner le gène majeur a été estimée (chapitre 3). D'autre part, de l'évolution expérimentale a été réalisée pour étudier l'impact des forces évolutives induites par la plante sur l'évolution du virus lorsqu'il est partiellement adapté au gène majeur (chapitre 4). Je vais maintenant discuter les avantages et les limites des deux approches et comparer les résultats obtenus.

2.2.1 La méthodologie employée

L'évolution expérimentale consiste à mesurer en temps réel l'évolution des populations en conditions contrôlées afin de tester des concepts de génétique des populations. Elle se place ainsi à la frontière entre les études réalisées sur le terrain et la modélisation (Bataillon et al. 2010). Contrairement aux études qui appréhendent l'évolution *a posteriori*, via par exemple l'analyse de fossiles, l'évolution expérimentale rend possible l'observation des processus évolutifs en direct et sur une période de temps accessible à l'Homme, allant de quelques

semaines à plusieurs années (Lenski 2004; Burke et al. 2010). Par rapport à la modélisation, l'évolution expérimentale présente l'avantage de mettre les attendus théoriques à l'épreuve de la réalité biologique. Classiquement, une expérience d'évolution est composée de trois étapes. Premièrement, une population d'origine est choisie pour débuter l'expérience. Il s'agit généralement d'une population génétiquement homogène, par exemple une population clonale (Barrick et al. 2009), ou d'une population composée de plusieurs génotypes de fréquences connues (Miralles et al. 1999; Griffin et al. 2004). La seconde étape consiste à faire évoluer la population dans différentes conditions choisies et contrôlées par l'expérimentateur. Une même condition est généralement imposée plusieurs fois à la population d'origine afin de générer des répétitions. Ces répétitions permettent de comparer la diversité des trajectoires évolutives empruntées par les populations au sein d'un même paysage adaptatif. Enfin, un ou plusieurs critères vont permettre d'évaluer la réponse des populations aux contraintes environnementales imposées par l'expérimentateur. La valeur sélective des populations évoluées est couramment mesurée et des compétitions entre populations ancestrales et évoluées peuvent également être effectuées (Bataillon et al. 2010). Les microorganismes, notamment les virus, ont été majoritairement utilisés pour ce type d'expériences car ils présentent des caractéristiques particulièrement avantageuses, comme un faible temps de génération, de grandes tailles de population, un taux de mutation élevé, de petits génomes et la possibilité de conserver les populations ancestrales pour les confronter aux générations ultérieures (Elena et Lenski 2003). Le pathosystème piment-PVY présentait donc des caractéristiques optimales pour réaliser des expériences d'évolution.

Par ailleurs, le concept d'évolution expérimentale regroupe trois grandes catégories d'expériences : les expériences d'accumulation de mutations, les expériences d'adaptation en cultures continues et les expériences d'adaptation par passages successifs. Chacune conduit à différentes dynamiques évolutives et permet donc de répondre à des objectifs qui lui sont propres (Barrick et Lenski 2013). La première catégorie consiste à laisser s'accumuler librement des mutations *de novo* au cours du temps. Des goulets d'étranglement sont régulièrement imposés aux populations, ce qui permet d'homogénéiser le nombre de descendants laissés par chaque individu d'une génération à une autre, indépendamment de leur valeur sélective (Figure 1a). L'effet de la sélection est ainsi minimisé et seules les mutations létales ou très fortement délétères sont purgées. Dans la seconde catégorie d'expérience, les populations sont maintenues dans un même milieu et à une taille relativement constante. Contrairement au premier type d'expérience, l'effet de la sélection est présent et des variants

mieux adaptés que ceux de départ sont susceptibles d'apparaître (Figure 1b). Enfin, la troisième catégorie d'expérience consiste à effectuer des passages en série en transférant périodiquement une proportion de la population dans un nouveau milieu. La taille de la population prélevée entre chaque passage est alors un paramètre important à évaluer. En effet, la taille de population doit être assez grande pour être suffisamment représentative de la population source, mais doit également laisser la possibilité à la population de s'accroître et d'évoluer à la suite du passage. Ce type d'expérience peut également conduire à une meilleure adaptation des populations au milieu dans lequel elles ont évolué (Figure 1c).

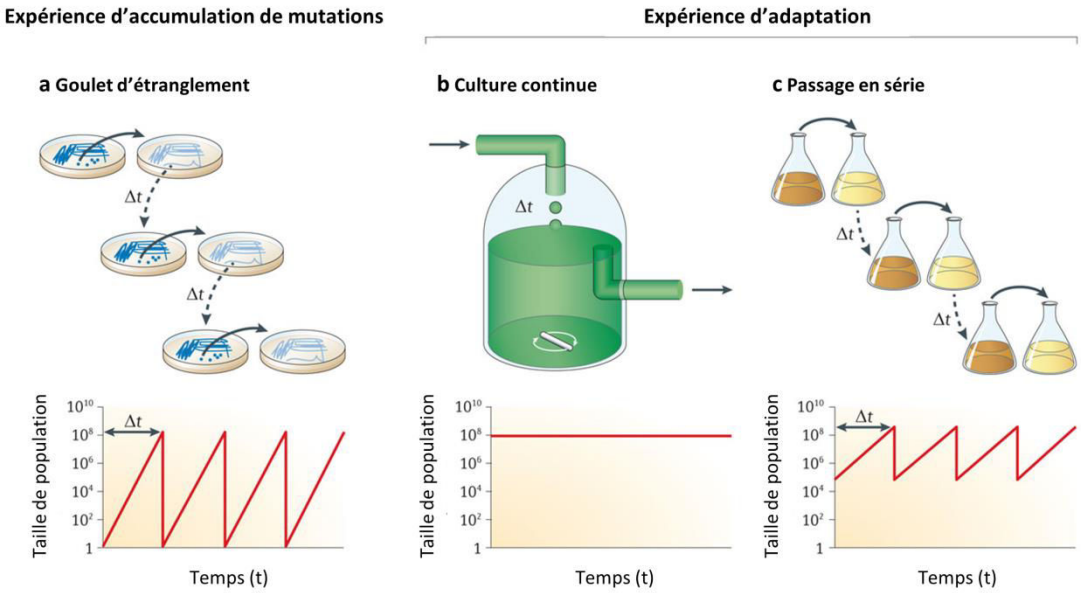


Figure 1 : Présentation des trois catégories principales d'évolution expérimentale. Les protocoles expérimentaux permettant de maintenir les populations au cours de l'expérience d'évolution sont illustrés en haut de la figure et les changements occasionnés sur les tailles des populations par chaque expérience sont représentés dans les graphiques en bas de la figure. (a) L'expérience d'accumulation de mutations consiste à induire des goulets d'étranglement répétés chez la population en réduisant drastiquement sa taille. (b) La culture continue maintient une taille de population relativement constante. (c) Les expériences d'évolution par passages en série consistent à prélever une proportion de la population source pour la faire évoluer dans un nouveau milieu de culture permettant son expansion. Modifié d'après Barrick et Lenski (2013).

L'expérience réalisée dans le chapitre 4 appartient évidemment à la catégorie des expériences d'adaptation par passages successifs. Les mesures de fréquences de contournement du gène *pvr2*^{3} de Quenouille et al. (2014) utilisées dans le chapitre 3 s'apparentent à ce que Barrick et Lenski (2013) définissent comme des expériences en cultures continues. En effet, les

populations virales ont évolué durant 38 jours dans une même plante et certaines populations se sont adaptées au gène majeur en fixant des mutations dans la région codant pour la VPg. Ces deux études peuvent donc être considérées comme de l'évolution expérimentale, même si elles imposent des contraintes différentes aux populations virales (Tableau 2). Chacune des expériences nous a permis d'étudier l'adaptation des populations virales aux lignées HD, mais les protocoles employés pour y parvenir étaient différents car les contextes de départs étaient différents. Dans le chapitre 3, les populations de départ n'étaient pas adaptées au gène majeur et il aurait été difficilement concevable de réaliser des passages en série, alors que l'adaptation partielle des populations de départ au gène majeur dans le chapitre 4 nous a permis d'avoir des populations initiales suffisamment adaptées pour effectuer ces passages. Enfin, il faut être prudent dans l'interprétation des résultats, particulièrement pour les données du chapitre 3 pour lesquelles on n'observe uniquement des corrélations qui ne sont pas forcément le reflet de la réalité biologique. En effet, comme expliqué précédemment, les variables explicatives et à expliquer ont été obtenues avec des variants viraux différents et non au cours d'une seule et même expérience.

Tableau 2 : Comparaisons des deux expériences d'évolution de ce manuscrit.

	"Culture continue" (*i.e.* mesure de la fréquence de contournement, chapitre 3)	**Transferts par passages en série** (chapitre 4)
Valeur sélective des populations virales de départ	Basse	Haute
Temps d'évolution	38 jours	7 mois
Fréquence de la dérive à l'inoculation	1 fois	7 fois
Nombre de phase de croissance exponentielle	1	7
Nombre de goulets d'étranglement	Moins nombreux	Nombreux

2.2.2 Les résultats obtenus

Un fait marquant lorsque l'on compare les résultats des deux études est que, malgré des contextes d'évolution différents (Tableau 2), l'impact des forces évolutives sur les populations virales est globalement le même entre les deux études. Ainsi, dans le chapitre 3, les lignées induisant un faible N_e (à l'inoculation ou global) ainsi qu'une faible VA montrent un taux de contournement du gène majeur plus faible. Les résultats obtenus dans le chapitre 4 sont globalement en adéquation avec ceux du chapitre 3, puisque lorsque la plante est déjà partiellement résistante au virus (VA est faible), un faible N_e et capable de s'opposer à la sélection en empêchant l'évolution de la population virale et peut même entraîner son extinction, contrairement à une lignée à forts N_e et s. Les deux études confirment donc que combiner un faible VA et un faible N_e dans une plante permet de contraindre l'évolution du virus.

Par ailleurs, des résultats plus complexes ont été obtenus pour l'effet de la sélection. Dans le chapitre 4, un impact significatif de la sélection n'a été détecté que pour l'une des deux variables à expliquer liée à l'adaptation des populations virales. Ce résultat ne signifie pas forcément que l'effet de la sélection sur l'évolution des virus est moindre que celui de la dérive génétique, mais plutôt que son estimation dans notre expérience n'est pas optimale. En effet, la sélection a été mesurée sur un ensemble de 5 variants viraux qui ne sont pas forcément représentatifs de notre contexte d'étude en évolution expérimentale, qui comprend ces 5 variants mais aussi beaucoup d'autres.

Dans le chapitre 3, un grand coefficient de sélection s, et donc une lignée dans laquelle certains variants viraux ont un taux d'accroissement très élevé, réduirait le taux de contournement. Ce résultat peut sembler étonnant mais est en fait imputable à notre méthode de mesure de la sélection. D'une part, la sélection a été mesurée dans des plantes inoculées avec les 5 variants G, N, K, KN et GK. Dans la grande majorité des lignées HD, les variants G et N sont contre-sélectionnés au profit des variants K, KN et GK. Une grande valeur de s signifie donc que G et N ont un faible taux d'accroissement dans la lignée HD. D'autre part, lors de l'expérience visant à mesurer le taux de contournement de $pvr2^3$, les variants contournant les plus fréquents étaient probablement les simples mutants G et N car une transition est à l'origine de ces mutations (contrairement au variant K) et les transitions sont plus fréquentes que les transversions. Ceci est d'autant plus probable pour les plantes présentant une faible

accumulation virale et pour lesquelles la probabilité d'apparition du mutant K est encore plus faible. De ce fait, on peut imaginer que lorsque ces variants apparaissent dans une plante à faible VA, ils ont un faible taux d'accroissement. Ils augmenteront donc lentement en fréquence et seront potentiellement éliminés par dérive génétique avant d'avoir pu se fixer. Malgré le grand coefficient de sélection mesuré chez ces lignées, c'est donc bien le faible coefficient de sélection des variants G et N associé à une forte dérive génétique qui permet de diminuer la fréquence de contournement. A l'inverse, pour des petites valeurs de *s*, les 5 variants ont des taux d'accroissement similaires. On peut donc imaginer que les mutants G et N seront sélectionnés plus rapidement dans des lignées présentant de petits *s*.

Enfin, ces résultats sont prometteurs pour une utilisation des forces évolutives imposées par la plante pour contrôler l'évolution des virus car ils suggèrent qu'il serait possible de prédire le devenir de la population virale en fonction du génotype de plante infecté. En effet, nos deux études s'accordent sur le fait qu'une plante réduisant conjointement l'accumulation virale et la taille efficace de population à différentes étapes de l'infection protège d'une part le gène majeur du contournement, et d'autre part contraint l'évolution de la population en empêchant un gain d'adaptation lorsque le virus possède déjà la capacité à contourner le gène majeur.

3 Par quels mécanismes le fonds génétique contraint-il l'évolution des pathogènes ?

Nous venons de voir que nos travaux ont démontré qu'il était possible de contraindre l'évolution des virus. Nos résultats montrent que cette contrainte est en partie due à des mécanismes stochastiques imposés par la plante, à savoir les goulets d'étranglement (*i.e.* les faibles N_e globaux mesurés) subis par les populations de PVY ainsi que potentiellement le cliquet de Muller qui peut en découler.

3.1 Effet des goulets d'étranglement sur les mutations adaptatives

Afin d'expliquer l'absence d'adaptation des populations virales aux lignées HD ainsi que l'absence de contournement du gène majeur chez certaines lignées, nous avons évoqué les nombreux goulets d'étranglement subis par les populations virales au cours de l'infection (chapitre 3, chapitre 4). Ces forts goulets d'étranglement entraineraient une diminution du N_e et donc une forte dérive génétique chez les populations virales. Cette forte dérive génétique serait à l'origine de l'absence d'adaptation observée chez des populations évoluant sur des

lignées à faibles N_e et VA. Plusieurs études théoriques ont mis en évidence que des goulets d'étranglement étroits réduisaient fortement la probabilité de survie et de fixation de mutations avantageuses dans les populations soumises à ces goulets (Wahl and Gerrish 2001; Wahl et al. 2002). D'autres études ont confirmé ces résultats par des approches d'évolution expérimentale en imposant différentes tailles de goulets d'étranglement aux populations étudiées. Par exemple, l'adaptation des populations de l'algue unicellulaire *Chlamydomonas reinhardtii* à des environnement à hautes teneurs en sel est déterministe au-delà d'un N_e de $4x10^5$, alors que pour un N_e de $5x10^3$ la sélection est moins efficace et le gain en valeur sélective plus faible chez ces populations (Lachapelle et al. 2015). Pour le bactériophage ID11, le passage d'une dynamique d'évolution déterministe à stochastique s'effectue entre des N_e de 10^4 et 10^5 (Miller et al. 2011). Dans notre cas, les valeurs de N_e mesurées sont plus faibles, et la transition entre ces deux états peut être estimée entre des N_e de 121 et 462 (chapitre 4, tableaux 1 et S1). D'autres études ont également obtenu des résultats plus nuancés concernant l'effet de la taille des goulets d'étranglement sur la capacité d'adaptation du pathogène. Par exemple, récemment Vogwill et al. (2016) ont démontré que chez la bactérie *Pseudomonas fluorescens*, les mutations adaptatives apparaissaient plus fréquemment lorsque les populations étaient soumises à de forts ou à de faibles goulets d'étranglement, et moins fréquemment pour des goulets de tailles intermédiaires.

L'originalité de nos travaux réside dans le fait que les goulets d'étranglement n'ont pas été imposés artificiellement par un expérimentateur mais par la plante elle-même. Ces goulets d'étranglement ont lieu lors de l'étape d'inoculation, avec la diminution du nombre de foyers d'infection primaires, mais également lors de la colonisation de la plante, comme l'attestent les faibles N_e mesurés chez certaines lignées HD. La succession de goulets d'étranglement imposés par l'ouverture/la fermeture des plasmodesmes ou encore les défenses de la plante lors du passage des particules virales d'un organe à un autre voire même d'une cellule à une autre est donc un premier mécanisme potentiellement à l'origine de la forte dérive génétique agissant sur les virus dans notre étude. Bien que, comme nous l'avons vu dans l'introduction de ce manuscrit, peu de données soient disponibles sur l'effet de ces goulets d'étranglement intra-plante chez d'autres pathogènes que les virus, il est fort probable qu'ils agissent également et entraînent des effets de dérive génétique chez d'autres espèces de pathogènes. Estimer l'effet de la taille des goulets d'étranglement chez ces autres espèces de pathogènes au sein d'un hôte, et non en milieu de culture, serait particulièrement intéressant pour évaluer la généricité de ce mécanisme.

3.2 Peut-on confirmer l'action du cliquet de Muller ?

Nous avions émis l'hypothèse que le cliquet de Muller imposé par le fonds génétique de la plante serait à l'origine des extinctions de PVY observées dans le chapitre 4. Ainsi, les résultats de l'évolution expérimentale montrent que lorsque les populations virales sont mal adaptées (faible VA) et sont soumises à des lignées imposant une forte dérive génétique (faible N_e), les populations ne sont pas en mesure de s'adapter au génotype de plante et peuvent même finir par s'éteindre. Avant de valider l'effet du cliquet de Muller, plusieurs conditions doivent être remplies. Tout d'abord, ce mécanisme ne peut avoir lieu qu'en l'absence de recombinaison car cette dernière pourrait permettre de recréer des variants sans mutations délétères. On peut émettre l'hypothèse que le taux de recombinaison dans notre expérience est faible. En effet, la création d'un nouveau variant par recombinaison nécessite la co-infection d'une cellule par deux variants viraux différant par au moins une mutation. Or, de nombreuses études ont rapporté que la colonisation d'une feuille par des virus à ARN génétiquement proches induisait une structuration spatiale de la population virale, les variants n'étant pas ou très peu fréquemment retrouvés dans la même cellule (Dietrich and Maiss 2003; Takeshita et al. 2004; Zwart et al. 2011; Gutiérrez et al. 2012b; Tromas et al. 2014). Au cours de notre expérience, les variants diffèrent par une ou deux mutations dans la VPg, et potentiellement par d'autres mutations dans le reste du génome qui n'a pas été séquencé. Les variants demeurent cependant très proches génétiquement, et la probabilité de co-infection, et par extension de recombinaison, est donc faible. Le cliquet de Muller postule également l'absence de réversions, qui élimineraient une mutation délétère, ainsi que l'absence de mutations qui compenseraient l'effet des mutations délétères sur la valeur sélective du virus. La probabilité qu'un événement de mutation se produise deux fois sur la même base est très faible ce qui ne remet pas en cause la possible action du cliquet de Muller. Néanmoins, il est plus difficile d'évaluer la proportion de mutations compensatoires qui pourraient se produire et nous faisons donc l'hypothèse qu'elles ne sont pas assez fréquentes pour systématiquement empêcher l'action du cliquet de Muller. Lors d'une expérience d'accumulation de mutations en laboratoire avec la levure *Saccharomyces cerevisiae*, seulement 2 populations sur 12 se sont éteintes par l'action du cliquet de Muller. Parmi les populations de levure qui se sont maintenues, certaines montraient des gains de valeur sélective, probablement grâce à des mutations compensatoires (Zeyl and DeVisser 2001). Si le cliquet de Muller a bien lieu dans nos lignées de piments, il est possible que le maintien de certaines populations virales au cours de l'expérience soit dû à la présence de mutations compensatoires uniquement chez les populations non éteintes, alors que d'autres

populations évoluant sur la même lignée se sont éteintes. Enfin, le cliquet de Muller nécessite qu'une accumulation de mutations délétères aient lieu. Nous avons uniquement séquencé la partie centrale de la VPg car il était probable que des mutations avantageuses se soient fixées dans cette région. En séquençant l'ensemble du génome des populations adaptées et des populations juste avant extinction, il serait potentiellement possible d'identifier des différences dues à l'accumulation de mutations délétères chez les populations éteintes. Les études portant sur le cliquet de Muller ne vont pas toujours jusqu'au séquençage des génomes des populations et se contentent parfois de prouver qu'une diminution de valeur sélective a eu lieu dans les populations de pathogènes. Les travaux de Escarmís et al. (1996, 2002) sont tout de même allés jusqu'à la détection de mutations délétères suite à des transferts de plage de lyse en plage de lyse du *Foot-and-mouth disease virus* (FMDV).

4 Conséquences pour la sélection variétale et l'application en champs

Le gène majeur *pvr2*^{3} présente une faible durabilité et peut être facilement contourné par le PVY, plusieurs mutations indépendantes dans la partie centrale de la VPg étant suffisantes au contournement (Ayme et al. 2006). Néanmoins, des études en conditions contrôlées ont prouvé que le fonds génétique de la plante pouvait protéger ce gène majeur du contournement via plusieurs mécanismes (Palloix et al. 2009; Quenouille et al. 2013a). Les travaux réalisés au cours de cette thèse ont démontré que la sélection et la dérive génétique faisaient partie de ces mécanismes. Bien que prometteurs pour parvenir au déploiement de résistances durables, ces résultats ne garantissent pas l'efficacité de ces mécanismes à l'échelle de la parcelle.

L'utilisation de la sélection imposée par le fonds génétique de la plante en sélection variétale n'est pas simple. Il faudrait dans un premier temps sélectionner des variétés sur la base d'un critère fiable permettant de différencier les plantes en fonction de l'intensité de sélection qu'elles imposent. Or, la mesure de ce caractère est fortement dépendante de la composition de la population de pathogènes et il est difficile de prédire à l'avance quels seront les variants rencontrés en champ et quels seront leurs taux d'accroissement respectifs. Les résultats du chapitre 3 montrent ainsi que des lignées de piment évaluées comme imposant une forte pression de sélection (grand *s*) en présence de certains variants peuvent également induire une faible pression de sélection en présence d'autres variants.

Par ailleurs, nous n'avons pas mis en évidence d'effets de sélection antagonistes : les variants viraux sélectionnés sont toujours les mêmes (GK, KN et K), c'est seulement l'intensité avec laquelle ils sont sélectionnés qui est variable selon les lignées. Une telle sélection

antagoniste aurait pourtant pu être très intéressante pour forcer les virus à des compromis évolutifs. La mesure de la sélection a seulement été effectuée sur une population biparentale. En élargissant la recherche dans notre core-collection, de nouveaux allèles contrôlant ces effets de sélection par le fonds génétique seraient peut-être détectés et des effets de sélection antagonistes pourraient être révélés. Par exemple, chez le pommier, il a été démontré que des QTL à spectre d'action étroit induisaient des pressions de sélection différentielle sur des populations de *Venturia inaequalis* (Lê Van et al. 2013). Cependant, des résultats suggèrent que de tels effets ne sont peut-être tout simplement pas induits par le fonds génétique du piment. J'ai également inoculé le mélange de 5 variants de PVY à 15 lignées de piments dépourvues du gène majeur *pvr2*^{3} afin de ne détecter que l'effet du fonds génétique sur l'évolution des fréquences des variants viraux (données non présentées). Sur ces 15 lignées, les coefficients de sélection étaient très faibles, contrairement aux effets de dérive génétique qui étaient variables d'une lignée à une autre. Toutes proportions gardées compte tenu du faible nombre de lignées, il est donc possible que des effets forts (en terme d'amplitude) et variés (en terme de direction) de sélection ne soient pas induits par le fonds génétique.

On peut émettre l'hypothèse que si des effets antagonistes de sélection n'ont pas été détectés, c'est parce que ce type d'effet est potentiellement plus répandu dans le cas de gènes majeurs de résistance. Ainsi, il serait envisageable de rechercher d'autres allèles de résistance au locus *pvr2* dans la collection de piment et d'évaluer s'ils occasionnent des effets de sélection antagonistes après inoculation d'un mélange de variants de PVY. On pourrait aussi élargir la recherche à d'autres espèces de plantes hôtes qui permettraient d'établir de la sélection disruptive en champ. L'utilisation d'un premier gène de résistance sélectionnerait un variant viral et l'utilisation d'un second gène de résistance (dans le temps ou dans l'espace) contre-sélectionnerait ce variant. Des approches de modélisation permettraient par ailleurs d'évaluer les meilleures stratégies de déploiement de ces gènes de résistance à l'échelle de la parcelle. Ce type de pratique a toutefois ses limites et pourrait se révéler contreproductif. En effet, la sélection imposée par une plante hôte peut parfois entraîner la fixation de mutations à effets pléiotropiques chez le pathogène, lui permettant d'infecter une autre espèce d'hôte ou d'acquérir plus facilement de nouvelles mutations lui conférant l'adaptation à ce nouvel hôte, comme cela a été prouvé avec le PVY chez les Solanacées (Moury et al. 2014).

De par leur nature stochastique, nous pouvons émettre l'hypothèse que les effets de dérive génétique des populations de pathogènes contrôlés génétiquement par les plantes ont un spectre d'action relativement large. En effet, par définition, la dérive génétique agit indépendamment

et de manière aléatoire sur l'ensemble des variants composant la population de pathogènes. Ce caractère devrait donc être moins dépendant de la composition de la population de pathogènes que la sélection, et son effet devrait être plus facilement prédictible. La cartographie des QTL de dérive génétique à l'inoculation réalisée dans le chapitre 1 tend à confirmer cette hypothèse. Des QTL contrôlant à la fois les N_e du PVY et du CMV, et potentiellement du ToMV, ont ainsi été identifiés. Par ailleurs, l'action de ces QTL pourrait être multiple, comme le QTL identifié sur le chromosome 6 qui diminue à la fois N_e à l'inoculation et l'accumulation virale dans la plante entière. Les QTL contrôlant N_e à l'inoculation présentent également plusieurs avantages pour une utilisation en sélection variétale. Qu'il ait été mesuré sur une population biparentale ou sur une collection de piments, le nombre de foyers à l'inoculation engendrés par le PVY a toujours été extrêmement héritable (h^2 = 0,93 et 0,98). Le nombre de lésions locales induites par le CMV était également un caractère hautement héritable (h^2 = 0,98), ce qui permettrait de l'utiliser comme critère de sélection. De plus, la mesure du caractère est relativement aisée et nécessite uniquement de dénombrer des foyers ou des lésions sur des feuilles. Il serait même envisageable d'appliquer de tels tests à d'autres pathogènes que les virus si des symptômes similaires sont causés par ces pathogènes.

Le déploiement de QTL affectant la dérive génétique à l'inoculation en champ serait donc une stratégie prometteuse pour protéger le gène majeur du contournement. Néanmoins, plusieurs conditions nécessaires à leur utilisation doivent encore être réunies. Tout d'abord, les QTL contrôlant la dérive génétique à l'inoculation ont été caractérisés en inoculant manuellement les virus. Leur efficacité à l'échelle du paysage n'est donc pas garantie puisque les principaux vecteurs du virus au champ sont les pucerons. Il serait donc pertinent d'étudier si les QTL de dérive génétique détectés en inoculation manuelle agissent également lorsque l'inoculation est effectuée par puceron. De plus, le phénotypage des lignées HD a été réalisé en conditions contrôlées et l'impact de l'environnement sur ces QTL n'est pas connu. L'efficacité des QTL de dérive génétique en champ devrait donc être testée. Des résultats encourageants ont tout de même été rapportés par une étude portant sur la création de variétés de piments porteuses de résistances partielles au CMV (Djian-Caporalino et al. 2006). Les géniteurs utilisés étaient porteurs de trois sources de résistance à différentes étapes du cycle infectieux : l'installation du virus dans les cellules de l'hôte, l'accumulation du virus dans l'hôte et le mouvement systémique (Figure 2a). L'installation dans les cellules de l'hôte a été mesurée en quantifiant le nombre de lésions locales induites par le CMV, ce qui correspond donc au N_e à l'inoculation. Des rétrocroisements avec des variétés à gros fruits suivis par des étapes

d'autofécondation ont permis d'obtenir des lignées pures porteuses des résistances seules ou combinées. Chez les lignées combinant la résistance au mouvement systémique et la capacité à réduire l'accumulation virale, l'infection est largement retardée et concerne entre 50 et 90 % des plantes inoculées (Figure 2b). Lorsque la réduction du N_e à l'inoculation est associée aux deux autres types de résistances dans une même plante, la résistance est beaucoup plus importante puisque l'infection n'a pas lieu chez certaines lignées et ne dépasse jamais 30 % des plantes inoculées (Figure 2c). En champ, ces génotypes sont proches de la résistance totale, attestant de l'effet bénéfique que peut conférer la combinaison de ces différents types de résistance, en particulier la réduction du N_e à l'inoculation et de l'accumulation virale.

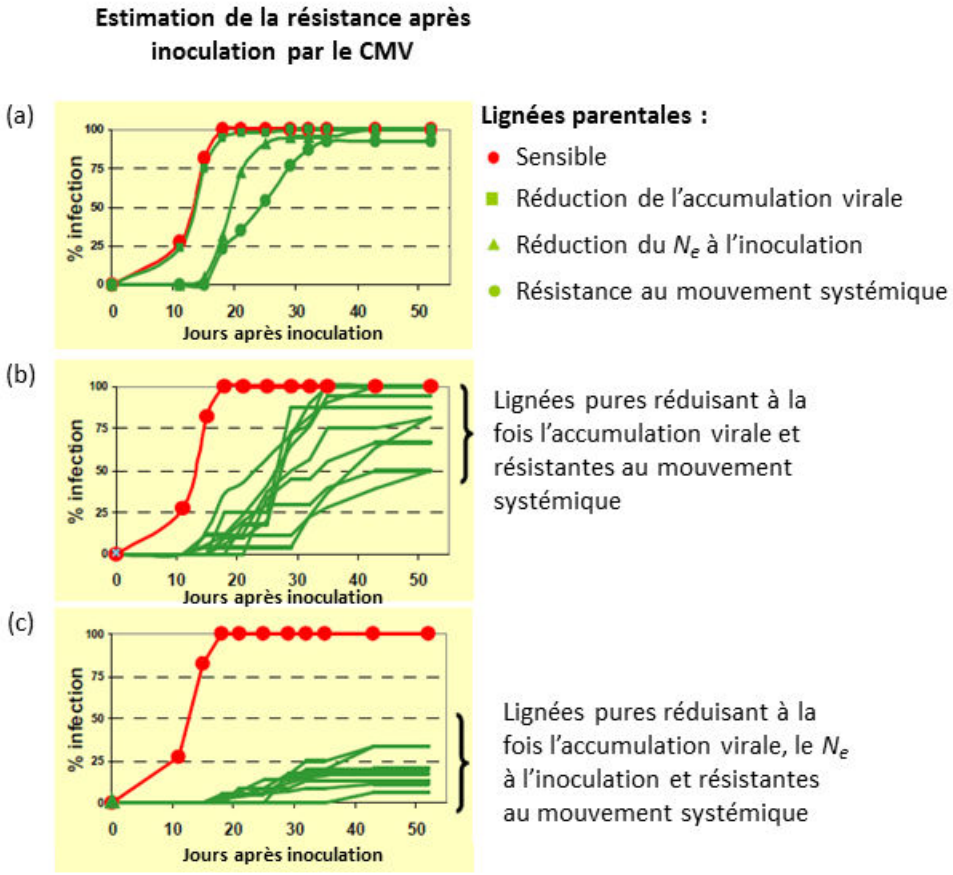


Figure 2 : Création de lignées résistantes au CMV chez le piment. Pourcentage de plantes infectées par le CMV (a) chez les lignées parentales sensibles (rouge) et porteuses de résistances partielles (vert), (b) chez les lignées combinant deux types de résistances (vert) comparées au parent sensible (rouge), (c) chez les lignées combinant les trois types de résistances (vert) comparées au parent sensible (rouge). Modifié d'après Djian-Caporalino et al. (2006).

5 Conclusion générale

En conclusion, cette thèse améliore nos connaissances sur les mécanismes mis en place par le fonds génétique de la plante pour protéger un gène majeur de résistance du contournement. D'une part, elle a permis d'identifier des déterminants génétiques de la plante qui influencent directement le potentiel évolutif du pathogène. D'autre part, elle a permis de démontrer que l'action combinée de certaines forces évolutives induites par la plante permettait de contraindre l'évolution du pathogène, d'empêcher le contournement du gène majeur voire d'entraîner l'extinction des populations d'agents pathogènes. Ces résultats ouvrent de nouvelles perspectives pour le déploiement de déterminants génétiques limitant l'évolution du pathogène et permettant de préserver la durabilité des gènes majeurs de résistance.

Références bibliographiques

A

Abang, M. M., Baum, M., Ceccarelli, S., Grando, S., Linde, C. C., Yahyaoui, A., Zhan, J. and McDonald, B. A. (2006). Differential selection on *Rhynchosporium secalis* during parasitic and saprophytic phases in the barley scald disease cycle. *Phytopathology* **96**, 1214–1222.

Abel, S., Abel zur Wiesch, P., Davis, B. M. and Waldor, M. K. (2015). Analysis of Bottlenecks in Experimental Models of Infection. *PLoS Pathog.* **11**, e1004823.

Acosta-Leal, R. and Xiong, Z. (2008). Complementary functions of two recessive R-genes determine resistance durability of tobacco 'Virgin A Mutant' (VAM) to *Potato virus Y*. *Virology* **379**, 275–283.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723.

Ali, S., Soubeyrand, S., Gladieux, P., Giraud, T., Leconte, M., Gautier, A., Mboup, M., Chen, W., Vallavieille-Pope, C. and Enjalbert, J. (2016). CLONCASE: estimation of sex frequency and effective population size by clonemate resampling in partially clonal organisms. *Mol. Ecol. Resour.* **16**, 845–861.

Anderson, E. C., Williamson, E. G. and Thompson, E. A. (2000). Monte Carlo evaluation of the likelihood for N_e from temporally spaced samples. *Genetics* **156**, 2109–2118.

Ayme, V., Souche, S., Caranta, C., Jacquemond, M., Chadœuf, J., Palloix, A. and Moury, B. (2006). Different mutations in the genome-linked protein VPg of *Potato virus Y* confer virulence on the *pvr2*^{3} resistance in pepper. *Mol. Plant. Microbe Interact.* **19**, 557–563.

Ayme, V., Petit-Pierre, J., Souche, S., Palloix, A. and Moury, B. (2007). Molecular dissection of the *Potato virus Y* VPg virulence factor reveals complex adaptations to the *pvr2* resistance allelic series in pepper. *J. Gen. Virol.* **88**, 1594–1601.

B

Bald, J. G. (1937). The use of numbers of infections for comparing the concentration of plant virus suspensions: dilution experiments with purified suspensions. *Ann. Appl. Biol.* **24**, 33–55.

Barchi, L., Lefebvre, V., Sage-Palloix, A.-M., Lanteri, S. and Palloix, A. (2009). QTL analysis of plant development and fruit traits in pepper and performance of selective phenotyping. *Theor. Appl. Genet.* **118**, 1157–1171.

Barrick, J. E. and Lenski, R. E. (2013). Genome dynamics during experimental evolution. *Nat. Rev. Genet.* **14**, 827.

Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E. and Kim, J. F. (2009). Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243.

Barrick, J. E., Kauth, M. R., Strelioff, C. C. and Lenski, R. E. (2010). *Escherichia coli* rpoB mutants have increased evolvability in proportion to their fitness defects. *Mol. Biol. Evol.* **27**, 1338–1347.

Bataillon, T., Dillmann, C., Gaba, S., Goldringer, I., Kaltz, O., Méry, F., Nidelet, T., Reboud, X., Schneider, D. and Sicard, D. (2010). Evolution expérimentale. In *Biologie évolutive*. De Boeck Superieur.

Bedhomme, S., Lafforgue, G. and Elena, S. F. (2012). Multihost Experimental Evolution of a Plant RNA Virus Reveals Local Adaptation and Host-Specific Mutations. *Mol. Biol. Evol.* **29**, 1481–1492.

Bergstrom, C. T., McElhany, P. and Real, L. A. (1999). Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. *Proc. Natl. Acad. Sci. USA.* **96**, 5095–5100.

Bergua, M., Zwart, M. P., El-Mohtar, C., Shilts, T., Elena, S. F. and Folimonova, S. Y. (2014). A viral protein mediates superinfection exclusion at the whole-organism level but is not required for exclusion at the cellular level. *J. Virol.* **88**, 11327–11338.

Berthier, P., Beaumont, M. A., Cornuet, J.-M. and Luikart, G. (2002). Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics* **160**, 741–751.

Betancourt, M., Fereres, A., Fraile, A. and Garcia-Arenal, F. (2008). Estimation of the Effective Number of Founders That Initiate an Infection after Aphid Transmission of a Multipartite Plant Virus. *J. Virol.* **82**, 12416–12421.

Boissot, N., Schoeny, A. and Vanlerberghe-Masutti, F. (2016). *Vat*, an amazing gene conferring resistance to aphids and viruses they carry: from molecular structure to field effects. *Front. Plant Sci.* **7**, 1-18.

Bollback, J. P., York, T. L. and Nielsen, R. (2008). Estimation of $2N_e s$ from temporal allele frequency data. *Genetics* **179**, 497–502.

Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C. S., Bliek, M., Boersma, M. R., Borghi, L. and Bruggmann, R. (2016). Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants* **2**, 16074.

Brachi, B., Morris, G. P. and Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* **12**, 232.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635.

Broman, K. W., Wu, H., Sen, Ś. and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890.

Brown, J. K. M. (2015). Durable Resistance of Crops to Disease: A Darwinian Perspective. *Annu. Rev. Phytopathol.* **53**, 513–539.

Brun, H., Chèvre, A.-M., Fitt, B. D., Powers, S., Besnard, A.-L., Ermel, M., Huteau, V., Marquer, B., Eber, F., Renard, M., et al. (2010). Quantitative resistance increases the durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. *New Phytol.* **185**, 285–299.

Burke, M. K., Dunham, J. P., Shahrestani, P., Thornton, K. R., Rose, M. R. and Long, A. D. (2010). Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* **467**, 587.

Bustin, S. A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* **25**, 169–193.

C

Caballero, A. (1994). Developments in the prediction of effective population size. *Heredity* **73**, 657–679.

Caffier, V., Lasserre-Zuber, P., Giraud, M., Lascostes, M., Stievenard, R., Lemarquand, A., van de Weg, E., Expert, P., Denancé, C., Didelot, F., et al. (2014). Erosion of quantitative host resistance in the apple×*Venturia inaequalis* pathosystem. *Infect. Genet. Evol.* **27**, 481–489.

Caldwell, K. S., Russell, J., Langridge, P. and Powell, W. (2006). Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics* **172**, 557–567.

Caranta, C. and Palloix, A. (1996). Both common and specific genetic factors are involved in polygenic resistance of pepper to several potyviruses. *Theor. Appl. Genet.* **92**, 15–20.

Caranta, C., Lefebvre, V. and Palloix, A. (1997a). Polygenic resistance of pepper to potyviruses consists of a combination of isolate-specific and broad-spectrum quantitative trait loci. *Mol. Plant. Microbe Interact.* **10**, 872–878.

Caranta, C., Palloix, A., Lefebvre, V. and Daubeze, A. M. (1997b). QTLs for a component of partial resistance to *Cucumber mosaic virus* in pepper: restriction of virus installation in host-cells. *Theor. Appl. Genet.* **94**, 431–438.

Carrasco, P., de la Iglesia, F. and Elena, S. F. (2007). Distribution of fitness and virulence effects caused by single-nucleotide substitutions in *Tobacco etch virus*. *J. Virol.* **81**, 12979–12984.

Chain, F., Riault, G., Trottet, M. and Jacquot, E. (2007). Evaluation of the durability of the *Barley yellow dwarf virus*-resistant Zhong ZH and TC14 wheat lines. *Eur. J. Plant Pathol.* **117**, 35–43.

Chao, L. (1990). Fitness of RNA virus decreased by Muller's rachet. *Nature* **348**, 454–455.

Charlesworth, B. (1978). Model for evolution of Y chromosomes and dosage compensation. *Proc. Natl. Acad. Sci. USA.* **75**, 5618–5622.

Charlesworth, B. (2009). Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205.

Charlesworth, B., Charlesworth, D. and Barton, N. H. (2003). The effects of genetic and geographic structure on neutral variation. *Annu. Rev. Ecol. Evol. Syst.* **34**, 99–125.

Charron, C., Nicolaï, M., Gallois, J.-L., Robaglia, C., Moury, B., Palloix, A. and Caranta, C. (2008). Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg: Co-evolution between eIF4E and VPg. *Plant J.* **54**, 56–68.

Chattopadhyay, S., Ali, K. A., Doss, S. G., Das, N. K., Aggarwal, R. K., Bandopadhyay, T. K., Sarkar, A. and Bajpai, A. (2011). Association of leaf micro-morphological characters with powdery mildew resistance in field-grown mulberry (*Morus* spp.) germplasm. *AoB Plants* **2011**, plr002.

Ching, A., Caldwell, K. S., Jung, M., Dolan, M., Smith, O., Tingey, S., Morgante, M. and Rafalski, A. J. (2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet.* **3**, 19.

Chung, B. Y.-W., Miller, W. A., Atkins, J. F. and Firth, A. E. (2008). An overlapping essential gene in the Potyviridae. *Proc. Natl. Acad. Sci. USA.* **105**, 5897–5902.

Cormier, F., Le Gouis, J., Dubreuil, P., Lafarge, S. and Praud, S. (2014). A genome-wide identification of chromosomal regions determining nitrogen use efficiency components in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **127**, 2679–2693.

Crow, J. F. and Denniston, C. (1988). Inbreeding and variance effective population numbers. *Evolution* **42**, 482–495.

Cruz, C. M. V., Bai, J., Oña, I., Leung, H., Nelson, R. J., Mew, T.-W. and Leach, J. E. (2000). Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. *Proc. Natl. Acad. Sci. USA.* **97**, 13500–13505.

D

da Silva, J. and Wyatt, S. (2014). Fitness valleys constrain HIV-1's adaptation to its secondary chemokine coreceptor. *J. Evol. Biol.* **27**, 604–615.

de la Iglesia, F. and Elena, S. F. (2007). Fitness Declines in *Tobacco Etch Virus* upon Serial Bottleneck Transfers. *J. Virol.* **81**, 4941–4947.

de la Iglesia, F., Martínez, F., Hillung, J., Cuevas, J. M., Gerrish, P. J., Daròs, J.-A. and Elena, S. F. (2012). Luria-Delbrück estimation of *Turnip mosaic virus* mutation rate in vivo. *J. Virol.* **86**, 3386–3388.

Díaz, J. A., Nieto, C., Moriones, E., Truniger, V. and Aranda, M. A. (2004). Molecular characterization of a *Melon necrotic spot virus* strain that overcomes the resistance in melon and nonhost plants. *Mol. Plant. Microbe Interact.* **17**, 668–675.

Dietrich, C. and Maiss, E. (2003). Fluorescent labelling reveals spatial separation of potyvirus populations in mixed infected *Nicotiana benthamiana* plants. *J. Gen. Virol.* **84**, 2871–2876.

Djian-Caporalino, C., Lefebvre, V., Sage-Daubèze, A.-M. and Palloix, A. (2006). *Capsicum*. In *Genetic Resources, Chromosome Engineering, and Crop Improvement: Vegetable crops*, pp. 185–243. R.J. Singh, ed. (Florida, USA: CRC Press).

Dogimont, C., Palloix, A., Daubze, A.-M., Marchoux, G., Selassie, K. G. and Pochard, E. (1996). Genetic analysis of broad spectrum resistance to potyviruses using doubled haploid lines of pepper (*Capsicum annuum* L.). *Euphytica* **88**, 231–239.

Dowkiw, A., Voisin, E. and Bastien, C. (2010). Potential of Eurasian poplar rust to overcome a major quantitative resistance factor. *Plant Pathol.* **59**, 523–534.

Duarte, E., Clarke, D., Moya, A., Domingo, E. and Holland, J. (1992). Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proc. Natl. Acad. Sci. USA.* **89**, 6015–6019.

E

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361.

Elena, S. F. and Lenski, R. E. (2003). Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* **4**, 457.

Elena, S. F. and Sanjuán, R. (2007). Virus Evolution: Insights from an Experimental Approach. *Annu. Rev. Ecol. Evol. Syst.* **38**, 27–52.

Elena, S. F., Bedhomme, S., Carrasco, P., Cuevas, J. M., De la Iglesia, F., Lafforgue, G., Lalic, J., Pròsper, À., Tromas, N. and Zwart, M. P. (2011). The evolutionary genetics of emerging plant RNA viruses. *Mol. Plant. Microbe Interact.* **24**, 287–293.

Engelstädter, J. (2008). Muller's ratchet and the degeneration of Y chromosomes: A simulation study. *Genetics* **180**, 957–967.

Ersoz, E. S., Yu, J. and Buckler, E. S. (2007). Applications of linkage disequilibrium and association mapping in crop plants. *Genomics-Assist. Crop Improv.* 97–119.

Escarmís, C., Dávila, M., Charpentier, N., Bracho, A., Moya, A. and Domingo, E. (1996). Genetic lesions associated with Muller's ratchet in an RNA virus. *J. Mol. Biol.* **264**, 255–267.

Escarmís, C., Gómez-Mariano, G., Dávila, M., Lázaro, E. and Domingo, E. (2002). Resistance to extinction of low fitness virus subjected to plaque-to-plaque transfers: diversification by mutation clustering. *J. Mol. Biol.* **315**, 647–661.

Escriu, F., Fraile, A. and García-Arenal, F. (2003). The evolution of virulence in a plant virus. *Evolution* **57**, 755–765.

Evanno, G., Regnaut, S. and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620.

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3**, 87–112.

Ewens, W. (1982). On the concept of the effective population size. *Theor. Popul. Biol.* **21**, 373–378.

Ewens, W. J. (2004). Mathematical population genetics. I. Theoretical introduction. Interdisciplinary applied mathematics, 27.

F

Fabre, F., Chadœuf, J., Costa, C., Lecoq, H. and Desbiez, C. (2010). Asymmetrical over-infection as a process of plant virus emergence. *J. Theor. Biol.* **265**, 377–388.

Fabre, F., Bruchou, C., Palloix, A. and Moury, B. (2009). Key determinants of resistance durability to plant viruses: Insights from a model linking within- and between-host dynamics. *Virus Res.* **141**, 140–149.

Fabre, F., Rousseau, E., Mailleret, L. and Moury, B. (2012a). Durable strategies to deploy plant resistance in agricultural landscapes. *New Phytol.* **193**, 1064–1075.

Fabre, F., Montarry, J., Coville, J., Senoussi, R., Simon, V. and Moury, B. (2012b). Modelling the Evolutionary Dynamics of Viruses within Their Hosts: A Case Study Using High-Throughput Sequencing. *PLoS Pathog.* **8**, e1002654.

Fabre, F., Moury, B., Johansen, E. I., Simon, V., Jacquemond, M. and Senoussi, R. (2014). Narrow bottlenecks affect *Pea seedborne mosaic virus* populations during vertical seed transmission but not during leaf colonization. *PLoS Pathog.* **10**, e1003833.

Fabre, F., Rousseau, E., Mailleret, L. and Moury, B. (2015). Epidemiological and evolutionary management of plant resistance: optimizing the deployment of cultivar mixtures in time and space in agricultural landscapes. *Evol. Appl.* **8**, 919–932.

Feder, A. F., Rhee, S.-Y., Holmes, S. P., Shafer, R. W., Petrov, D. A. and Pennings, P. S. (2016). More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. *Elife* **5**, e10670.

Flint-Garcia, S. A., Thornsberry, J. M. and Buckler IV, E. S. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* **54**, 357–374.

Folimonova, S. Y. (2013). Developing an understanding of cross-protection by *Citrus tristeza virus*. *Front. Microbiol.* **4**, 76.

Foll, M., Shim, H. and Jensen, J. D. (2015). WFABC: a Wright–Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Mol. Ecol. Resour.* **15**, 87–98.

Fournet, S., Kerlan, M. C., Renault, L., Dantec, J. P., Rouaux, C. and Montarry, J. (2013). Selection of nematodes by resistant plants has implications for local adaptation and cross-virulence: Local adaptation and cross-virulence in *Globodera pallida*. *Plant Pathol.* **62**, 184–193.

French, R. and Stenger, D. C. (2003). Evolution of *Wheat streak mosaic virus*: dynamics of population growth within plants may explain limited variation. *Annu. Rev. Phytopathol.* **41**, 199–214.

G

Gago, S., Elena, S. F., Flores, R. and Sanjuán, R. (2009). Extremely high mutation rate of a hammerhead viroid. *Science* **323**, 1308–1308.

García-Arenal, F. and Fraile, A. (2011). Population Dynamics and Genetics of Plant Infection by Viruses. In *Recent Advances in Plant Virology*, pp. 263–281. Norfolk, United Kingdom: Caister Academic Press.

García-Arenal, F. and McDonald, B. A. (2003). An analysis of the durability of resistance to plant viruses. *Phytopathology* **93**, 941–952.

García-Arenal, F., Fraile, A. and Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annu. Rev. Phytopathol.* **39**, 157–186.

García-Arenal, F., Fraile, A. and Malpica, J. M. (2003). Variation and evolution of plant virus populations. *Int. Microbiol.* **6**, 225–232.

Garrett, K. and Mundt, C. (2000). Host diversity can reduce potato late blight severity for focal and general patterns of primary inoculum. *Phytopathology* **90**, 1307–1312.

Gibbs, A., Gibbs, M., Ohshima, K. and Garcia-Arenal, F. (2008a). More about plant virus evolution; past, present and future. In *Origin and Evolution of Viruses,* 229–50.

Gibbs, A. J., Ohshima, K., Phillips, M. J. and Gibbs, M. J. (2008b). The prehistory of potyviruses: their initial radiation was during the dawn of agriculture. *PLoS One* **3**, e2523.

Gómez, P., Rodríguez-Hernández, A., Moury, B. and Aranda, M. (2009). Genetic resistance for the sustainable control of plant virus diseases: breeding, mechanisms and durability. *Eur. J. Plant Pathol.* **125**, 1–22.

González-Jara, P., Fraile, A., Canto, T. and Garcia-Arenal, F. (2009). The multiplicity of infection of a plant virus varies during colonization of its eukaryotic host. *J. Virol.* **83**, 7487–7494.

González-Jara, P., Fraile, A., Canto, T. and García-Arenal, F. (2013). The multiplicity of infection of a plant virus varies during colonization of its eukaryotic host. *J. Virol.* **87**, 2374–2374.

Gordo, I. and Charlesworth, B. (2000). On the speed of Muller's ratchet. *Genetics* **156**, 2137–2140.

Gouesnard, B., Bataillon, T. M., Decoux, G., Rozale, C., Schoen, D. J. and David, J. L. (2001). MSTRAT: An Algorithm for Building Germ Plasm Core Collections by Maximizing Allelic or Phenotypic Richness. *J. Hered.* **92**, 93–94.

Griffin, A. S., West, S. A. and Buckling, A. (2004). Cooperation and competition in pathogenic bacteria. *Nature* **430**, 1024–1027.

Grubaugh, N. D., Weger-Lucarelli, J., Murrieta, R. A., Fauver, J. R., Garcia-Luna, S. M., Prasad, A. N., Black, W. C. and Ebel, G. D. (2016). Genetic drift during systemic arbovirus infection of mosquito vectors leads to decreased relative fitness during host switching. *Cell Host Microbe* **19**, 481–492.

Grube, R., Blauth, J., Andrés, M. A., Caranta, C. and Jahn, M. (2000). Identification and comparative mapping of a dominant potyvirus resistance gene cluster in *Capsicum*. *Theor. Appl. Genet.* **101**, 852–859.

Guidot, A., Jiang, W., Ferdy, J.-B., Thébaud, C., Barberis, P., Gouzy, J. and Genin, S. (2014). Multihost experimental evolution of the pathogen *Ralstonia solanacearum* unveils genes involved in adaptation to plants. *Mol. Biol. Evol.* **31**, 2913–2928.

Gutiérrez, S., Yvon, M., Thébaud, G., Monsion, B., Michalakis, Y. and Blanc, S. (2010). Dynamics of the multiplicity of cellular infection in a plant virus. *PLoS Pathog.* **6**, e1001113.

Gutiérrez, S., Yvon, M., Pirolles, E., Garzo, E., Fereres, A., Michalakis, Y. and Blanc, S. (2012a). Circulating virus load determines the size of bottlenecks in viral populations progressing within a host. *PLoS Pathog.* **8**, e1003009.

Gutiérrez, S., Michalakis, Y. and Blanc, S. (2012b). Virus population bottlenecks during within-host progression and host-to-host transmission. *Curr. Opin. Virol.* **2**, 546–555.

Gutiérrez, S., Pirolles, E., Yvon, M., Baecker, V., Michalakis, Y. and Blanc, S. (2015). The Multiplicity of Cellular Infection Changes Depending on the Route of Cell Infection in a Plant Virus. *J. Virol.* **89**, 9665–9675.

H

Hall, J. S., French, R., Hein, G. L., Morris, T. J. and Stenger, D. C. (2001). Three distinct mechanisms facilitate genetic isolation of sympatric wheat streak mosaic virus lineages. *Virology* **282**, 230–236.

Hamblin, M. T., Buckler, E. S. and Jannink, J.-L. (2011). Population genetics of genomics-based crop improvement methods. *Trends Genet.* **27**, 98–106.

Harrison, B. D. (2002). Virus variation in relation to resistance-breaking in plants. *Euphytica* **124**, 181–192.

Hillung, J., Cuevas, J. M., Valverde, S. and Elena, S. F. (2014). Experimental evolution of an emerging plant virus in host genotypes that differ in their susceptibility to infection. *Evolution* **68**, 2467–2480.

I

Iwasa, Y., Michor, F. and Nowak, M. A. (2004). Stochastic tunnels in evolutionary dynamics. *Genetics* **166**, 1571–1579.

J

Jan, P., Gracianne, C., Fournet, S., Olivier, E., Arnaud, J., Porte, C., Bardou-Valette, S., Denis, M. and Petit, E. J. (2016). Temporal sampling helps unravel the genetic structure of naturally occurring populations of a phytoparasitic nematode. 1. Insights from the estimation of effective population sizes. *Evol. Appl.* **9**, 489–501.

Janzac, B., Fabre, M.-F., Palloix, A. and Moury, B. (2008). Characterization of a new potyvirus infecting pepper crops in Ecuador. *Arch. Virol.* **153**, 1543.

Janzac, B., Fabre, M.-F., Palloix, A. and Moury, B. (2009a). Phenotype and spectrum of action of the *Pvr4* resistance in pepper against potyviruses, and selection for virulent variants. *Plant Pathol.* **58**, 443–449.

Janzac, B., Fabre, F., Palloix, A. and Moury, B. (2009b). Constraints on evolution of virus avirulence factors predict the durability of corresponding plant resistances. *Mol. Plant Pathol.* **10**, 599–610.

Janzac, B., Montarry, J., Palloix, A., Navaud, O. and Moury, B. (2010). A point mutation in the polymerase of *Potato virus Y* confers virulence toward the *Pvr4* resistance of pepper and a high competitiveness cost in susceptible cultivar. *Mol. Plant. Microbe Interact.* **23**, 823–830.

Janzac, B., Willemsen, A., Cuevas, J. M., Glais, L., Tribodet, M., Verrier, J., Elena, S. and Jacquot, E. (2015). Brazilian *Potato virus Y* isolates identified as members of a new clade facilitate the reconstruction of evolutionary traits within this species. *Plant Pathol.* **64**, 799–807.

Jaramillo, N., Domingo, E., Muñoz-Egea, M. C., Tabares, E. and Gadea, I. (2013). Evidence of Muller's ratchet in *Herpes simplex virus* type 1. *J. Gen. Virol.* **94**, 366–375.

Jiang, G., Peyraud, R., Remigi, P., Guidot, A., Ding, W., Genin, S. and Peeters, N. (2016). Modeling and experimental determination of infection bottleneck and within-host dynamics of a soil-borne bacterial plant pathogen. *bioRxiv*.

Jorde, P. E. and Ryman, N. (2007). Unbiased estimator for genetic drift and effective population size. *Genetics* **177**, 927–935.

K

Kerlan, C. (2006). *Potato virus Y*. Descriptions of plant viruses, no. 414. *Wellesbourne UK Assoc. Appl. Biol.*

Khelifa, M., Massé, D., Blanc, S. and Drucker, M. (2010). Evaluation of the minimal replication time of *Cauliflower mosaic virus* in different hosts. *Virology* **396**, 238–245.

Kim, S., Park, M., Yeom, S.-I., Kim, Y.-M., Lee, J. M., Lee, H.-A., Seo, E., Choi, J., Cheong, K., Kim, K.-T., et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat Genet* **46**, 270–278.

Kim, S., Kang, W., Huy, H. N., Yeom, S., An, J., Kim, S., Kang, M., Kim, H. J., Jo, Y. D. and Ha, Y. (2017). Divergent evolution of multiple virus-resistance genes from a progenitor in *Capsicum* spp. *New Phytol.* **213**, 886–899.

Kimura, M. (1970). Stochastic processes in population genetics, with special reference to distribution of gene frequencies and probability of gene fixation. In *Mathematical topics in population genetics*, pp. 178–209. Springer.

Kimura, M. and Crow, J. F. (1963). The measurement of effective population number. *Evolution* **17**, 279–288.

Kimura, M. and Ohta, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763.

Kinkel, L. L., Bakker, M. G. and Schlatter, D. C. (2011). A coevolutionary framework for managing disease-suppressive soils. *Annu. Rev. Phytopathol.* **49**, 47–67.

Kleczkowski, A. (1950). Interpreting relationships between the concentrations of plant viruses and numbers of local lesions. *Microbiology* **4**, 53–69.

Korte, A. and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* **9**, 29.

Kutnjak, D., Elena, S. F. and Ravnikar, M. (2017). Time-sampled population sequencing reveals the interplay of selection and genetic drift in experimental evolution of *Potato virus Y*. *J. Virol.* JVI-00690.

L

Lacerda, M. and Seoighe, C. (2014). Population genetics inference for longitudinally-sampled mutants under strong selection. *Genetics*. **198**, 1237-1250.

Lachapelle, J., Reid, J. and Colegrave, N. (2015). Repeatability of adaptation in experimental populations of different sizes. *Proc. R. Soc. Lond. B Biol. Sci.* **282**, 20143033.

Lafforgue, G., Tromas, N., Elena, S. F. and Zwart, M. P. (2012). Dynamics of the establishment of systemic potyvirus infection: independent yet cumulative action of primary infection sites. *J. Virol.* **86**, 12912–12922.

Lam, L. H. and Monack, D. M. (2014). Intraspecies competition for niches in the distal gut dictate transmission during persistent *Salmonell*a infection. *PLoS Pathog.* **10**, e1004527.

Lande, R. and Barrowclough, G. F. (1987). Effective population size, genetic variation, and their use in population management. *Viable Popul. Conserv.* **87**, 124.

Lanfear, R., Kokko, H. and Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends Ecol. Evol.* **29**, 33–41.

Lannou, C. (2012). Variation and Selection of Quantitative Traits in Plant Pathogens. *Annu. Rev. Phytopathol.* **50**, 319–338.

Lauffer, M. A. and Price, W. (1945). Infection by Viruses. *Arch. Biochem.* 449–450.

Le Guen, V., Garcia, D., Mattos, C. R. R., Doaré, F., Lespinasse, D. and Seguin, M. (2007). Bypassing of a polygenic *Microcyclus ulei* resistance in rubber tree, analyzed by QTL detection. *New Phytol.* **173**, 335–345.

Lê Van, A., Caffier, V., Lasserre-Zuber, P., Chauveau, A., Brunel, D., Le Cam, B. and Durel, C.-E. (2013). Differential selection pressures exerted by host resistance quantitative trait loci on a pathogen population: a case study in an apple × *Venturia inaequalis* pathosystem. *New Phytol.* **197**, 899–908.

Leach, J. E., Vera Cruz, C. M., Bai, J. and Leung, H. (2001). Pathogen fitness penalty as a predictor of durability of disease resistance genes. *Annu. Rev. Phytopathol.* **39**, 187–224.

Lecoq, H., Lemaire, J. and Wipf-Scheibel, C. (1991). Control of *Zucchini yellow mosaic virus* in squash by cross protection. *Plant Dis.* **75**, 208–211.

Lecoq, H., Moury, B., Desbiez, C., Palloix, A. and Pitrat, M. (2004). Durable virus resistance in plants through conventional approaches: a challenge. *Virus Res.* **100**, 31–39.

Lee, J.-Y. and Lu, H. (2011). Plasmodesmata: the battleground against intruders. *Trends Plant Sci.* **16**, 201–210.

Lefevre, T., Raymond, M. and Thomas, F. (2016). *Biologie évolutive*. De Boeck Superieur.

Lenski, R. (2004). Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. *Plant Breed. Rev. Vol-24 Part 2 Long-Term Sel. Crops Anim. Bact.*

Lequime, S., Fontaine, A., Gouilh, M. A., Moltini-Conclois, I. and Lambrechts, L. (2016). Genetic drift, purifying selection and vector genotype shape dengue virus intra-host genetic diversity in mosquitoes. *PLoS Genet.* **12**, e1006111.

Li, H. and Roossinck, M. J. (2004). Genetic Bottlenecks Reduce Population Variation in an Experimental RNA Virus Population. *J. Virol.* **78**, 10582–10587.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S. and Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399.

Luna, A. and Nicodemus, K. K. (2007). snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. *Bioinformatics* **23**, 774–776.

Lynch, M., Bürger, R., Butcher, D. and Gabriel, W. (1993). The mutational meltdown in asexual populations. *J. Hered.* **84**, 339–344.

M

Malaspinas, A.-S., Malaspinas, O., Evans, S. N. and Slatkin, M. (2012). Estimating allele age and selection coefficient from time-serial data. *Genetics* **192**, 599–607.

Malpica, J. M., Fraile, A., Moreno, I., Obies, C. I., Drake, J. W. and García-Arenal, F. (2002). The rate and character of spontaneous mutation in an RNA virus. *Genetics* **162**, 1505–1511.

Marczewski, W., Flis, B., Syller, J., Schäfer-Pregl, R. and Gebhardt, C. (2001). A major quantitative trait locus for resistance to *Potato leafroll virus* is located in a resistance hotspot on potato chromosome XI and is tightly linked to N-gene-like markers. *Mol. Plant. Microbe Interact.* **14**, 1420–1425.

Martínez, F., Sardanyés, J., Elena, S. F. and Daròs, J.-A. (2011). Dynamics of a plant RNA virus intracellular accumulation: stamping machine vs. geometric replication. *Genetics* **188**, 637–646.

Mathieson, I. and McVean, G. (2013). Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics* **193**, 973–984.

McDonald, B. A. and Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* **40**, 349–379.

Melotto, M., Underwood, W., Koczan, J., Nomura, K. and He, S. Y. (2006). Plant stomata function in innate immunity against bacterial invasion. *Cell* **126**, 969–980.

Miller, C. R., Joyce, P. and Wichman, H. A. (2011). Mutational effects and population dynamics during viral adaptation challenge current models. *Genetics* **187**, 185–202.

Miralles, R., Gerrish, P. J., Moya, A. and Elena, S. F. (1999). Clonal interference and the evolution of RNA viruses. *Science* **285**, 1745–1747.

Miras, M., Sempere, R. N., Kraft, J. J., Miller, W. A., Aranda, M. A. and Truniger, V. (2014). Interfamilial recombination between viruses led to acquisition of a novel translation-enhancing RNA element that allows resistance breaking. *New Phytol.* **202**, 233–246.

Miyashita, S. and Kishino, H. (2010). Estimation of the size of genetic bottlenecks in cell-to-cell movement of *Soil-borne wheat mosaic virus* and the possible role of the bottlenecks in speeding up selection of variations in trans-acting genes or elements. *J. Virol.* **84**, 1828–1837.

Monsion, B., Froissart, R., Michalakis, Y. and Blanc, S. (2008). Large bottleneck size in *Cauliflower mosaic virus* populations during host plant colonization. *PLoS Pathog.* **4**, e1000174.

Montarry, J., Doumayrou, J., Simon, V. and Moury, B. (2011). Genetic background matters: a plant-virus gene-for-gene interaction is strongly influenced by genetic contexts: Viral genetic background matters. *Mol. Plant Pathol.* **12**, 911–920.

Montarry, J., Cartier, E., Jacquemond, M., Palloix, A. and Moury, B. (2012). Virus adaptation to quantitative plant resistance: erosion or breakdown? *J. Evol. Biol.* **25**, 2242–2252.

Moscone, E. A., Scaldaferro, M. A., Grabiele, M., Cecchini, N. M., Sánchez García, Y., Jarret, R., Daviña, J. R., Ducasse, D. A., Barboza, G. E. and Ehrendorfer, F. (2006). The evolution of chili peppers (*Capsicum*-Solanaceae): a cytogenetic perspective.pp. 137–170.

Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C., Palloix, A. and Jacquemond, M. (2004). Mutations in *Potato virus Y* genome-linked protein determine virulence toward recessive resistances in *Capsicum annuum* and Lycopersicon hirsutum. *Mol. Plant. Microbe Interact.* **17**, 322–329.

Moury, B., Palloix, A., Caranta, C., Gognalons, P., Souche, S., Selassie, K. G. and Marchoux, G. (2005). Serological, molecular, and pathotype diversity of *Pepper veinal mottle virus* and *Chili veinal mottle virus*. *Phytopathology* **95**, 227–232.

Moury, B., Fabre, F. and Senoussi, R. (2007). Estimation of the number of virus particles transmitted by an insect vector. *Proc. Natl. Acad. Sci. USA.* **104**, 17891–17896.

Moury, B., Fereres, A., García-Arenal, F. and Lecoq, H. (2011). Sustainable management of plant resistance to viruses. In *Recent advances in plant virology*, pp. 219–336. Norwich, UK.

Moury, B., Janzac, B., Ruellan, Y., Simon, V., Ben Khalifa, M., Fakhfakh, H., Fabre, F. and Palloix, A. (2014). Interaction Patterns between *Potato Virus Y* and eIF4E-Mediated Recessive Resistance in the Solanaceae. *J. Virol.* **88**, 9799–9807.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutat. Res. Mol. Mech. Mutagen.* **1**, 2–9.

Mundt, C. (2002). Use of multiline cultivars and cultivar mixtures for disease management. *Annu. Rev. Phytopathol.* **40**, 381–410.

Murphy, J. F., Blauth, J. R., Livingstone, K. D., Lackney, V. K. and Jahn, M. K. (1998). Genetic mapping of the pvr1 locus in *Capsicum* spp. and evidence that distinct potyvirus resistance loci control responses that differ at the whole plant and cellular levels. *Mol. Plant. Microbe Interact.* **11**, 943–951.

N

Nei, M. and Tajima, F. (1981). Genetic drift and estimation of effective population size. *Genetics* **98**, 625–640.

Nicolaï, M., Pisani, C., Bouchet, J.-P., Vuylsteke, M. and Palloix, A. (2012). Short Communication Discovery of a large set of SNP and SSR genetic markers by high-throughput sequencing of pepper (*Capsicum annuum*). *Genet. Mol. Res.* **11**, 2295–2300.

Nicolaï, M., Cantet, M., Lefebvre, V., Sage-Palloix, A.-M. and Palloix, A. (2013). Genotyping a large collection of pepper (*Capsicum* spp.) with SSR loci brings new evidence for the wild origin of cultivated *C. annuum* and the structuring of genetic diversity by human selection of cultivar types. *Genet. Resour. Crop Evol.* **60**, 2375–2390.

Nordborg, M. and Krone, S. M. (2002). Separation of time scales and convergence to the coalescent in structured populations. In *Modern Developments in Population Genetics. The Legacy of Gustave Malécot,* 194–232. Oxford, UK: Oxford Univ. Press.

Novella, I. S., Reissig, D. D. and Wilke, C. O. (2004). Density-dependent selection in *Vesicular stomatitis virus*. *J. Virol.* **78**, 5799–5804.

P

Padgett, H. S., Watanabe, Y. and Beachy, R. N. (1997). Identification of the TMV replicase sequence that activates the N gene-mediated hypersensitive response. *Mol. Plant. Microbe Interact.* **10**, 709–715.

Pagny, G., Paulstephenraj, P. S., Poque, S., Sicard, O., Cosson, P., Eyquard, J., Caballero, M., Chague, A., Gourdon, G. and Negrel, L. (2012). Family-based linkage and association mapping reveals novel genes affecting *Plum pox virus* infection in *Arabidopsis thaliana*. *New Phytol.* **196**, 873–886.

Palloix, A., Ayme, V. and Moury, B. (2009). Durability of plant major resistance genes to pathogens depends on the genetic background, experimental evidence and consequences for breeding strategies. *New Phytol.* **183**, 190–199.

Parlevliet, J. E. (2002). Durability of resistance against fungal, bacterial and viral pathogens; present situation. *Euphytica* **124**, 147–156.

Pascual, L., Desplat, N., Huang, B. E., Desgroux, A., Bruguier, L., Bouchet, J., Le, Q. H., Chauchard, B., Verschave, P. and Causse, M. (2015). Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnol. J.* **13**, 565–577.

Patwa, Z. and Wahl, L. M. (2008). The fixation probability of beneficial mutations. *J. R. Soc. Interface* **5**, 1279–1289.

Peakall, R. and Smouse, P. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28, 2537e2539.

Perrier, X. and Jacquemoud-Collet, J. (2006). DARwin (http://darwin. cirad. fr/darwin).

Perry, L., Dickau, R., Zarrillo, S., Holst, I., Pearsall, D. M., Piperno, D. R., Berman, M. J., Cooke, R. G., Rademaker, K. and Ranere, A. J. (2007). Starch fossils and the domestication and dispersal of chili peppers (*Capsicum* spp. L.) in the Americas. *Science* **315**, 986–988.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One* **7**, e37135.

Pink, D. A. (2002). Strategies using genes for non-durable disease resistance. *Euphytica* **124**, 227–236.

Poland, J. A., Balint-Kurti, P. J., Wisser, R. J., Pratt, R. C. and Nelson, R. J. (2009). Shades of gray: the world of quantitative disease resistance. *Trends Plant Sci.* **14**, 21–29.

Potato Genome Sequencing Consortium (2011). Genome sequence and analysis of the tuber crop potato. *Nature* **475**, 189–195.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

Q

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., Luo, Y., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl. Acad. Sci. USA.* **111**, 5135–5140.

Quenouille, J., Montarry, J., Palloix, A. and Moury, B. (2013a). Farther, slower, stronger: how the plant genetic background protects a major resistance gene from breakdown: Mechanisms of polygenic resistance durability. *Mol. Plant Pathol.* **14**, 109–118.

Quenouille, J., Vassilakos, N. and Moury, B. (2013b). *Potato virus Y*: a major crop pathogen that has provided major insights into the evolution of viral pathogenicity. *Mol. Plant Pathol.* **14**, 439–452.

Quenouille, J., Paulhiac, E., Moury, B. and Palloix, A. (2014). Quantitative trait loci from the host genetic background modulate the durability of a resistance gene: a rational basis for sustainable resistance breeding in plants. *Heredity* **112**, 579–587.

Quenouille, J., Saint-Felix, L., Moury, B. and Palloix, A. (2015). Diversity of genetic backgrounds modulating the durability of a major resistance gene. Analysis of a core collection of pepper landraces resistant to *Potato virus Y*. *Mol. Plant Pathol.* **17**, 296–302.

Quinn, G. P. and Keough, M. J. (2002). Experimental design and data analysis for biologists. Cambridge University Press.

R

Råberg, L., Sim, D. and Read, A. F. (2007). Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. *Science* **318**, 812–814.

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. M. and Buckler, E. S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA.* **98**, 11479–11484.

Restif, O. and Koella, J. C. (2004). Concurrent evolution of resistance and tolerance to pathogens. *Am. Nat.* **164**, E90–E102.

Richardson, K., Vales, M., Kling, J., Mundt, C. and Hayes, P. (2006). Pyramiding and dissecting disease resistance QTL to barley stripe rust. *Theor. Appl. Genet.* **113**, 485–495.

Rieux, A., De Bellaire, L. D. L., Zapater, M.-F., Ravigné, V. and Carlier, J. (2013). Recent range expansion and agricultural landscape heterogeneity have only minimal effect on the spatial genetic structure of the plant pathogenic fungus *Mycosphaerella fijiensis*. *Heredity* **110**, 29–38.

Rodrigo, G., Zwart, M. P. and Elena, S. F. (2014). Onset of virus systemic infection in plants is determined by speed of cell-to-cell movement and number of primary infection foci. *J. R. Soc. Interface* **11**, 20140555.

Rojas, M. R. and Gilbertson, R. L. (2008). Emerging plant viruses: a diversity of mechanisms and opportunities. In *Plant virus evolution*, pp. 27–51. Springer.

Rolland, M., Lacroix, C., Blanchard, A., Baldwin, T., Kerlan, C. and Jacquot, E. (2008). *Potato virus Y* (PVY): From its discovery to the latest outbreaks. *Virologie* **12**, 261–273.

Rousseau, E., Fabre, F., Senoussi, R., Mailleret, L., Palloix, A., Simon, V., Valière, S., Moury, B. and Grognard, F. (In press). Estimating virus effective population size and selection without neutral markers. *PLoS Pathog.*

Rouzine, I. M., Rodrigo, A. and Coffin, J. M. (2001). Transition between Stochastic Evolution and Deterministic Evolution in the Presence of Selection: General Theory and Application to Virology. *Microbiol. Mol. Biol. Rev.* **65**, 151–185.

Rubio, M., Nicolai, M., Caranta, C. and Palloix, A. (2009). Allele mining in the pepper gene pool provided new complementation effects between *pvr2*-eIF4E and *pvr6*-eIF(iso)4E alleles for resistance to *Pepper veinal mottle virus*. *J. Gen. Virol.* **90**, 2808–2814.

Ruffel, S., Dussault, M., Palloix, A., Moury, B., Bendahmane, A., Robaglia, C. and Caranta, C. (2002). A natural recessive resistance gene against *Potato virus Y* in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E). *Plant J.* **32**, 1067–1075.

Ruffel, S., Dussault, M., Lesage, M., Moretti, A., Palloix, A., Daunay, M., Moury, B., Bendahmane, A., Robaglia, C. and Caranta, C. (2004). Involvement of the eukaryotic translation initiation factor eIF4E in Solanaceae-Potyvirus interactions.pp. 167–170. European Association for Research on Plant Breeding (EUCARPIA).

Ruffel, S., Gallois, J.-L., Moury, B., Robaglia, C., Palloix, A. and Caranta, C. (2006). Simultaneous mutations in translation initiation factors eIF4E and eIF (iso) 4E are required to prevent *Pepper veinal mottle virus* infection of pepper. *J. Gen. Virol.* **87**, 2089–2098.

S

Sacristán, S., Malpica, J. M., Fraile, A. and Garcia-Arenal, F. (2003). Estimation of Population Bottlenecks during Systemic Movement of *Tobacco mosaic virus* in Tobacco Plants. *J. Virol.* **77**, 9906–9911.

Sacristán, S., Diaz, M., Fraile, A. and Garcia-Arenal, F. (2011). Contact Transmission of *Tobacco mosaic virus*: a Quantitative Analysis of Parameters Relevant for Virus Evolution. *J. Virol.* **85**, 4974–4981.

Sage-Palloix, A.-M., Jourdan, F., Phaly, T., Nemouchi, G., Lefebvre, V. and Palloix, A. (2007). Analysis of diversity in pepper genetic resources: distribution of horticultural and resistance traits in the INRA pepper germplasm. In *Progress in research on Capsicum & Eggplant*, pp. 32–42. Warsaw (Pologne).

Sanjuan, R., Nebot, M. R., Chirico, N., Mansky, L. M. and Belshaw, R. (2010). Viral mutation rates. *J. Virol.* **84**, 9733–9748.

Sierro, N., Battey, J. N., Ouadi, S., Bovet, L., Goepfert, S., Bakaher, N., Peitsch, M. C. and Ivanov, N. V. (2013). Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol.* **14**, R60.

Silander, O. K., Tenaillon, O. and Chao, L. (2007). Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. *PLoS Biol.* **5**, e94.

Smith, D. R., Adams, A. P., Kenney, J. L., Wang, E. and Weaver, S. C. (2008). *Venezuelan equine encephalitis virus* in the mosquito vector *Aedes taeniorhynchus*: infection initiated by a small number of susceptible epithelial cells and a population bottleneck. *Virology* **372**, 176–186.

Sommerhalder, R. J., McDonald, B. A., Mascher, F. and Zhan, J. (2011). Effect of hosts on competition among clones and evidence of differential selection between pathogenic and saprophytic phases in experimental populations of the wheat pathogen *Phaeosphaeria nodorum*. *BMC Evol. Biol.* **11**, 188.

Sonah, H., O'Donoughue, L., Cober, E., Rajcan, I. and Belzile, F. (2015). Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol. J.* **13**, 211–221.

Stefansson, T. S., McDonald, B. A. and Willi, Y. (2014). The influence of genetic drift and selection on quantitative traits in a plant pathogenic fungus. *PLoS One* **9**, e112523.

Steinrücken, M., Bhaskar, A. and Song, Y. S. (2014). A novel spectral method for inferring general diploid selection from time series genetic data. *Ann. Appl. Stat.* **8**, 2203.

Stekhoven, D. J. and Bühlmann, P. (2011). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112–118.

Syller, J. (2016). Antagonistic within-host interactions between plant viruses: molecular basis and impact on viral and host fitness. *Mol. Plant Pathol.* **17**, 769–782.

T

Takeshita, M., Shigemune, N., Kikuhara, K., Furuya, N. and Takanami, Y. (2004). Spatial analysis for exclusive interactions between subgroups I and II of *Cucumber mosaic virus* in cowpea. *Virology* **328**, 45–51.

Tamisier, L., Rousseau, E., Barraillé, S., Nemouchi, G., Szadkowski, M., Mailleret, L., Grognard, F., Fabre, F., Moury, B. and Palloix, A. (2017). Quantitative trait loci in pepper control the effective population size of two RNA viruses at inoculation. *J. Gen. Virol.* **98**, 1923–1931.

Taranto, F., D'Agostino, N., Greco, B., Cardi, T. and Tripodi, P. (2016). Genome-wide SNP discovery and population structure analysis in pepper (*Capsicum annuum*) using genotyping by sequencing. *BMC Genomics* **17**, 943.

Tenaillon, O. (2014). The utility of Fisher's geometric model in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* **45**, 179–201.

Tenaillon, M. I., Sawkins, M. C., Long, A. D., Gaut, R. L., Doebley, J. F. and Gaut, B. S. (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. mays L.). *Proc. Natl. Acad. Sci. USA.* **98**, 9161–9166.

Terhorst, J., Schlötterer, C. and Song, Y. S. (2015). Multi-locus analysis of genomic time series data from experimental evolution. *PLoS Genet.* **11**, e1005069.

Thrall, P. H., Oakeshott, J. G., Fitt, G., Southerton, S., Burdon, J. J., Sheppard, A., Russell, R. J., Zalucki, M., Heino, M. and Ford Denison, R. (2011). Evolution in agriculture: the application of evolutionary approaches to the management of biotic interactions in agro-ecosystems. *Evol. Appl.* **4**, 200–215.

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635.

Tromas, N. and Elena, S. F. (2010). The rate and spectrum of spontaneous mutations in a plant RNA virus. *Genetics* **185**, 983–989.

Tromas, N., Zwart, M. P., Lafforgue, G. and Elena, S. F. (2014). Within-Host Spatiotemporal Dynamics of Plant Virus Infection at the Cellular Level. *PLoS Genet.* **10**, e1004186.

U

Urcuqui-Inchima, S., Haenni, A.-L. and Bernardi, F. (2001). Potyvirus proteins: a wealth of functions. *Virus Res.* **74**, 157–175.

V

Van Den Berg, F., Bacaër, N., Metz, J., Lannou, C. and Van Den Bosch, F. (2011). Periodic host absence can select for higher or lower parasite transmission rates. *Evol. Ecol.* **25**, 121–137.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423.

Vassilakos, N., Simon, V., Tzima, A., Johansen, E. and Moury, B. (2016). Genetic Determinism and Evolutionary Reconstruction of a Host Jump in a Plant Virus. *Mol. Biol. Evol.* **33**, 541–553.

Venkatesh, J., An, J., Kang, W.-H., Jahn, M. and Kang, B. C. (In press). Fine Mapping of the Dominant Potyvirus Resistance Gene *Pvr7* Reveals a Relationship with *Pvr4* in *Capsicum annuum*. *Phytopathology*.

Vogwill, T., Phillips, R. L., Gifford, D. R. and MacLean, R. C. (2016). Divergent evolution peaks under intermediate population bottlenecks during bacterial experimental evolution. *Proceedings of the Royal Society of London B: Biological Sciences*, **283**, 20160749.

Voorrips, R. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **93**, 77–78.

W

Wahl, L. M. and Gerrish, P. J. (2001). The probability that beneficial mutations are lost in populations with periodic bottlenecks. *Evolution* **55**, 2606–2610.

Wahl, L. M., Gerrish, P. J. and Saika-Voivod, I. (2002). Evaluating the impact of population bottlenecks in experimental evolution. *Genetics* **162**, 961–971.

Wang, J. (2001). A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genet. Res.* **78**, 243–257.

Wang, J. (2005). Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. **360**, 1395–1409.

Waples, R. S. (1989). A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**, 379–391.

Webster, B., Ott, M. and Greene, W. C. (2013). Evasion of superinfection exclusion and elimination of primary viral RNA by an adapted strain of *hepatitis C virus*. *J. Virol.* **87**, 13354–13369.

Williamson, E. G. and Slatkin, M. (1999). Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics* **152**, 755–761.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.

Y

Yeh, S. and Gonsalves, D. (1984). Evaluation of induced mutants of *Papaya ringspot virus* for control by cross protection. *Phytopathology* **74**, 1086–1091.

Yuste, E., Sánchez-Palomino, S., Casado, C., Domingo, E. and López-Galíndez, C. (1999). Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. *J. Virol.* **73**, 2745–2751.

Z

Zeyl, C. and DeVisser, J. A. G. (2001). Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. *Genetics* **157**, 53–61.

Zhan, J., Mundt, C. C. and McDonald, B. A. (2001). Using restriction fragment length polymorphisms to assess temporal variation and estimate the number of ascospores that initiate epidemics in field populations of *Mycosphaerella graminicola*. *Phytopathology* **91**, 1011–1017.

Zhan, J., Mundt, C., Hoffer, M. and McDonald, B. (2002). Local adaptation and effect of host genotype on the rate of pathogen evolution: an experimental test in a plant pathosystem. *J. Evol. Biol.* **15**, 634–647.

Zhan, J., Thrall, P. H. and Burdon, J. J. (2014). Achieving sustainable plant disease management through evolutionary principles. *Trends Plant Sci.* **19**, 570–575.

Zhan, J., Thrall, P. H., Papaïx, J., Xie, L. and Burdon, J. J. (2015). Playing on a Pathogen's Weakness: Using Evolution to Guide Sustainable Plant Disease Control Strategies. *Annu. Rev. Phytopathol.* **53**, 19–43.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K. and Ordovas, J. M. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360.

Zhu, Y., Chen, H., Fan, J. and Wang, Y. (2000). Genetic diversity and disease control in rice. *Nature* **406**, 718.

Zwart, M. P. and Elena, S. F. (2015). Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annu. Rev. Virol.* **2**, 161–179.

Zwart, M. P., Daròs, J.-A. and Elena, S. F. (2011). One Is Enough: In Vivo Effective Population Size Is Dose-Dependent for a Plant RNA Virus. *PLoS Pathog.* **7**, e1002122.

Zwart, M. P., Daros, J.-A. and Elena, S. F. (2012). Effects of Potyvirus Effective Population Size in Inoculated Leaves on Viral Accumulation and the Onset of Symptoms. *J. Virol.* **86**, 9737–9747.

Annexe

Le document en annexe de ce manuscrit est l'article « Estimating virus effective population size and selection without neutral markers » de Rousseau et al. (2017) qui vient d'être accepté pour publication dans PLOS Pathogens. Cet article présente le modèle utilisé dans cette thèse pour caractériser les lignées haploïdes doublées de piment en fonction des effets de dérive génétique et de sélection qu'elles imposent aux populations virales (chapitres 3 et 4). Etant donné que cet article est souvent cité dans ce manuscrit mais qu'il n'est pas encore disponible en ligne, il a été ajouté en annexe.

Estimating virus effective population size and selection without neutral markers

Short-title: Estimation of genetic drift and selection

Elsa Rousseau[1¶*,2,3], Benoît Moury[3], Ludovic Mailleret[1,2], Rachid Senoussi[4], Alain Palloix[5], Vincent Simon[3,6], Sophie Valière[7,8], Frédéric Grognard[1], Frédéric Fabre[9¶*]

1 Université Côte d'Azur, Inria, INRA, CNRS, UPMC Univ Paris 06, Biocore team, Sophia Antipolis, France

2 Université Côte d'Azur, INRA, CNRS, ISA, France

3 Pathologie Végétale, INRA, 84140 Montfavet, France

4 UR BioSp, INRA, Avignon, France

5 UR GAFL, INRA, Montfavet, France

6 UMR BFP, INRA, Villenave d'Ornon, France

7 GeT-PlaGe, INRA, Genotoul, Castanet-tolosan, France

8 UAR DEPT GA, INRA, Castanet-Tolosan, France

9 UMR SAVE, INRA, Villenave d'Ornon, France

¶These authors contributed equally to this work.

*** Corresponding authors**

elsa.rousseau7@gmail.com (ER)

frederic.fabre@inra.fr (FF)

Abstract

By combining high-throughput sequencing (HTS) with experimental evolution, we can observe the within-host dynamics of pathogen variants of biomedical or ecological interest. We studied the evolutionary dynamics of five variants of *Potato virus Y* (PVY) in 15 doubled-haploid lines of pepper. All plants were inoculated with the same mixture of virus variants and, variant frequencies were determined by HTS in eight plants of each pepper line, at each of six sampling dates. We developed a method for estimating the intensities of selection and genetic drift in a multi-allelic Wright-Fisher model, applicable whether these forces are strong or weak, and in the absence of neutral markers. This method required variant frequency determination at several time points, in independent hosts. The parameters are the selection coefficients for each PVY variant and four effective population sizes N_e at different time-points of the experiment. Numerical simulations of asexual haploid Wright-Fisher populations subjected to contrasting genetic drift ($N_e \in [10, 2000]$) and selection ($|s| \in [0, 0.15]$) regimes were used to validate the method proposed. This experiment in closely related host genotypes revealed that viruses experienced a considerable diversity of selection and genetic drift regimes. The resulting variant dynamics were accurately described by Wright-Fisher models. The fitness ranks of the variants were almost identical between host genotypes. By contrast, the dynamics of N_e were highly variable, although a bottleneck was often identified during the systemic movement of the virus. We demonstrated that, for a fixed initial PVY population, effective PVY population size is a heritable trait in plants. These findings pave the way for the breeding of plant varieties exposing viruses to stronger genetic drift, thereby slowing virus adaptation.

Key words

Genetic drift, Experimental evolution, High-throughput sequencing, Parameter estimation, MiSeq Illumina, Plant virus, Selection, Virus evolution, Wright-Fisher model

Author Summary

A growing number of experimental evolution studies are using an "evolve-and-resequence" approach to observe the within-host dynamics of pathogen variants of biomedical or ecological interest. The resulting data are particularly appropriate for studying the effects of evolutionary forces, such as selection and genetic drift, on the emergence of new pathogen variants. However, it remains challenging to unravel the effects of selection and genetic drift in the absence of neutral markers, a situation frequently encountered for microbes, such as viruses, due to their small constrained genomes. Using such an approach on a plant virus, we observed that the same set of virus variants displayed highly diverse dynamics in closely related plant genotypes. We developed and validated a method that does not require neutral markers, for estimating selection coefficients and effective population sizes from these experimental evolution data. We found that the viruses experienced considerable diversity in genetic drift regimes, depending on host genotype. Importantly, genetic drift experienced by virus populations was shown to be a heritable plant trait. These findings pave the way for the breeding of plant varieties exposing viruses to strong genetic drift, thereby slowing virus adaptation.

Introduction

Evolution in isolated populations results from the interplay between several forces, including mutation, selection, and genetic drift. Mutation creates genetic diversity within a population. Subsequent selection and genetic drift drive the evolution of diversity within the population. Selection is a deterministic force that increases the frequency of the fittest variants at the expense of the weakest ones. It can be characterized by the selection coefficient s, commonly calculated, at a specific locus, as the relative difference in fitness conferred by two alleles. Genetic drift, unlike selection,

acts equally on all variants. It is the outcome of random sampling effects between generations, resulting in stochastic fluctuations in variant frequencies [1]. The strength of genetic drift is frequently evaluated by determining the effective population size N_e [1]. N_e is defined as the size of an ideal panmictic population of constant size with non-overlapping generations that would display the same degree of randomness in allele frequencies as the population studied [2]. N_e is often much lower than the census population size [3, 4], but it can be seen as its evolutionary analog [5]. When N_e is small, sampling effects are magnified between generations, and allele frequencies therefore fluctuate strongly. For populations varying in size over time, the effective population size over a given number of generations can be approximated by the harmonic mean $\bar{N}_e$ of effective population sizes at each generation. This approximation holds provided that the number of generations is much smaller than $\bar{N}_e$ [6–8] and that mutation can be neglected [9]. Population size may vary over time due to bottlenecks, which are common in natural populations. As they greatly decrease population size, they have a disproportionate effect on the overall value of $\bar{N}_e$ [1].

When selection and genetic drift act simultaneously, the probability of fixation of a new mutation (with a selection coefficient s), and, more generally, its evolutionary dynamics, is controlled by the product $N_e \times s$ [1, 10]. If $N_e \times s \ll 1$, then genetic drift predominates over selection and evolution is mostly stochastic. If $N_e \times s \gg 1$, then selection becomes effective and evolution is mostly deterministic [10]. This rule of thumb can be applied to the evolutionary dynamics of pathogen variants of biomedical or ecological interest, during the course of infection of a single host, for microbe variants escaping the immune response of their host, or becoming resistant to drug therapy (e.g. [11]) or, in the case of plant pathogens, for variants adapting to host resistance genes (e.g. [12]). In this study, we combined high-throughput sequencing (HTS) with experimental evolution to measure the within-host dynamics of adaptation of five

variants of *Potato virus Y* (PVY, genus *Potyvirus*, family Potyviridae) in closely related plant genotypes [13].

It remains challenging to unravel the effects of genetic drift and selection in the absence of neutral markers, in studies of the adaptation dynamics of pathogens. This situation is frequently encountered for pathogens with small genomes, especially viruses [14, 15]. Various approaches based on moment [16, 17] or likelihood [18–20] methods have been proposed for estimating N_e, but all require the genetic markers studied to be neutral. Various methods have also been proposed for detecting selection and estimating selection coefficients. These methods require at least some prior information about N_e (e.g. [21]) or assume that genetic drift is negligible (e.g. [22]). However, in the absence of neutral markers and without prior estimates of N_e, both selection and genetic drift must be taken into account, as these two forces act simultaneously on evolution. This greatly complicates the estimation of N_e and s. Only a few methods have been proposed for the joint estimation of N_e and s from time-sampled data (see [11] and [23] for a review). For large effective population sizes (typically $N_e > 5000$) and small selection coefficients (typically $|s| < 0.01$), several likelihood methods based on diffusion approximations of the Wright-Fisher model [1, 11] are available [24–27]. In the situations in which these methods are valid, the ranges of N_e and s values obtained are rather restrictive for many microorganisms, particularly viruses [28–31]. Foll *et al.* [32] recently proposed the use of approximate Bayesian computation (ABC) for the joint estimation of N_e and s in a Wright-Fisher model. Their method can deal with both weak and strong selection regimes, but still requires multilocus genome-wide data with mostly neutral loci to estimate N_e accurately.

In this study, we investigated the evolutionary dynamics of five variants of PVY in 15 closely related pepper genotypes. All plants were inoculated with the same mixture of virus variants and variant frequencies were determined with HTS in eight plants of each

genotype at each of six sampling dates after inoculation. A diverse range of evolutionary patterns was observed. We developed a method for estimating the parameters of a multi-allelic Wright-Fisher model with selection and genetic drift, to investigate the underlying evolutionary processes. This method has two main advantages: it applies to a large range of selection and genetic drift intensities, whether strong or weak, and it works efficiently in the absence of neutral markers. The parameters of the Wright-Fisher model (i.e. selection coefficients for each virus variant and effective population sizes at given time points) can be estimated by coupling maximum likelihood and ABC methods and applying them to a set of variant frequencies determined at several time points in independent hosts. We tested the method with numerical simulations mimicking the datasets obtained with HTS in evolve-and-resequence experiments [33]. The simulations covered an extensive range of N_e and s values. We were then able to estimate the selection coefficient of each PVY variant in each pepper genotype and the changes in effective population size over time during the colonization of the plant by the virus. Finally, by varying pepper genotypes and fixing the initial PVY population, we provided evidence that the effective population size of PVY is a heritable plant trait. This finding paves the way for the breeding of plant cultivars exposing viruses to greater genetic drift and/or smaller selection effects.

Materials and Methods

Biological experiment

Plant and virus material We used 15 doubled-haploid (DH) lines of pepper (*Capsicum annuum*, family Solanaceae). All the plants of a given genotype were thus genetically identical. All DH lines carried the major resistance gene *pvr2*^{3} and differed in terms of their genetic background [12]. They are issued from the F_1 hybrid between

two pepper lines, Perennial and Yolo Wonder. Thus, on average, any pair of DH lines had 50% of alleles in common, at markers distinguishing between Perennial and Yolo Wonder. Each DH line therefore constituted a different host environment for plant colonization by PVY. These lines were chosen for study on the basis of quantitative differences in three previously measured factors, so as to generate different intensities of genetic drift and selection acting on PVY populations: (i) relative within-plant viral accumulation, (ii) resistance breakdown (RB) frequency [12] and (iii) the number of primary infection foci after mechanical inoculation with the virus (Fig S1).

All plants were mechanically inoculated with the same initial virus population comprising an equimolar mixture, based on quantitative double-antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA), of the five PVY variants G, N, K, GK and KN [34]. Single- and double-letter names indicate single and double mutants, respectively, of the infectious clone SON41p (Fig. 1A). Three mutations located close together in the PVY genome differentiate the five variants, and these mutations are named after the amino-acid substitutions observed at positions 101 for the S (serine) to G (glycine) substitution, 115 for the T (threonine) to K (lysine) substitution, and 119 for the D (aspartic acid) to N (asparagine) substitution, in the VPg (viral protein genome-linked) protein. The G and N variants displayed a low level of adaptation to the major resistance gene *pvr2*^{3} carried by all plant genotypes, whereas variants K, GK and KN displayed higher levels of adaptation [12].

Experimental set-up and plant sampling For each pepper genotype, 48 plants were arranged in randomized blocks, to minimize environmental effects. The first true leaf of each plant was inoculated 29 days after sowing. We then analyzed eight plants per DH line at 6, 10, 14, 20, 27 and 34 days post-inoculation (dpi) (Fig. 1B). The inoculated leaf was sampled at 6 dpi, and, on subsequent sampling dates, three uninoculated leaves, corresponding to the three youngest unfolded leaves, were sampled

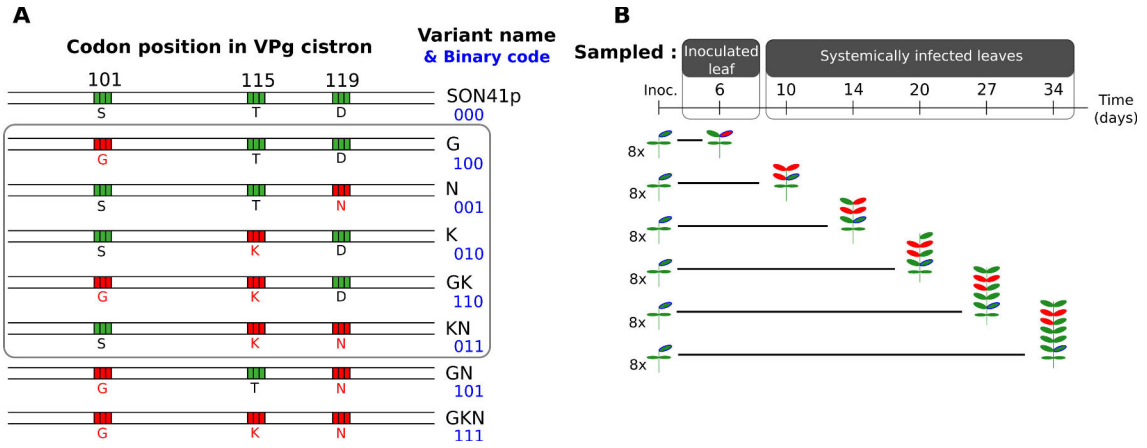


Fig 1. Virus variants inoculated to pepper plants and sampling protocol. (A) The five virus variants (in the gray box) were derived from the SON41p PVY clone and differed only at codon positions 101, 115 and 119 of the VPg cistron. These positions are shown in green if they correspond to the SON41p clone and in red if a non-synonymous substitution was introduced by site-directed mutagenesis. Single-letter amino acid abbreviations are presented below each position and PVY variant. Variant names and the corresponding binary code for the three point mutations of interest are given on the right of the sequences, with the binary code of the SON41p variant set to 000. The two additional possible variants, based on the three-digit binary code, are also shown at the bottom. (B) Sampling protocol for one pepper genotype. We inoculated 48 plants with the virus. Eight plants were sampled at each sampling time, from 6 to 34 days post-inoculation. The leaf circled in blue is the leaf inoculated with the virus. The leaves sampled are shown in plain red.

and pooled together. As the plants were removed after sampling, the virus populations obtained from each plant sample were independent. This prevented possible effects on virus population dynamics due to the removal of infected leaves and subsequent re-sampling. Each leaf sample was ground in four volumes of 0.03 M phosphate buffer (pH 7.0) supplemented with 2% (w/v) diethyldithiocarbamate, as previously described [34].

High-Throughput Sequencing and determination of PVY variant frequencies Total RNA was purified from individual plant samples with the Tri-Reagent kit (Sigma-Aldrich). It was subjected to reverse transcription-polymerase chain reaction (RT-PCR) with tagged primers for the amplification, over 35 cycles, of a 104-nucleotide region encompassing the polymorphic region of the PVY VPg cistron. Eight differently tagged primers were used, corresponding to the eight different plant

replicates of the same plant genotype for each sampling date (Table S1). Amplified DNAs corresponding to the eight plant replicates were pooled together on the basis of their intensity on electrophoresis gels.

HTS was performed at the Genomic Platform of INRA Toulouse. For this purpose, 2×150 base-pair (bp) libraries with multiplex adapters were prepared, and all the RT-PCR-amplified products were pooled into a single large sample (12 cycles). This sample was run on a MiSeq Illumina paired-end sequencer with the MiSeq Reagent Kit v2, for 500 cycles. We chose to use MiSeq Illumina sequencing, because this technology has a much lower error rate than other high-throughput sequencing technologies, such as 454 sequencing [35]. By using tagged primers and subsequent multiplex adapters, we were able to assign a plant genotype and a sampling date to each sequence.

In the initial sequence analysis, we used FLASH software to obtain the consensus sequence from reads 1 and 2 with a minimum overlap length between the two reads of 63, a maximum overlap length of 153 and a maximum allowed ratio of the number of mismatched base pairs to overlap length of 0.2 [36]. The sequences were then sorted by adapter and by tag. Finally, the sequences corresponding to each PVY variant were determined with the 'agrep' function in R software [37], and sequence counts were used to assess the composition of the virus population in each sample. After sequence sorting, we had 374 to 14141 sequences per sample, with a mean of 3295 sequences per sample.

We carried out two complementary sequence analyses to detect PVY mutations (see Text S1 for details). It was important to perform these analyses, as the presence of mutants might have affected virus population dynamics and the intensities of the evolutionary forces studied. In the first analysis, we looked for all eight possible variants based on the three codon positions of interest in the VPg cistron, *i.e.* the five analyzed variants, G, N, K, GK and KN, together with the SON41p, GN and GKN variants (corresponding to all the possible binary codes in Fig. 1A). The sum of the frequencies

of all three additional variants, SON41p, GN and GKN, remained below 5%, and these variants were not, therefore, considered in N_e and s estimations. The raw data from this analysis (*i.e.* number of sequences of each variant in each of the 677 samples analyzed) are available from Table S2. We then calculated the frequencies of *de novo* substitutions in each sample and at each nucleotide position of all sequences, by comparison with the sequence of the SON41p reference clone (equivalent to comparison with sequences of the G, N, K, GK and KN clones). In all, only PVY populations sampled from six of the 677 plants studied presented a *de novo* substitution with a frequency exceeding 5% (Text S1). These PVY populations were removed for subsequent analyses. Furthermore, numerical simulations showed that a sixth unaccounted for variant present at a mean frequency of 7% in virus populations had no significant impact on estimates of N_e and s (see below), justifying our use of a 5% threshold.

Genetic analysis and heritability estimation With this experimental design, we studied the phenotype of each pepper DH line in terms of its effect on PVY populations. The plants within each pepper DH line were genetically identical and experimental conditions were set up so as to ensure an absence of differences in environmental effects between DH lines. We were therefore able to estimate the heritability of plant traits of interest, corresponding to the evolutionary forces exerted by the plant on PVY populations, by constituting two replicates for each DH line dataset. More precisely, we assessed the heritability of the intrinsic rates of increase for the five PVY variants and of the effective population sizes of PVY. As the initial population of PVY was fixed and identical for all plant genotypes, we did not consider the effect of PVY population composition on the heritability of effective PVY population size. We split the dataset for the 48 plants for each DH line in two, by randomly selecting four of the eight plants at each sampling date. The broad-sense heritability of any plant trait of interest can be estimated as $h^2 = \sigma_G^2/(\sigma_G^2 + \sigma_e^2/n)$, where σ_G^2 corresponds to the genotypic variance, σ_e^2

to the phenotypic variance and n to the number of replicates [38], the variance being set to the sum of the squared deviations from the mean.

Estimation of selection and genetic drift intensities

We developed a method for estimating the parameters of a multi-allelic Wright-Fisher model with selection and genetic drift for a haploid population. The parameters and state variables of the model and the observed variables are summarized in Table 1.

Table 1. Main notations for the observations and the model.

	Designation (unit) [reference value]
Observed variables	
$\boldsymbol{x}^p(t_d) = (x_1^p(t_d), \ldots, x_{n_{var}}^p(t_d))$	Variant sequence counts in virus population p at sampling time t_d (seq [a])
$\boldsymbol{f}^p(t_d) = (f_1^p(t_d), \ldots, f_{n_{var}}^p(t_d))$	Variant frequencies in virus population p at sampling time t_d (no unit)
State variables	
$\boldsymbol{\lambda}^p(t) = (\lambda_1^p(t), \ldots, \lambda_{n_{var}}^p(t))$	Theoretical variant frequencies in virus population p at time t of a Wright-Fisher model (no unit)
Parameters of interest	
$\boldsymbol{r} = (r_1, \ldots, r_{n_{var}})$	Variant relative intrinsic rates of increase ($generation^{-1}$) [b]
$\boldsymbol{\eta}_e = \left(\eta_e^{IO}, \eta_e^{S_1}, \eta_e^{S_2}, \eta_e^{S_3}\right)$	Successive virus effective population sizes ($individuals$) [c]
Fixed parameters	
$\boldsymbol{\lambda}^{inoc}$	Vector of variant frequencies in the virus inoculum (no unit)
$\boldsymbol{T}$	Vector of measurement dates (day) $[(0, 6, 10, 14, 20, 27, 34)]$
Additional notations	
$\boldsymbol{N_e} = (N_e(1), \ldots, N_e(34))$	Vector of virus effective population sizes (piecewise constant function of $\boldsymbol{\eta}_e$)
$N_e^h(t_d)$	Harmonic mean of virus effective population sizes at sampling time t_d
$\sigma[f_i^\bullet(t_d)]$	Standard deviation of the frequencies of virus variant i at sampling time t_d over the virus populations p
$\lambda_i^{det}(t)$	Deterministic frequency of variant i at time t for an infinite size Wright-Fisher model

[a] The abbreviation "seq" is the number of sequences representing the virus population or a given variant in this population. [b] The mean intrinsic rate of increase $\bar{r}$ of all virus variants is one. [c] With the full model $\mathfrak{M}_4$, η_e^{IO} in the inoculated organ for $t \in [1, 6]$, η_e^{S1} at the onset of systemic infection for $t \in [7, 10]$, η_e^{S2} for $t \in [11, 14]$ and η_e^{S3} for $t \in [15, 34]$.

Notation: observed variables, state variables and parameters of interest

$\boldsymbol{\lambda}^{inoc}$ denotes the vector of the observed variant frequencies in the parental virus population, i.e. the inoculum used to inoculate all host plants in the experiment. Thereafter, the measurement date vector $\boldsymbol{T} = (0, 6, 10, 14, 20, 27, 34)$ is indexed by t_d $(d = 0, \ldots, 6)$. In particular, $t_0 = 0$ is the inoculation date and t_d $(d = 1, \ldots, 6)$ are the sampling dates. The time, indexed by $t = 1, \ldots, 34$, is the number of days

post-inoculation and indicates the viral generation, as the generation time was assumed to be one day [39]. For a given plant genotype, at each sampling date t_d a sample of $n_{inf}(t_d)$ infected plants was observed, for which we measured the vectors $\boldsymbol{x}^p(t_d) = (x_1^p(t_d), \ldots, x_{n_{var}}^p(t_d))$, with $x_i^p(t_d)$ the number of sequences obtained for virus variant i ($1 \leq i \leq n_{var}$) in population p ($1 \leq p \leq n_{inf}(t_d)$). Thereafter, replacing an index in a notation by $\bullet$ is equivalent to summing over the corresponding index set. The total number of sequences obtained from virus population p at time t_d is thus $x_\bullet^p(t_d)$. Finally, $\boldsymbol{f}^p(t_d) = (f_1^p(t_d), \ldots, f_{n_{var}}^p(t_d))$, with $f_i^p(t_d)$ the observed frequency of virus variant i in population p at sampling date t_d. It is calculated as $\boldsymbol{x}^p(t_d)/x_\bullet^p(t_d)$.

The state variable of interest is $\boldsymbol{\lambda}^p(t) = (\lambda_1^p(t), \ldots, \lambda_{n_{var}}^p(t))$, with $\lambda_i^p(t)$ the frequency of virus variant i ($1 \leq i \leq n_{var}$) in virus population p at any date $t = 1, \ldots, 34$. Virus variant dynamics are represented by a Wright-Fisher model (see below) to infer $\boldsymbol{\theta} = (\boldsymbol{r}, \boldsymbol{\eta}_e)$, the vector of parameters describing the underlying evolutionary forces. $\boldsymbol{r} = (r_1, \ldots, r_{n_{var}})$ is the vector of the intrinsic rates of increase r_i of each virus variant i. We assumed that the mean intrinsic rate of increase $\bar{\boldsymbol{r}}$ over all variants was one, as we were interested only in the relative intrinsic rates of increase. The selection coefficient of a variant i is usually computed as $s_i = r_i - 1$. The vector parameter $\boldsymbol{\eta}_e$ defines a piecewise function describing effective population sizes $N_e(t)$. We determined the temporal variation of effective population sizes, using four models with $\boldsymbol{\eta}_e$ with one to four parameters. With the more general model $\mathfrak{M}_4$, $\boldsymbol{\eta}_e = \left(\eta_e^{IO}, \eta_e^{S_1}, \eta_e^{S_2}, \eta_e^{S_3}\right)$. η_e^{IO} is the effective population size of the viral population in the inoculated organ; this stage lasts $t_1 = 6$ days in our experimental design. $\eta_e^{S_1}$ is the effective population size during the onset of systemic infection; this stage lasts $t_2 - t_1 = 4$ days. $\eta_e^{S_2}$ is the effective population size during the next $t_3 - t_2 = 4$ days and $\eta_e^{S_3}$ the effective population size later on, during the last $t_6 - t_3 = 20$ days of survey. Accordingly, we define $N_e(t)$ as follows:

$$N_e(t) = \begin{cases} \eta_e^{IO} & t \in [1, \ldots, t_1] \\ \eta_e^{S_1} & t \in [t_1 + 1, \ldots, t_2] \\ \eta_e^{S_2} & t \in [t_2 + 1, \ldots, t_3] \\ \eta_e^{S_3} & t \in [t_3 + 1, \ldots, t_6] \end{cases} \quad (1)$$

With model $\mathfrak{M}_3$, $\boldsymbol{\eta}_e = \left(\eta_e^{IO}, \eta_e^{S_1}, \eta_e^{S_2}\right)$. $N_e(t)$ has three parameters: (i) $N_e(t) = \eta_e^{IO}$ when $t \in [1, 6]$, (ii) $N_e(t) = \eta_e^{S_1}$ when $t \in [7, 14]$ and (iii) $N_e(t) = \eta_e^{S_2}$ when $t \in [15, 34]$. With model $\mathfrak{M}_2$, $\boldsymbol{\eta}_e = \left(\eta_e^{IO}, \eta_e^{S}\right)$. $N_e(t)$ has two parameters: (i) $N_e(t) = \eta_e^{IO}$ when $t \in [1, 6]$ and (ii) $N_e(t) = \eta_e^{S}$ when $t \in [7, 34]$. Finally, with model $\mathfrak{M}_1$, $\boldsymbol{\eta}_e = \left(\eta_e\right)$: the effective population size for the virus remains constant throughout the experiment ($N_e(t) = \eta_e$ for $t \in [1, 34]$). The effective population size at any sampling date of interest t_d is given, approximately, by the harmonic mean of the effective sizes of the successive generations $N_e^h(t_d) = \left(\frac{1}{t}\sum_{j=1}^{t}\frac{1}{N_e(j)}\right)^{-1}$.

The multi-allelic Wright-Fisher model with selection and genetic drift The Wright-Fisher model occupies a central position in population genetics ([40]). It assumes an ideal population: a randomly mating haploid population of finite size reproducing in discrete non-overlapping generations, with no structure. By definition, N_e is the size of an ideal population (i.e. obeying previous assumptions) that would display the same degree of randomness in variant frequencies as the real population studied ([2]). As for any model-based approach, the use of this concept requires the actual population not too far from an ideal Wright-Fisher model with suitable parameters ([10]). Formally, the Wright-Fisher model is very similar to the quasispecies model describing the evolution of DNA (or RNA) sequences in finite populations ([41]). In practice, the Wright-Fisher model has been used to infer the evolutionary history of viruses (*e.g.* [11]) and it can also be used to describe the stochastic dynamics of the frequency of the n_{var} virus variants considered here. Let $\boldsymbol{z}(t)$ be the vector of the

number of each virus variant i ($1 \leq i \leq n_{var}$) in generation t and, with previous notations, let $\boldsymbol{\lambda}(t) = \dfrac{\boldsymbol{z}(t)}{N_e(t)}$ be the corresponding vector of variant frequencies. The dynamics of $\boldsymbol{z}(t)$ are shaped by random genetic drift and selection. Let $\boldsymbol{pr}(t) = (pr_1(t), \ldots, pr_{n_{var}}(t))$ be the vector of the probabilities of sampling each virus variants from generation t to generation $t+1$. For $t \geq 1$, the distribution of $\boldsymbol{z}(t+1)$ conditionally on $\boldsymbol{z}(t)$ follows a multinomial distribution [40]:

$$\begin{cases} \boldsymbol{z}(t+1)|\boldsymbol{z}(t) \sim \text{Mult}\big(\text{size} = N_e(t), \text{prob} = \boldsymbol{pr}(t)\big) & (2) \\ \boldsymbol{pr}_i(t) = \dfrac{r_i \lambda_i(t)}{\sum_{j=1}^{n_{var}} r_j \lambda_j(t)} \quad \text{with} \quad \lambda_i(t) = \dfrac{z_i(t)}{N_e(t)} \quad \text{and} \quad \boldsymbol{\lambda}(0) = \boldsymbol{\lambda}^{inoc} & (3) \end{cases}$$

In the model, selection reweights the different genotypes according to their constant fitness. Fitness does not depend on the composition of the population (*i.e.* there is no frequency-dependent selection). As population size tends to ∞ (*i.e.* genetic drift becomes negligible), the stochastic process of variant frequencies described by eq. (2) converges on deterministic recursion describing $\lambda_i^{det}(t)$ which approximates the mean frequency of variant i at generation t. For $t \geq 1$:

$$\lambda_i^{det}(t+1)|\boldsymbol{r} = \frac{r_i \lambda_i^{det}(t)}{\sum_{j=1}^{n_{var}} r_j \lambda_j^{det}(t)} \quad \text{with} \quad \boldsymbol{\lambda}^{det}(0) = \boldsymbol{\lambda}^{inoc} \qquad (4)$$

Parameter estimation We propose an approach combining a first step relying on maximum-likelihood followed by a step relying on ABC, to estimate the parameters of interest $\boldsymbol{\theta} = (\boldsymbol{r}, \boldsymbol{\eta}_e)$. The first step estimates the vector of the relative intrinsic rates of increase of each virus variant $\boldsymbol{r}$ by maximum-likelihood methods. Let the vector $\boldsymbol{x}^{\bullet}(t_d) = (x_1^{\bullet}(t_d), \ldots, x_{n_{var}}^{\bullet}(t_d))$ be the total number of sequences of virus variant i obtained in the $n_{inf}(t_d)$ infected plants (of a given plant genotype) at sampling date t_d. This step assumes that $\boldsymbol{x}^{\bullet}(t_d) \sim Mult\left(\text{size} = \sum_{i=1}^{n_{var}} x_i^{\bullet}(t_d), \text{prob} = \boldsymbol{\lambda}^{det}(t_d)|\boldsymbol{r}\right)$. Let $\boldsymbol{x}$ denote the vector of all total sequence counts, at all sampling time-points, constituting

one dataset. As the samples are independent between sampling dates, the likelihood function is:

$$l(\boldsymbol{x}|\boldsymbol{r}) = \prod_{d=1}^{6} dM\left(\text{size} = \sum_{i=1}^{n_{var}} x_i^{\bullet}(t_d), \text{prob} = \boldsymbol{\lambda}^{det}(t_d)|\boldsymbol{r}\right),$$

dM being the probability density function (pdf) of the multinomial distribution. Under these hypotheses, $\boldsymbol{r}$ can be inferred by minimizing $-log\big(l(\boldsymbol{x}|\boldsymbol{r})\big)$, assuming that the mean intrinsic rate of increase $\bar{\boldsymbol{r}}$ of all variants is one. The estimate of $\boldsymbol{r}$, denoted $\hat{\boldsymbol{r}}$, was obtained straightforwardly, using the 'nlminb' optimization routine implemented in R software version 3.0.2 [37].

The second step estimates the vector of effective population sizes $\boldsymbol{\eta}_e$ of a given model $\mathfrak{M}_j$ $(j = 1, \cdots, 4)$ with ABC, conditionally to $\hat{\boldsymbol{r}}$. All ABC algorithms involve the simulation of a large number of possible datasets by sampling the parameters of interest (here $\boldsymbol{\eta}_e$) from prior probability distributions $\pi(\boldsymbol{\eta}_e)$. We used independent log-uniform priors on $[10, 2500]$ for each parameter of $\boldsymbol{\eta}_e$. For a given $\boldsymbol{\eta}_e^{sim}$ sampled in $\pi(\boldsymbol{\eta}_e)$, a dataset was simulated as follows. The first step was to simulate 48 (=6 sampling dates * 8 plants/date) independent dynamics of evolution of virus variant frequencies lasting 34 days $\big(\max(\boldsymbol{T})\big)$ with the Wright-Fisher model (equations 1, 2 and 3) parameterized by $(\hat{\boldsymbol{r}}, \boldsymbol{\eta}_e^{sim})$. Let $\boldsymbol{\lambda}_{sim}^{l}(t)$ $(l = 1, \ldots, 48\ ;\ t = 1, \ldots, 34)$ be this set of simulated dynamics. The second step was to simulate the experimental design. Eight plants were analyzed with HTS at each sampling date, $x_{sim}^{tot,l}$ being the total number of sequences obtained for plant l. Variant counts $\boldsymbol{x}_{sim}^{l}(t_d)$ for HTS are sampled from multinomial distributions. A single sample was obtained for each dynamic l : at 6 dpi, $\boldsymbol{x}_{sim}^{l}(6) \sim \text{Mult}\big(\text{size} = x_{sim}^{tot,l}, \text{prob} = \boldsymbol{\lambda}_{sim}^{l}(6)\big)$ with $l \in [1, 8]$; at 10 dpi, $\boldsymbol{x}_{sim}^{l}(10)$ is obtained similarly with $l \in [9, 16]$, and so on, until 34 dpi, with $l \in [41, 48]$. The observed frequencies are $\boldsymbol{f}_{sim}^{l}(t_d) = \dfrac{\boldsymbol{x}_{sim}^{l}(t_d)}{x_{sim}^{tot,l}}$. The last step was to calculate the vector of summary statistics from the simulated dataset, $\boldsymbol{S}_{sim} = \big(S_{sim}^{t_1}, \ldots, S_{sim}^{t_6}\big)$. A single

summary statistic was calculated for each sampling date. This statistic is the inverse of the mean of the standard deviation of the variant frequencies at sampling date t_d. Formally, $S_{sim}^{t_d} = \left(\frac{1}{n_{var}} \sum_{i=1}^{n_{var}} \sigma[f^{\bullet}_{sim,i}(t_d)] \right)^{-1}$ where $\sigma[f^{\bullet}_{sim,i}(t_d)]$ is the standard deviation (over the infected hosts at sampling date t_d) of $f^{l}_{sim,i}(t_d)$, the observed frequency of variant i. In practice, estimation was performed with the adaptive ABC algorithm of Lenormand *et al.* [42] implemented in the R package EasyABC with tuning parameters $nb_{simul} = 5000$, $p_{accmin} = 0.04$ and $\alpha = 0.5$. Models $\mathfrak{M}_1$, $\mathfrak{M}_2$, $\mathfrak{M}_3$ and $\mathfrak{M}_4$ (embedding $N_e(t)$ functions with one to four parameters) were compared, using the multinomial logistic regression method implemented in the ABC package (function postpr with 2.10^5, 2.5×10^5, 7.5×10^5 and 1.5×10^6 simulated summary statistics under models $\mathfrak{M}_1$, $\mathfrak{M}_2$, $\mathfrak{M}_3$ and $\mathfrak{M}_4$, respectively, and tuning parameter tol $= 5 \times 10^{-4}$). The estimation code will be made available upon request.

Numerical simulations

Before using the estimation method on the datasets corresponding to the biological experiment, we performed several batches of simulations to assess its ability to infer effective population sizes and selection coefficients accurately (see Text S2 for details). Briefly, in experiment 1, we first simulated the changes in frequencies of five virus variants under 750 selection and genetic drift regimes with a Wright-Fisher model for haploid individuals. The simulations were designed to fit the experimental setup of our datasets (48 independent host plants regularly analyzed at 6 sampling dates). For each of the 750 datasets obtained, the true parameters $\boldsymbol{\theta}_{true}$ were known and could be compared with the estimated parameters $\hat{\boldsymbol{\theta}}$. In experiment 2, we assessed the sensitivity of the estimation method to the presence of a sixth undetected virus variant. This sixth variant was selectively neutral (its selection coefficient is zero), present in the inoculum at a frequency of 3% and still present at the last sampling date (34 dpi) in all plants

analyzed, at frequencies ranging from 1% to 6%. It affected the dynamics of the five variants of interest in all plants but was not detected, so the variant frequencies measured by HTS (and used to estimate $\hat{\boldsymbol{\theta}}$) are noisy with respect to their true values. In all, 350 simulated datasets were analyzed in this second test.

Results

In this section, we will (i) describe the virus dynamics observed in the biological experiment with 15 plant genotypes, (ii) validate the method for estimating selection and genetic drift with numerical simulations and (iii) describe the estimates obtained in the biological experiment.

Virus variant dynamics in the 15 plant genotypes

The frequencies of the five virus variants were assessed in completely isolated populations during the course of infection, in 15 different plant genotypes. For each of these 15 pepper genotypes, 48 plants were inoculated with the same equimolar mixture of the five variants, and the frequencies of the virus variants were determined in eight plants at each of six sampling dates, from 6 to 34 days post-inoculation (Fig. 2). In a few cases, no viruses were detected in plant samples (lacking bars in Fig. 2). These negative samples may reflect the presence of an extreme bottleneck at inoculation, leading to virus population extinction, or a long time lag to systemic infection of the plant (for measurements from 10 to 34 dpi), resulting in the sampling of leaves not yet infected (e.g. DH line 2321). Negative samples were most frequent for the first two dates on which systemically infected leaves were analyzed, *i.e.* at 10 and 14 dpi, probably indicating a time lag to systemic infection in some DH lines. Negative samples were observed in only four DH lines (e.g. DH lines 219 and 2321). No infection was observed in a mean of 3.5 (resp. 2.0) plant samples 10 (resp. 14) dpi for the four DH

lines concerned.

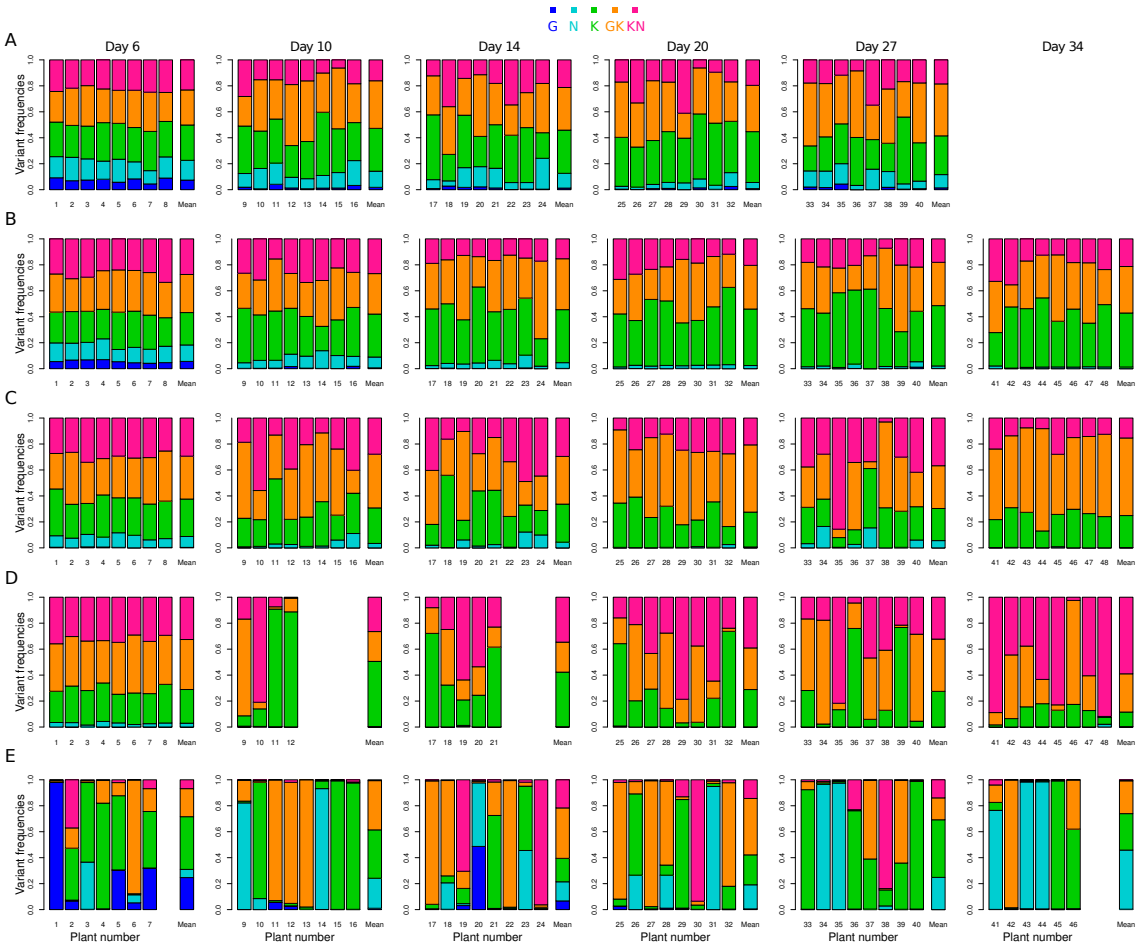


Fig 2. Five contrasting datasets obtained in the biological experiment. Each line of bar plots represents the dynamics of virus variants in a single DH line over time: (A) 240, (B) 2430, (C) 2344, (D) 2321 and (E) 219. We inoculated 48 plants per DH line, and we sampled eight plants, which were subsequently removed from the experiment, at each of the six sampling dates (6, 10, 14, 20, 27 and 34 days post-inoculation). Within each bar plot, the frequencies of the five variants (see top of the figure for the color code) in each infected plant sample are represented by single bars (labeled from 1 to 48). The missing bars correspond to plant samples for which no viruses were detected. The last bar indicates the mean viral composition in the infected plants. Each individual bar plot corresponds to a single sampling date, indicated at the top of each column of barplots. The five DH lines displayed contrasting virus variant dynamics, consistent with contrasting patterns of selection and genetic drift. We could not sample plants of DH line 240 (A) 34 days post-inoculation, because severe necrosis symptoms invading the stem led to the death of all plants at this sampling date.

The virus populations present in all infected plants and in the common inoculum were analyzed by HTS, to determine the frequencies of the five PVY variants. Inoculum analysis confirmed that all variants were present in roughly equimolar proportions, with

22.6% of variant G, 17.5% of N, 20.6% of K, 17.1% of GK and 22.2% of KN.

The raw data for variant frequency dynamics provided considerably different patterns between the 15 pepper genotypes (Fig. 2, Fig S2 and Fig S3). Variant frequencies were similar between virus populations sampled on the same date in some plant genotypes, consistent with weak genetic drift (e.g. DH lines 240 and 2430, Fig. 2A & B), whereas they differed in other plant genotypes (*e.g.* DH lines 2321 and 219, Fig. 2D & E). Furthermore, the heterogeneity of variant frequencies between the eight plants analyzed fluctuated between dates, probably due to changes in effective population size during the course of infection (*e.g.* DH line 2344, Fig. 2C). The four pepper genotypes for which some samples were virus-negative were also characterized by the highest heterogeneity in variant frequencies, consistent with an extreme bottleneck at inoculation (see DH lines 219, 2256, 2321 and 2400 in Fig S1). Selection regimes also differed between lines. In some DH lines, all variants remained present at all dates (*e.g.* DH line 240, Fig. 2A), whereas one variant (e.g. DH line 219, Fig. 2E), or up to two variants (*e.g.* DH lines 2430, 2344, 2321, Fig. 2B-D) became extinct in others.

Validation of the estimation method with numerical simulations

Before its application to the experimental dataset, we validated the estimation method proposed by numerical simulations of a Wright-Fisher model with selection and genetic drift for haploid individuals.

Range of selection and genetic drift intensities explored The 750 datasets generated in experiment 1 corresponded to very different selection and genetic drift regimes (Fig S4). Randomly sampled effective population sizes in the inoculated organ and at the onset of systemic infection were combined independently, to simulate highly diverse dynamics of effective population size (see Text S2). This led to the exploration of a large range of harmonic means of effective population size $N_e^h(t_d)$, ranging from 10

to 1996 (5% quantile = 22, mean = 230, 95% quantile = 873). Relative fitness values (r_i) ranged from 0.75 to 1.27, independently of effective population size (Fig S4). They reflect a mean absolute selection coefficient $|s|$ of 0.08 (5% quantile = 0.007, 95% quantile = 0.18). As a result, highly diverse simulated datasets were obtained. Overall, the patterns encountered in the experimental datasets (Fig. 2) were similar to some of those obtained for the simulated datasets. The simulated datasets also included a number of datasets with more extreme patterns of selection and genetic drift regimes. We illustrate the differences in the genetic drift regimes observed in Fig. 3. For a given dataset, variant frequencies could be roughly similar (Fig. 3A) or very different (Fig. 3D) between populations, at all sampling dates. Moreover, the independent sampling of effective population sizes in the inoculated organ and during systemic infection generated genetic drift regimes that varied over time. For example, we observed strong similarities between populations at the first sampling date, but greater heterogeneity at subsequent dates (Fig. 3B) and the opposite pattern (Fig. 3C).

Parameter estimation accuracy Effective population sizes $N_e^h(t_d)$ and intrinsic rates of increase of each variant r_i were inferred for each of the 750 datasets simulated in experiment 1 (with 5 virus variants) and the 350 datasets simulated in experiment 2 (with the 5 virus variants and an additional undetected sixth variant) using the more general model $\mathfrak{M}_4$. True parameters (*i.e.* known parameter values used in the simulations) and estimated parameter values were compared, to assess estimation accuracy.

In numerical experiment 1, HTS analysis provided samples of the true frequencies of the virus variants in the simulated Wright-Fisher populations. The estimates of the intrinsic rates of increase $\hat{r}_i$ were very accurate, with an R^2 of the best-fit line of 0.93, a slope close to 1 (0.98) and an intercept of 0.02 (Table 2, Fig. 4A). The estimates of the harmonic mean of the effective population size $\hat{N}_e^h(t_d)$ were also accurate, with a best-fit

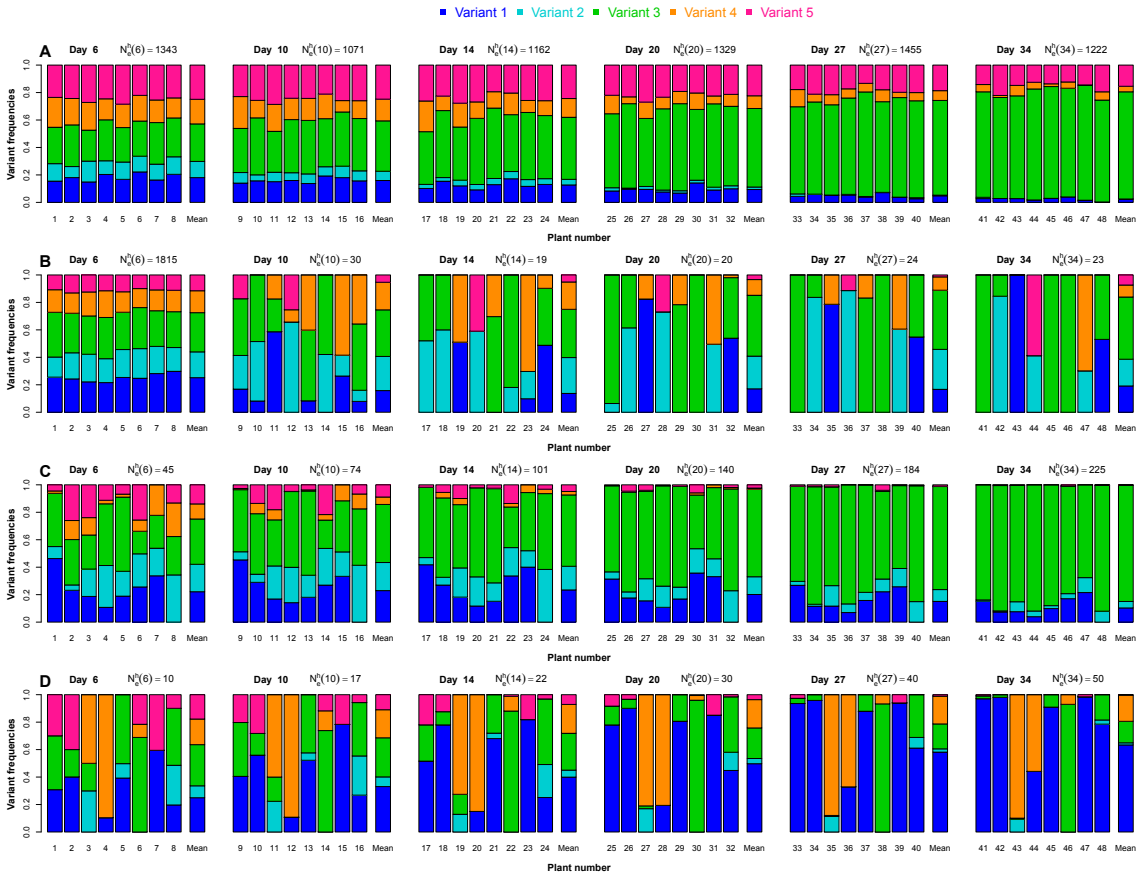


Fig 3. Contrasting datasets obtained in numerical experiment 1. For each dataset (series A to D), the composition of eight populations was observed at six sampling dates, from 6 to 34 days post-inoculation, in independently sampled hosts. Within each plot, each bar represents the composition of the population in one plant at one date, and the last bar shows the mean frequencies over these populations. The color code at the top is used to distinguish the five variants. The harmonic mean of effective population size is indicated in the main title of each plot. The parameter values used for the simulations are: series (A) $\boldsymbol{r} = (0.971, 0.92, 1.09, 0.992, 1.027)$, $N_e^{IO} = 1343$, $N_e^{S_1} = 822$; series (B) $\boldsymbol{r} = (1.05, 1.005, 1.077, 0.963, 0.904)$, $N_e^{IO} = 1815$, $N_e^{S_1} = 12$; series (C) $\boldsymbol{r} = (1.045, 1.031, 1.12, 0.879, 0.924)$, $N_e^{IO} = 45$, $N_e^{S_1} = 1473$; series (D) $\boldsymbol{r} = (1.105, 0.943, 0.999, 1.041, 0.912)$, $N_e^{IO} = 10$, $N_e^{S_1} = 1025$. Note that $N_e^{S_1}$ is used for the iterative computation of a sequence of effective population sizes varying each five generations during the systemic infection stage.

line close to the first bisector (Table 2, Fig. 4B) ($R^2 = 0.85$), despite a slight trend towards overestimation (slope=0.91, intercept=28). In both cases, mean relative bias was small and its 95% confidence interval included zero. The 90% confidence intervals of all estimated parameters were highly accurate. They included the true parameter values in nearly 90% (resp. 91 %) of the cases for $\hat{N}_e^h(t_d)$ (resp. $\hat{r}_i$) (Table 2).

Table 2. Performance of the estimators of the harmonic mean of effective population sizes $N_e^h(t_d)$ and variant fitness r obtained with the two numerical experiments

Experiment[a]	Parameter[b]	R^2	Intercept	Slope	Accuracy of 90% CI	Mean bias [95% CI]
Experiment 1	$\hat{r}_i$ (all variants)	0.93	0.02	0.98	91%	10^{-4} [-0.05;0.05]
Experiment 1	$\hat{N}_e^h$ (all dates)	0.86	28	0.91	90%	0.18 [-0.42; 1.48]
Experiment 2	$\hat{r}_i$ (all variants)	0.92	0.06	0.94	89%	8.10^{-4} [-0.07;0.07]
Experiment 2	$\hat{N}_e^h$ (all dates)	0.85	20	0.87	87%	0.03 [-0.53; 1.17]

[a] Experiment 1: 750 simulated datasets with 5 variants under a wide range of selection and genetic drift regimes. Experiment 2: 350 simulated datasets with an additional and undetected sixth variant.

[b] For each parameter, the determination coefficient R^2, the slope and the intercept of the best linear model fit between predicted and true values are given, together with the percentage of true parameter values included in a 90% confidence interval and the mean relative bias and its 95% confidence interval.

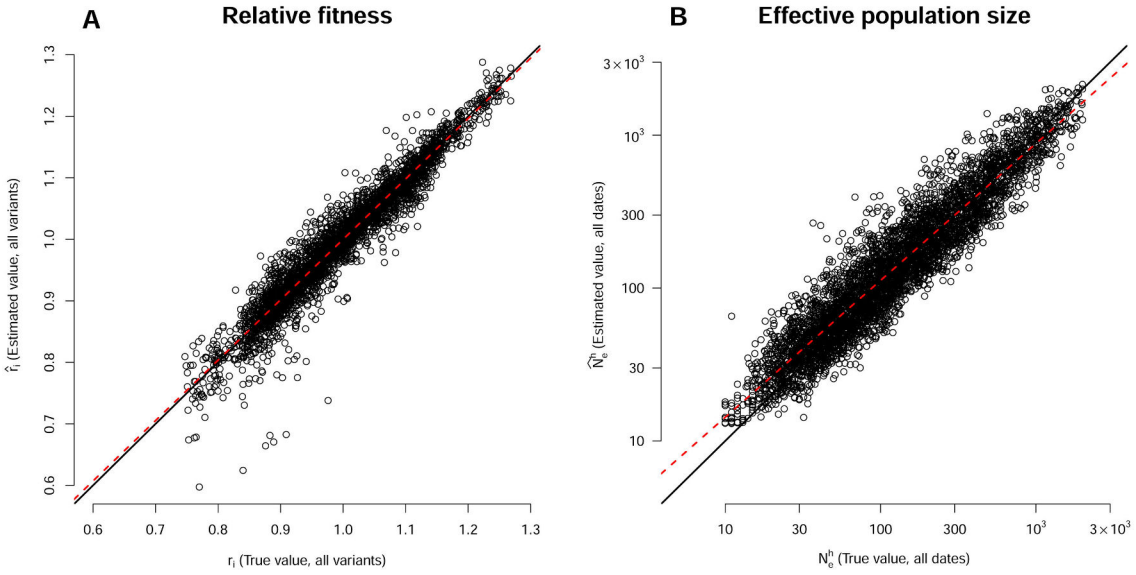


Fig 4. Inferences for variant fitness r and for the harmonic mean of effective population size $N_e^h(t_d)$, for the 750 datasets simulated with five virus variants. (A) Correlation between true r_i (x-axis) and estimated $\hat{r}_i$ (y-axis) (all variants considered together). (B) Correlation between true N_e^h (x-axis) and estimated $\hat{N}_e^h$ (y-axis) (all sampling dates considered together, logarithmic scale). In both panels, the black line is the first bisector and the red dashed line is the best-fitting linear model. In panel A, the 9 points with $\hat{r}_i < 0.7$ correspond to datasets in which a highly counterselected variant was observed in only a few plants (5, on average, of the 48 plants) due to an initial low effective population size.

In numerical experiment 2, we assessed the sensitivity of the estimation method to the presence of a sixth undetected virus variant (see Text S2). This additional variant was neutral and initially present in the inoculum at a frequency of 3%. It affected virus population dynamics in all 48 host plants of the dataset, because we retained only Wright-Fisher simulations in which the frequency of this sixth variant ranged from 0.01

to 0.06 at 34 dpi. In the 350 simulated datasets, the mean frequency of the sixth variant at all sampling dates and in all plants was 0.07 (5% quantile = 0.01, median = 0.04, 95% quantile = 0.24). However, this variant was considered to be undetected by the HTS method. Thus, HTS analysis provided noisy estimates of the true frequencies of the five virus variants of interest: the mean relative difference between their true frequencies in the simulated population and their measured frequencies was 0.08 (5% quantile = 0.01, median = 0.05, 95% quantile = 0.29). Moreover, inference was performed assuming, as in numerical experiment 1, that the inoculum was an equimolar mixture of the five variants of interest. Despite this detection bias, the estimates of both $N_e^h(t_d)$ and r_i remained consistent (Table 2). The systematic presence of an undetected virus variant at a mean frequency of 0.07 only slightly affected the performance of the estimators (mean relative bias confidence intervals systematically included zero, the R^2 of the best-fit lines remained unchanged, 90% of confidence intervals remained highly precise).

In a nutshell, from these numerical simulations with known parameter values, we can conclude that the proposed inference method provides accurate estimates of the intrinsic rates of increase r_i of each variant i, and, thus, of their selection coefficient, together with the dynamics of effective population size $N_e^h(t_d)$ during the time course of the experiment.

Estimation of effective population sizes and variant fitness in the 15 plant genotypes

We estimated the $N_e(t)$ and r_i of the PVY populations in each DH line with a Wright-Fisher model including selection and genetic drift. By contrast to the numerical experiments, the evolutionary parameters underlying the true dynamics of virus populations in their hosts were unknown. The Wright-Fisher model fitted the data very poorly (Fig. 5). The best-fit line between observed and fitted mean variant frequencies

(averaged over all virus populations and sampling times) was very close to the first bisector (Fig. 5A; slope = 0.92, intercept = 0.01, $R^2 = 0.92$). This was also the case for the variability of variant frequencies between virus populations at each sampling date t_d (Fig. 5B; slope = 0.92, intercept =-0.09, $R^2 = 0.84$). A Wright-Fisher model including selection and genetic drift accurately described the mean evolutionary dynamics of a virus population and the variability of these dynamics between hosts. Due to an identifiability issue (we observed the relative frequencies of variants rather than variant densities), we had to fix the number of generations per day γ. We set this number to 1, a value close to that reported by Khelifa *et al.* [39]. Different γ values would change r_i and $N_e(t)$ estimates to $r_i^{1/\gamma}$ and $\gamma N_e(t)$, but would have no effect on their ranking.

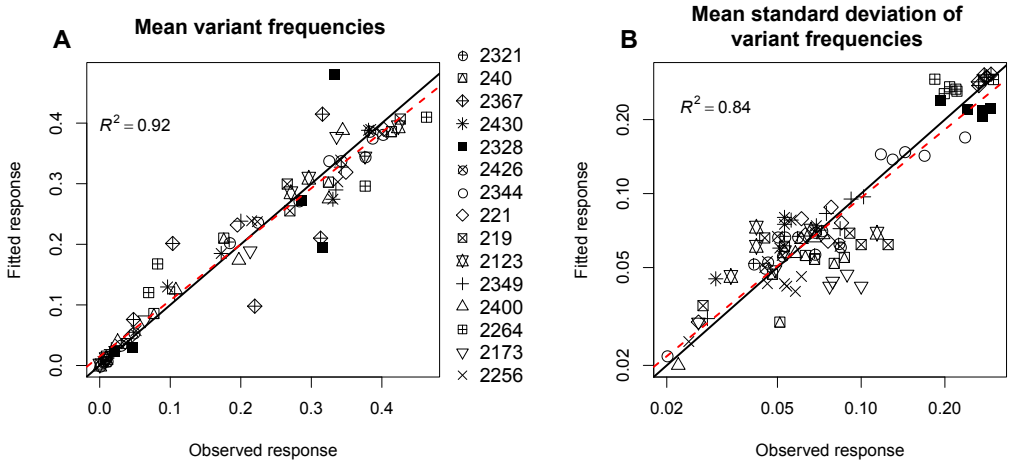


Fig 5. Goodness-of-fit of the Wright-Fisher model with the data of the biological experiments (A) Correlation between the observed mean frequencies of the five virus variants (averaged over all virus populations and sampling times($mean[f_i^\bullet(\bullet)]$) and their fitted values ($n = 75$). (B) Correlation between the logarithm of the observed mean (averaged over the variants) standard deviation of variant frequencies (between virus populations) at each sampling date t_d ($1/n_{var} \sum_{i=1}^{n_{var}} \sigma[f_i^\bullet(t_d)]$) and their fitted values with model $\mathfrak{M}_4$ ($n = 87$). In both panels, the black line is the first bisector and the red dashed line is the best-fitting linear model.

Relative fitness values (r_i) ranged from 0.43 to 1.25 (corresponding to $|s|$: 5% quantile=0.004, mean=0.12, 95% quantile=0.27) and were associated with narrow 90% confidence intervals (Table S3). The fitness ranks of the PVY variants were very similar in most DH lines (Fig. 6A,C). Variant G was the weakest in all DH lines, followed by

variant N in 13 DH lines. Variant GK was the fittest variant in 13 DH lines, with variant K the fittest variant in the remaining two lines (DH lines 2256 and 2430). Overall, variants K and GK are the two fittest variants in 12 DH lines; variants GK and KN are the two fittest in DH lines 2349 and 2321, and variants N and GK the two fittest in DH line 219. The fitness difference between the weakest and the fittest variants ranged from 0.14 for DH line 219, to 0.81 for DH line 2349.

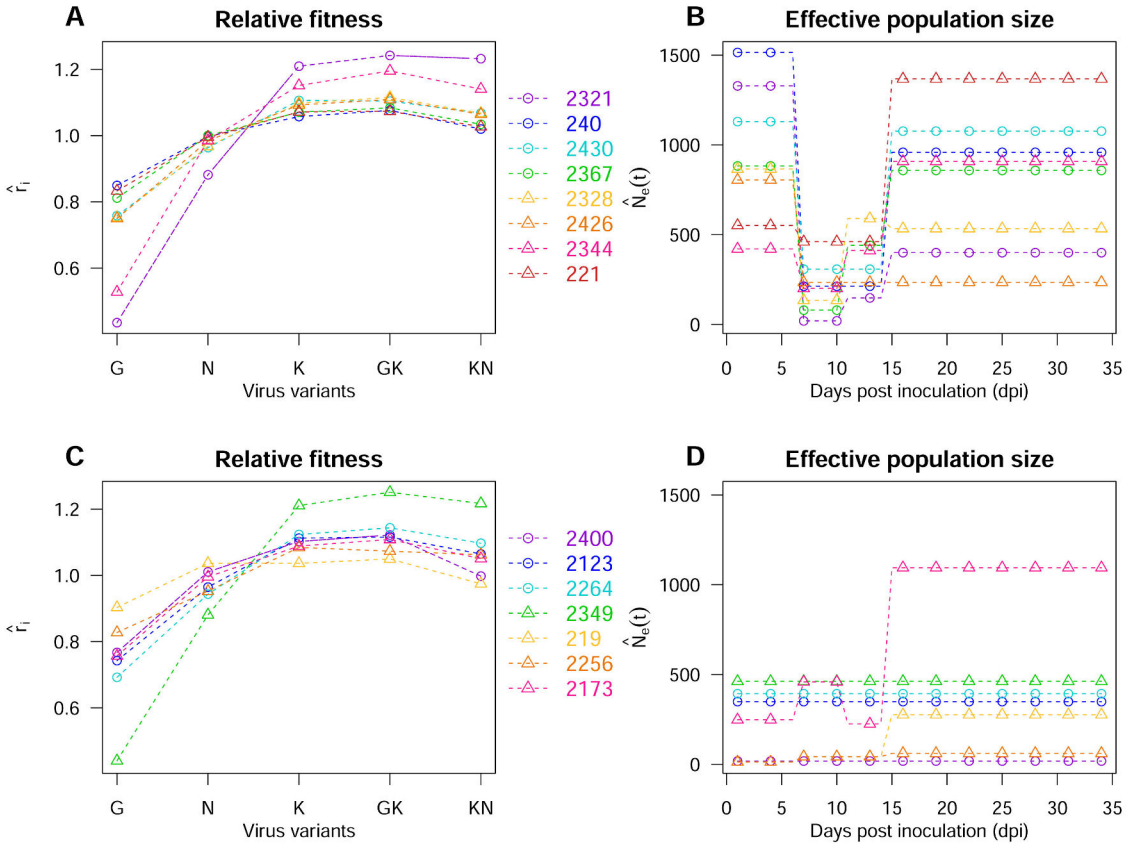


Fig 6. Fitness of virus variants and effective population size estimates for the 15 plant genotypes. (A) Estimates of intrinsic rates of increase $\hat{r}_i$ for each variant i for the DH lines 2321, 240, 2430, 2367, 2328, 2426, 2344 and 221. (B) Estimates of effective population size $\hat{N}_e(t)$ during the time course of the experiment for the DH lines listed in (A) and the model best supported by the data. (C) As for (A) for DH lines 2400, 2123, 2264, 2349, 219, 2256 and 2173. (D) As for (B) for the DH lines listed in (C).

We further estimated the dynamics of effective population size over the time course of the experiment, as modeled by a piecewise function $N_e(t)$, using a model selection procedure. Four models with from one to four parameters were considered. The most

general model $\mathfrak{M}_4$ distinguished four successive effective population sizes (one in the inoculated organ and three during systemic infection). $\mathfrak{M}_4$ was the model best supported by the data for five DH lines (2173, 2321, 2328, 2344 and 2367). Model $\mathfrak{M}_3$ distinguished three successive effective population sizes (one in the inoculated organ and two during systemic infection). It was best supported by the data for five DH lines (219, 221, 2256, 240 and 2430). Model $\mathfrak{M}_2$, which distinguished two successive effective population sizes (one in the inoculated organ and one during systemic infection), was selected for a single DH line (2426). Finally, with $\mathfrak{M}_1$, the effective population size of the virus population remained constant. This model was selected in the four remaining DH lines (2123, 2264, 2349 and 2400). The corresponding posterior probabilities of each model are shown in Table S4, together with effective population size estimates and 90% credibility intervals.

At the first sampling date, considerable variability was observed (Fig. 6B,D), with effective population sizes ranging from 13 for DH lines 219 and 2256 to 1515 for DH line 240. This was not surprising, given that we chose the DH lines on the basis of the density of primary infection foci in inoculated organs (Fig S1). A much narrower range of effective population sizes, from 20 to 461, was observed across all plant genotypes at 10 dpi, the first date on which systemic infection was observed. From 6 to 10 dpi, effective population sizes decreased in eight DH lines (Fig. 6B), remained approximately constant in six DH lines (Fig. 6D) and increased slightly in a single plant genotype (DH line 2173, Fig. 6D). Later on, from 10 to 34 dpi, effective population size increased in eight DH lines (mostly DH lines displaying a bottleneck from 6 to 10 dpi, Fig. 6B) and remained approximately constant in the others (mostly in DH lines with lower, *i.e.* < 500, effective population sizes in the inoculated organ, Fig. 6D).

Heritability of the intensities of selection and genetic drift exerted by plants on virus populations

By creating two dataset replicates of 24 randomly chosen plants for each DH line, we estimated the heritability of two plant traits corresponding to the evolutionary forces exerted by the plant on virus populations: selection and genetic drift. These forces were estimated by (i) intrinsic rates of increase in viral variants and (ii) effective population sizes for PVY. With 24 plants in each dataset, we used the function $N_e(t)$ of model $\mathfrak{M}_2$ with two parameters . In this approach, we used the contrasting behavior of PVY populations, which were fixed and identical at the time of inoculation in all plants, on different pepper genotypes to characterize the phenotype of each host. Very high heritability estimates were obtained for the intrinsic rates of increase (mean heritability over the five variant estimates: $h^2 = 0.94$). Somewhat lower, but nevertheless substantial heritability estimates were obtained for effective population size in the inoculated organ (mean heritability, $h^2 = 0.64$) and for effective population size during systemic infection (mean heritability, $h^2 = 0.63$). The details of the calculation are provided in supplementary Text S3.

Discussion

Advances in sequencing technologies are revolutionizing the study of microbial evolution [13]. To our knowledge, this study is, for example, the first to suggest such strong variability in the selection and genetic drift regimes experienced by plant viruses in closely related host genotypes (Fig. 2). This new type of data paves the way for the estimation of population genetics parameters influencing the fate of pathogen variants of special interest in medicine and agriculture (e.g. variants resistant to pesticides and drugs [43] or, as in this study, variants adapted to host resistance genes). However,

estimation methods encompassing the whole range of variation of these parameters are still lacking.

A method for estimating genetic drift and selection from microbial experimental evolution

We present here a method for the estimation of selection and genetic drift in a haploid and asexual organism, as modeled by a Wright-Fisher process. As for any model-based approach, the population of interest must not be too far from an ideal Wright-Fisher model with suitable parameters [10]. The estimation method did not require neutral markers. It was validated for small effective population sizes ($N_e \ll 1000$) and a wide range of both positive and negative selection coefficients (weak ($|s| \simeq 0.01$) or strong ($|s| \simeq 0.15$) selection), using simulated datasets. Recent reviews [23, 32] have highlighted the small number of methods available for the inference of selection and genetic drift over the whole range of variation, particularly in the case of small effective population sizes ($N_e \ll 1000$) and strong selection coefficients ($|s| \simeq 0.1$). Indeed, these conditions do not fulfill the hypothesis underlying most approximations of the Wright-Fisher model. The classical approximation, with a standard diffusion process, requires both selection and genetic drift to be weak [23]. Approximations based on Gaussian diffusion require the stochastic effects of genetic drift to decrease more rapidly than the effects of selection [23]. The work of Foll *et al.* [11, 32] constituted a major step forward, but their method requires a large proportion of the genetic markers studied to be neutral. This assumption is not valid for many pathogens with small genomes, such as viruses. For example, only 22.7% of 66 randomly chosen mutations in the genome of *Tobacco etch virus* (TEV, genus *Potyvirus*), a plant RNA virus, were found to be consistent with neutrality [44]. As the statistical power to detect departure from neutrality is limited, the true proportion of neutral mutations is probably much lower. Similar results have

been obtained for bacteria (e.g. [45]).

The estimation method proposed does not require neutral markers, an appealing feature for studying pathogens with small genomes. Lacerda and Seoighe [46] recently developed another method that does not require neutral markers. Their method provided satisfactory estimates of both N_e and s (estimated at a single locus) for a relatively small effective population size of 1000 individuals and values of s up to 0.5. They did not test the performance of their method for $N_e \ll 1000$. By comparison, the method developed here was effective for much lower N_e values, in the range of a few tens of individuals, and for inferring the time course of N_e over a few tens of generations. However, although the range of selection coefficients s included cases of strong selection ($|s| \simeq 0.1$, as defined by Malaspinas [23]), none of the simulation experiments included values as high as 0.5. It may be possible to infer such high selection coefficients with the estimation method proposed, provided that the first generations are sampled more densely, typically every day after inoculation in our set-up. Lacerda and Seoighe [46], for example, used samples taken at each generation, for 20 generations. This makes it possible to record the trajectories of variant frequencies before variant loss or fixation.

The use of the proposed estimation method requires observation of the evolution of isolated populations derived from the same parental population, each population being sampled only once. This design is particularly suitable for studying within-host microbial evolution, as several hosts (48 plants for each pepper genotype in our case study) can easily be included in the experiment. With this experimental design, we observed a set of variant frequencies at several time points, in independent hosts. This set contained footprints of selection and genetic drift. In the method developed, selection is evaluated from the mean trajectories of variant frequency. Genetic drift is evaluated at several time points, by assessing differences in variant frequencies between the replicated populations during the time-course of the experiment. Even for

populations with small effective sizes, for which genetic drift and selection have confounding effects on the fate of variants (Fig. 2), a moderate number of replicates contains sufficient information to disentangle the two mechanisms. Here, we estimated four selection coefficients and four effective population sizes (*i.e.* 8 parameters) with 48 samples (6 sampling dates $\times$ 8 replicates).

The proposed estimation method could be improved further. It explicitly accounts for the technical sampling noise resulting from the assessment of variant frequencies from finite counts of virus sequences. However, HTS also introduces sequencing errors, albeit at a low rate of about 1 substitution per 400 bases for MiSeq technology [47], which were not explicitly accounted for in our framework. Several models have been proposed for separating true genetic variation from technical artifacts [47], and these models could be integrated into the method through a hierarchical Bayesian modeling framework [48], for example. Finally, the method could be extended to take mutation and recombination into account, particularly for experiments over longer periods, in which new variants might appear and displace those currently most abundant. In our short-term experiment, we have already observed *de novo* substitutions in a few plants (removed plant samples, see Text S1). The inclusion of recombination is not relevant for our case study, as the nucleotide positions differentiating the variants are located only a few codons apart. Recombination can thus be ignored in this study [49], particularly given the small number of generations considered [32].

Plant genotypes modulate genetic drift and selection within virus populations.

On the host side, our experiment involved 15 DH lines of pepper, all carrying the major resistance gene *pvr2*^{3}, but differing in terms of their genetic backgrounds [12]. These DH lines were derived from the F_1 hybrid between two pepper lines, Perennial and Yolo

Wonder. Consequently, on average, any pair of DH lines have 50 percent of alleles in common for markers differentiating between Perennial and Yolo Wonder. This is the first study, to our knowledge, to show such a high level of diversity in selection and genetic drift regimes experienced by virus populations from the same viral inoculum in closely related host genotypes (Fig. 2, Fig S2 and Fig S3). On the pathogen side, we used five virus variants: the G and N variants displayed weaker adaptation to *pvr2*^{3} than the K, GK and KN variants. The ranking of the selection coefficients of the five variants was mostly identical in the 15 plant genotypes. We were therefore unable to identify any host genotype, among those tested, able to counterselect against the virus variants best adapted to *pvr2*^{3}. This may be due to (i) the strong selective effect exerted by the major-effect resistance gene *pvr2*^{3}, which is present in all the DH lines studied here and probably exceeds the additional selective effect of the plant genetic background and/or (ii) the close genetic relatedness of the DH lines analyzed. Other genetic resources for pepper should be explored, to identify genotypes capable of counterselecting against the K, GK and KN variants, which were the fittest in our study. The best candidates for this would be pepper genotypes carrying *pvr2* resistance alleles other than *pvr2*^{3}, with a different specificity in the face of PVY diversity [50], or pepper genotypes devoid of resistance alleles at the *pvr2* locus, as shown by Quenouille *et al.* [12]. Combinations of plant genotypes exerting opposite selective pressures on pathogen populations are particularly interesting for the sustainable management of plant resistance at landscape level, and can be implemented in cultivar rotations, mixtures or mosaics [51]. However, in our study, the difference in fitness between the weakest and fittest variants differed between host genotypes. The dynamics of selection for the fittest variants were under plant genetic control and could therefore be modulated by the choice of plant genotypes grown. For example, growing the pepper DH lines with the smallest differential selection between the five PVY variants would be

particularly useful for delaying PVY adaptation in *pvr2*^{3}-carrying plants, in which a two-step mutational trajectory may be required [12]. Indeed, the G and N variants are most likely to appear initially, because they require transitions, whereas the K variant requires a transversion, and transitions are more frequent than transversions [52]. However, an additional substitution, in a second step, is required to confer a sufficient level of fitness for the emergence of GK and KN variants. These mutational trajectories were observed in PVY adaptation to the Perennial pepper genotype, the resistant parent of all the DH lines studied here [12].

We also inferred the time course of the genetic drift experienced by the viruses in the 15 host environments during the experiment. Genetic drift intensities were highly variable with time and between plant genotypes, revealing an unprecedented level of variability between closely related host genotypes. Our estimates of $N_e(t)$ ranged from 20 to 458 just after the colonization of apical leaves at 10 dpi, and from 13 to 1469 in the inoculated leaves four days previously (at 6 dpi). Eight of the 15 DH lines displayed a high N_e in the inoculated leaves at 6 dpi (from 421 to 1515), a decrease at 10 dpi (Ne (10 dpi) values of 1.5 to 83.5 % of the value at 6 dpi) and a subsequent increase (Fig. 6B). This pattern suggests a founder effect, in which a new PVY population in apical leaves is set up by a few members of the original population in the inoculated leaf. In the remaining seven DH lines, the N_e of the inoculated leaves at 6 dpi was much lower (from 13 to 462), and N_e values often remained low in the apical leaves (Fig. 6D). However, an increase in N_e was observed in DH lines 219 and 2173, after 14 dpi. This result sheds new light on the importance of the within-host bottlenecks experienced by virus populations, as discussed in a recent article by Zwart *et al.* [53], who reported that the N_e of TEV in the first systemically infected leaf of tobacco plants was determined largely by inoculum viral load. They then hypothesized that genetic drift occurred mostly during the inoculation process. Previous estimations of N_e for viruses did not

focus on N_e dynamics at the whole-plant study as in this study. Instead, they considered the multiplicity of infection (MOI) during cell-to-cell movement or N_e during the colonization of apical leaves (for a comprehensive review, see Gutiérrez *et al.* [4]). Direct comparisons with these studies are, therefore, not appropriate. Gutiérrez *et al.* [54] recently showed that *turnip mosaic virus* (genus *Potyvirus*) infections are characterized by a very low MOI ($\simeq 1$) when cells are infected with virus particles moving in the plant vasculature, and a much higher MOI ($\simeq 30$) during subsequent cell-to-cell movement in the mesophyll. The general picture that emerges when we consider both these MOI patterns and plant growth dynamics is consistent with our observations. Indeed, the lowest N_e values were observed at 10 dpi, corresponding to the onset of systemic infection, when plants were small and consisted essentially of a few infected leaves. N_e tends often to increase with time, because (i) increasing numbers of leaves are infected and behave as virus sources as the plant grows and (ii) leaf areas increase, probably increasing the relative proportion of cell-to-cell, as opposed to long-distance, virus movement.

One of the key results of this study is the finding that the effective population size of PVY is a heritable plant trait. The high heritability estimated for N_e (partially due to the use of a DH progeny of pepper genotypes) indicates that plant resistance could potentially be improved through breeding programs. Indeed, our findings pave the way for the breeding of plant cultivars exposing viruses to greater genetic drift. This would provide a twofold benefit against viruses. First, in asexual populations, genetic drift favors the accumulation of deleterious mutations, decreasing viral fitness (Muller's ratchet) [55]. Second, genetic drift decreases the fixation probability of beneficial mutations, such as those responsible for overcoming plant resistance genes [56]. Breeding for greater genetic drift in virus populations would thus constitute a novel approach to increasing the durability of resistance to plant viruses in agricultural

landscapes [51, 57, 58]. Another key result is the finding that the Wright-Fisher model accurately captures the major processes driving the within-host dynamics of a set of virus variants (Fig. 5), despite being much simpler than the underlying mechanisms involved in the infection of highly structured hosts. Over longer periods, mutation and recombination increase in importance and this can easily be encompassed in the Wright-Fisher model [59]. This model can thus serve as a valuable cornerstone for linking the within- and between-host scales of disease dynamics and studying, for example, how breeding for greater genetic drift can delay the emergence of a new pathogen variant.

Supporting Information

Text S1 Sequence analyses to detect PVY mutations

Text S2 Numerical experiments

Text S3 Heritability of the intensity of selection and genetic drift exerted by plants on virus populations

Fig S1. Resistance-breakdown (RB) frequency, viral accumulation and mean number of primary infection foci for the 15 DH lines studied. Pepper genotypes are represented as points, with their nomenclature (DH line number) given above each point. We estimated the mean number of primary infection foci for the 15 DH lines with the *Potato virus Y* (PVY, genus *Potyvirus*) variant K, carrying a green fluorescent marker (green fluorescent protein, GFP). The resistance-breakdown (RB) frequency and the relative viral accumulation were estimated by Quenouille *et al.* [12]. The RB frequency corresponds to the percentage of infected plants when inoculated with an avirulent variant regarding the allele of resistance $pvr2^3$, carried by all DH lines.

The relative viral accumulation, or relative viral concentration, was measured by double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA).

Fig S2 Five datasets obtained by high-throughput sequencing in the biological experiment. Each line of bar plots represents the dynamics of virus variants in a single DH line over time: (A) 221, (B) 2123, (C) 2173, (D) 2256 and (E) 2264. Within each bar plot, the frequencies of the five variants (see top of the figure for the color code) in each infected plant sample are represented by single bars (labeled from 1 to 48). The missing bars correspond to plant samples for which no viruses were detected. The last bar indicates the mean viral composition in the infected plants. Each individual bar plot corresponds to a single sampling date, indicated at the top of each column of barplots.

Fig S3 Five datasets obtained by high-throughput sequencing in the biological experiment. Each line of bar plots represents the dynamics of virus variants in a single DH line over time: (F) 2328, (G) 2349, (H) 2367, (I) 2400 and (J) 2426. Within each bar plot, the frequencies of the five variants (see top of the figure for the color code) in each infected plant sample are represented by single bars (labeled from 1 to 48). The missing bars correspond to plant samples for which no viruses were detected. The last bar indicates the mean viral composition in the infected plants. Each individual bar plot corresponds to a single sampling date, indicated at the top of each column of barplots.

Fig S4 Variability of the selection and genetic drift regimes obtained among the simulated datasets in numerical experiment 1. In the diagonal, each histogram represents the distribution of input parameters r_1 (intrinsic rate of increase of variant 1), N_e^{IO} (effective population size in the inoculated organ) and $N_e^{S_1}$ (effective population size at the onset of the systemic infection) used to simulate the 750

datasets. Off-diagonal scatter plots are two by two combinations of parameters.

Table S1 Tag sequences used to distinguish each plant sample after pooling and MiSeq Illumina high-throughput sequencing. The forward (Fwd.) primer sequence was the same for all amplifications and was bound to the sequence tag, just after it. Its binding site corresponds to positions 5971 to 5990 of PVY isolate SON41p (accession number AJ439544). The binding site of the reverse (Rev.) primer sequence corresponds to positions 6095 to 6114 of PVY isolate SON41p. RT-PCR amplifications were done according to the following profile: 1h at 42°C, 10 min at 95°C, 35 times the following sequence (45s at 95°C, 30s at 50°C and 20s at 72°C) and finally 10 min at 72°C.

Table S2 Number of sequences and composition of the virus populations in each sample of the biological experiment. In all, 708 samples were analyzed: 15 doubled-haploid (DH) lines of pepper × 6 sampling dates (dpi: days post-inoculation) × 8 plants per date, except for virus-negative samples, and 4 samples for the initial inoculum. Columns indicate (i) the name of each DH line, (ii) the sampling date in dpi, (iii) the number of the sequence tag used (see Table S1), (iv) the plant number (as in Fig 2, Fig S2 and Fig S3), (v) the infection status of each sample (0: not infected / 1: infected), (vi) the number of cleaned sequences assigned to each sample after filtering and (vii-xi) the number of sequences for each viral variant (G, N, K, GK and KN).

Table S3 Estimations of the relative intrinsic rates of increase of the virus variants for the 15 plant genotypes. Virus variants are indexed as follow: (i) r_1 virus variant G, (ii) r_2 virus variant N, (iii) r_3 virus variant K, (iv) r_4 virus variant GK, (v) r_5 virus variant KN. The 90% confidence intervals are calculated as $\hat{r_i} \pm 1.645.\hat{\sigma_i}$

Table S4 Model selection and estimations of the effective population sizes

for the 15 plant genotypes. The posterior probabilities of the four models considered for the piecewise function describing the temporal variation of the effective population sizes during the time course of the experiment (models $\mathfrak{M}_1$, $\mathfrak{M}_2$, $\mathfrak{M}_3$ and $\mathfrak{M}_4$) are first indicated. The bold value corresponds to the model that is best supported by the data. The next columns indicate the estimation of the effective population sizes of the model selected and the extent of a 90% credibility intervals.

Acknowledgments

This work was supported by the SMaCH (Sustainable Management of Crop Health) metaprogramme of INRA and by a grant over-seen by the French National Research Agency (ANR) as part of the "Blanc2013" program (ANR-13-BSV7-0011, FunFit project). We would like to thank Fabio Zanini (Max Planck Institute) and Nicolas Parisey (INRA Rennes) for their invaluable help with FFPopSim and Rcpp softwares, respectively, used in an earlier version of this manuscript. We also would like to thank Grégory Girardot and Baptiste Lederer for their precious help during experiments. The simulations were carried out with the Avakas (Bordeaux University) and Migale (INRA Jouy en Josas) computer clusters.

Author Contributions

Conceived and designed the experiments: ER BM AP FF. Performed the experiments: ER BM VS SV FF. Analyzed the data: ER BM LM FG FF. Contributed reagents/materials/analysis tools: ER BM RS AP VS SV FF. Wrote the paper: ER BM LM RS FG FF

References

1. Charlesworth B. Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics. 2009;10:195–205.
2. Wright S. Evolution in Mendelian populations. Genetics. 1931;16(2):97–159.
3. Vucetich JA, Waite TA, Nunney L. Fluctuating population size and the ratio of effective to census population size. Evolution. 1997;51(6):2017–2021.
4. Gutiérrez S, Michalakis Y, Blanc S. Virus population bottlenecks during within-host progression and host-to-host transmission. Current Opinion in Virology. 2012;2:546–555.
5. Waples RS, Antao T, Luikart G. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. Genetics. 2014;197:769–780.
6. Wright S. Statistical genetics in relation to evolution. In: Hermann, editor. Exposés de Biométrie et de Statistique Biologique. Paris; 1939.
7. Kimura M, Crow JF. The measurment of effective population number. Evolution. 1963;17(3):279–288.
8. Caballero A. Developments in the prediction of effective population size. Heredity. 1994;73:657–679.
9. Motro U, Thomson G. On heterozygosity and the effective size of populations subject to size changes. Evolution. 1982;36(5):1059–1066.
10. Rouzine IM, Rodrigo A, Coffin JM. Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application in virology. Microbiol Mol Biol Rev. 2001;65:151–185.

11. Foll M, Poh YP, Renzette N, Ferrer-Admettla A, Bank C, Shim H, et al. Influenza virus drug resistance: a time-sampled population genetics perspective. PLoS Genet. 2014;10(2):e1004185.

12. Quenouille J, Montarry J, Palloix A, Moury B. Farther, slower, stronger: how the plant genetic background protects a major resistance gene from breakdown. Molecular Plant Pathology. 2013;14:109–118.

13. Brockhurst MA, Colegrave N, Rozen DE. Next-generation sequencing as a tool to study microbial evolution. Molecular Ecology. 2011;20:972–980.

14. Sanjuán R, Moya A, Elena SF. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. PNAS. 2004;101(22):8396–8401.

15. Elena SF, Fraile A, García-Arenal F. Evolution and Emergence of Plant Viruses. Murphy KMFA, editor. Burlington: Academic Press; 2014.

16. Nei M, Tajima F. Genetic drift and estimation of effective population size. Genetics. 1981;98:625–640.

17. Waples RS. A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics. 1989;121:379–391.

18. Williamson EG, Slatkin M. Using maximum likelihood to estimate population size from temporal changes in allele frequencies. Genetics. 1999;152:755–761.

19. Anderson EC, Williamson EG, Thompson EA. Monte Carlo evaluation of the likelihood for N_e from temporal spaced samples. Genetics. 2000;156:2109–2118.

20. Berthier P, Beaumont MA, Cornuet JM, Luikart G. Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. Genetics. 2002;160:741–751.

21. Vitalis R, Gautier M, Dawson KJ, Beaumont MA. Detecting and measuring selection from gene frequency data. Genetics. 2014;196:799–817.

22. Illingworth CJR, Parts L, Schiffels S, Liti G, Mustonen V. Quantifying selection acting on a complex trait using allele frequency time series data. Molecular Biology and Evolution. 2012;29(4):1187–1197. Available from: http://mbe.oxfordjournals.org/content/29/4/1187.abstract.

23. Malaspinas AS. Methods to characterize selective sweeps using time serial samples: an ancient DNA perspective. Molecular Ecology. 2016;25(1):24–41. Available from: http://dx.doi.org/10.1111/mec.13492.

24. Bollback JP, York TL, Nielsen R. Estimation of $2N_e s$ from temporal allele frequency data. Genetics. 2008;179:497–502.

25. Malaspinas AS, Malaspinas O, Evans SN, Slatkin M. Estimating allele age and selection coefficient from time-serial data. Genetics. 2012;192:599–607.

26. Mathieson I, McVean G. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. Genetics. 2013;193:973–984.

27. Steinrücken M, Bhaskar A, Song YS. A novel spectral method for inferring general diploid selection from time series genetic data. Ann Appl Stat. 2014 12;8(4):2203–2222. Available from: http://dx.doi.org/10.1214/14-AOAS764.

28. French RC, Stenger DC. Evolution of Wheat streak mosaic virus: dynamics of population growth within plants may explain limited variation. Annu Rev Phytopathol. 2003;41:199–214.

29. García-Arenal F, Fraile A, Malpica JM. Variation and evolution of plant virus populations. Int Microbiol. 2003;6:225–232.

30. Sacristán S, Malpica JM, Fraile A, García-Arenal F. Estimation of population bottlenecks during systemic movement of *Tobacco mosaic virus* in tobacco plants. J Virol. 2003;77(18):9906–9911.

31. Elena SF, Bedhomme S, Carrasco P, Cuevas JM, de la Iglesia F, Lafforgue G, et al. The evolutionary genetics of emerging plant RNA Viruses. MPMI. 2011;24(3):287–293.

32. Foll M, Shim H, Jensen JD. WFABC : a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. Molecular Ecology Resources. 2014; 15(1): 87-98.

33. Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. PLoS Genet. 2011;7(3):e1001336. Available from: `http://dx.doi.org/10.1371%2Fjournal.pgen.1001336`.

34. Fabre F, Montarry J, Coville J, Senoussi R, Simon V, Moury B. Modelling the evolutionary dynamics of viruses within their hosts : a case study using high-throughput sequencing. PLoS Pathog. 2012;8:1–9.

35. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biology. 2007;8(7):R143 Available from: `10.1186/gb-2007-8-7-r143`

36. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27(21):2957–2963.

37. R Core Team. R: a language and environment for statistical computing. Vienna, Austria; 2013. Available from: `http://www.R-project.org/`.

38. Gallais A. Théorie de la sélection en amélioration des plantes. Paris: Masson; 1990.

39. Khelifa M, Massé D, Blanc S, Drucker M. Evaluation of the minimal time of *Cauliflower mosaic virus* in different hosts. Virology. 2010;396:238–245.

40. Ewens WJ. Mathematical population genetics 1 - Theoretical introduction. Antman S, Marsden J, Sirovich L, Wiggins S, editors. Springer-Verlag; 2004.

41. Musso F. On the relation between the Eigen model and the asexual Wright–Fisher model. Bulletin of Mathematical Biology. 2012;74 (1):103–115.

42. Lenormand M, Jabot F, Deffuant G Adaptive approximate Bayesian computation for complex models. Comput. Stat., 28, 2777–2796 Computational Statistics. 2013;28:2777-–2796

43. Consortium, R. E. X. Heterogeneity of selection and the evolution of resistance. Trends in Ecology & Evolution. 2013;2:110–118. Available from: http://www.sciencedirect.com/science/article/pii/S0169534712002352.

44. Carrasco P, de la Iglesia F, Elena SF. Distribution of fitness and virulence effects caused by single-nucleotide substitutions in *Tobacco etch virus*. J Virol. 2007;81(23):12979–12984.

45. Kassen R, Bataillon T. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nature Genetics. 2006;38:484–488. 4. Available from: ://000236340500024.

46. Lacerda M, Seoighe C. Population genetics inference for longitudinally-sampled mutants under strong selection. Genetics. 2014;198(3):1237–1250. Available from: http://genetics.org/content/198/3/1237.

47. Laehnemann D, Borkhardt A, McHardy AC. Denoising DNA deep sequencing data— high-throughput sequencing errors and their correction. Briefings in Bioinformatics. 2016;17(1):154–179.

48. Clark JS. Why environmental scientists are becoming Bayesians. Ecology Letters. 2005;8:2–14. 1. Available from: ://000225750300001.

49. Rhodes TD, Nikolaitchik O, Chen J, Powell D, Hu WS. Genetic recombination of Human immunodeficiency virus Type 1 in one round of viral replication: Effects of genetic distance, target cells, accessory genes, and lack of high negative interference in crossover events. J Virol. 2005;79(3):1666–1677.

50. Moury B, Janzac B, Ruellan Y, Simon V, Khalifa MB, Fakhfakh H, Fabre F, Palloix A. Interaction Patterns between *Potato Virus Y* and eIF4E-Mediated Recessive Resistance in the *Solanaceae*. J Virol. 2014;88(17):9799–9807. Available from: 10.1128/JVI.00930-14.

51. Djidjou-Demasse R, Moury B, Fabre F. Mosaics often outperform pyramids: insights from a model comparing strategies for the deployment of plant resistance genes against viruses in agricultural landscapes. New Phytologist. 2017;216:239–253. Available from: 10.1111/nph.14701

52. Ayme V, Souche S, Caranta C, Jacquemond M, Chadoeuf J, Palloix A, Moury B. Different mutations in the genome-linked protein VPg of *Potato virus Y* confer virulence on the *pvr2*^{3} resistance in pepper. Mol Plant Microbe Interact. 2006;19:557–563.

53. Zwart MP, Daròs JA, Elena SF. One is enough: *in vivo* effective population size is dose-dependent for a plant RNA virus. PLoS Pathog. 2011;7:1–12.

54. Gutiérrez S, Pirolles E, Yvon M, Baecker V, Michalakis Y, Blanc S. The multiplicity of cellular infection changes depending on the route of cell infection

in a plant virus. J Virol. 2015;89(18):9665–9675. Available from: http://jvi.asm.org/content/89/18/9665.

55. de la Iglesia F, Elena SF. Fitness declines in *Tobacco etch virus* upon serial bottleneck transfer. J Virol. 2007;81(10):4941–4947.

56. Patwa Z, Wahl LM. The fixation probability of beneficial mutations. J R Soc Interface. 2008;5:1279–1289.

57. Fabre F, Rousseau E, Mailleret L, Moury B. Durable strategies to deploy plant resistance in agricultural landscapes. New Phytologist. 2012;193:1064–1075.

58. Fabre F, Rousseau E, Mailleret L, Moury B. Epidemiological and evolutionary management of plant resistance: optimizing the deployment of cultivar mixtures in time and space in agricultural landscapes. Evolutionary Applications. 2015;8(10):919–932. Available from: http://dx.doi.org/10.1111/eva.12304.

59. Zanini F, Neher RA. FFPopSim: an efficient forward simulation package for the evolution of large populations. Bioinformatics. 2012;28 (24):3332–3333.

Supplementary Information – Text S1: Sequence analyses to detect PVY mutations

Viruses being characterized by high mutation rates, we conducted a sequence analysis to detect potential mutations in virus populations representing the common inoculum and the infected plants. In all, 677 samples were analyzed, corresponding to all infected plants from the 15 doubled-haploid (DH) lines of pepper, and 4 samples representing replicates of the initial inoculum. This analysis is important because the presence of mutants could affect the dynamics of virus populations and the intensities of the evolutionary forces at stake. We analyzed each nucleotide position in all sequences, starting from the end of the forward primer until the beginning of the reverse primer, i.e. 99 nucleotide positions per sequence. First, we focused on the three single-nucleotide polymorphisms (SNPs) located at codon positions 101, 115 and 119 of the VPg cistron which distinguish the five variants mixed to make the inoculum, i.e. variants G, N, K, GK and KN (see Fig S2A). For each sequence, we determined the corresponding variant among the eight possible (2^3) at the three nucleotide positions of interest. By doing so, we could estimate the frequencies of the five variants included in the inoculum and of the three possible variants carrying other SNP combinations at the three nucleotide positions of interest (i.e. the wild-type variant SON41p and variants GN and GKN). Then, for each of the 677 PVY populations, we determined the relative frequencies of these eight PVY variants. The additional three possible variants SON41p, GN and GKN could have appeared by mutation or recombination, either *in vivo* or *in vitro*, and should thus be surveyed. In a second step, we calculated the frequencies of all remaining nucleotide substitutions in each virus population by comparison with the sequence of the SON41p reference clone (equivalent to comparison with sequences of the G, N, K, GK and KN clones).

Sequence counts of the eight variants corresponding to the SNPs present in the initial inoculum We assigned each sequence to one of the eight potential PVY variants defined by the three SNPs of interest (variants G, N, K, GK, KN, SON41p, GN or GKN, see Fig S2A). Sequence counts are available in Table S2.

The sum of the frequencies of the three variants SON41p, GN and GKN that had not been included into the inoculum remained below 5% in all virus populations analyzed (Table 1). Additionally, 93.35% of samples showed a sum of frequencies below 2%, and 99.41% were below 3%. In the inoculum, the sum of the mean frequencies of these three variants was 2.32% (SON41p: 1.13%, GN: 0.50%, GKN: 0.69%), while it was 1.03% in plant samples (SON41p:

0.14%, GN: 0.11%, GKN: 0.78%). Given the low frequencies recorded, we cannot tell if those sequences indeed correspond to variants present in the virus population, or if they are artifacts due to errors during RT-PCR or sequencing. Negroni and Buc reported estimates of the recombination frequency by reverse transcriptases between $4 \cdot 10^{-5}$ and $2.4 \cdot 10^{-4}$ per nucleotide [6]. Our experimental framework involved one reverse transcription (RT) step and a maximum distance of 55 nucleotides between SNPs (corresponding to the G and N substitutions), yielding a recombination probability of 0.2 to 1.3%. We recorded 76.66 to 99.41% of samples with frequencies of variant SON41p below 0.2 to 1.3%, respectively. For the same variant frequencies (0.2% and 1.3%), the sample frequencies reached 79.32 to 99.85% for variant GN and 13.15 to 84.64% for variant GKN. Additionally, Potapov and Ong [7] estimated a polymerase chain reaction (PCR)-mediated recombination rate by *Taq* polymerase of $1.1 \cdot 10^{-4}$ per nucleotide and per doubling. For one doubling cycle, and considering again the most distant SNPs (55 nucleotides), the probability for one recombination event would be of 0.6%. We performed 47 PCR cycles, 35 to amplify the raw PCR products plus 12 to introduce the indices that identify each sequenced virus population. The expected recombination probability would be of 24.64%. Hence there is a high probability of artifactual presence of recombinants. This figure corresponds to the recombination probability between two types of DNA molecules present in equal proportion, which is not the case in our experiment comprising five different kinds of DNA molecules with highly variable frequencies. Moreover, recombinant DNA molecules generated *in vitro* may themselves recombine with other DNA molecules in the following PCR cycles. Consequently, although we expect high probabilities of *in vitro* recombination, these probabilities should be lower than the 24.64% estimated. In addition, given that these three variants could also have been generated by mutation of some of the five initial variants, and given the mutation error rate of RT, PCR or Illumina sequencing (a maximum estimated value of 0.86%; see below), it is not possible to distinguish these three PVY variants from the error background of laboratory enzymes. Additional arguments come from the comparison between the frequencies of the SON41p, GN and GKN variants in the inoculum and in infected plants. Concerning the inoculum population, there was no opportunity for *in vivo* recombination between the G, N, K, GK and KN variants, since these five variants were multiplied separately in different plants before being mixed to make the inoculum. Consequently, the SON41p, GN and GKN variants detected in the inoculum have been generated either by *in vitro* recombination or by mutation (*in vivo* or *in vitro*). In contrast, *in vivo* recombination may have been involved in the generation of these variants in the infected plants of the experiment, in addition to the previous two mechanisms. The fact that the frequency of the three putative recombinants is

lower in infected plants than in the inoculum indicates that *in vivo* recombination was negligible. Furthermore, even if some of these recombinants had been truly generated *in vivo*, their decrease in frequency between the inoculum and infected plants indicates that those variants did not play a substantial role in virus dynamics.

Table 1: **Counts of the sum of the frequencies of the three variants SON41p, GN and GKN in virus populations. Only intervals where at least one case was observed are reported.**

Interval (%)	]0-1]	]1-2]	]2-3]	]3-4]	]4-5]	**Total**
Counts	352	280	41	3	1	677

Frequencies of *de novo* nucleotide substitutions The complementary analysis consisted in looking at substitutions at all remaining 96 nucleotide positions, and at the two possible remaining substitutions at the three SNP positions of interest, i.e. those that were absent in the inoculum (first part). We focused on the sum of the frequencies of potential *de novo* substitutions at each nucleotide position, in plant samples or in the inoculum (Table 2). In infected plants, 5.15% (3452/67023) of nucleotide positions did not show any *de novo* substitution (i.e. frequency of 0%). Additionally, 99.91% of nucleotide positions showed a frequency of *de novo* substitutions below 1%, and 99.99% showed a frequency below 5%.

Table 2: **Counts in various intervals of the sum of the frequencies of all potential *de novo* substitutions at each nucleotide position, in plant samples and in the inoculum (inoc.). Only intervals where at least one case was observed are reported.**

Interval (%)	0	]0-0.1]	]0.1-0.2]	]0.2-0.3]	]0.3-0.4]	]0.4-0.5]	]0.5-1]	]1-2]	]2-3]
Counts in plants	3452	14817	22413	15488	7212	2493	1091	39	6
Counts in inoc.	6	87	171	94	28	8	2	0	0

Interval (%)	]3-4]	]4-5]	]5-6]	]6-7]	]14-15]	]26-27]	]71-72]	**Total**
Counts in plants	2	3	2	2	1	1	1	67023
Counts in inoc.	0	0	0	0	0	0	0	396

Illumina MiSeq sequencing errors comprise mostly nucleotide mismatches and a much lower rate of insertions or deletions [9]. The mismatch error rate was estimated between 0.25 and 0.46% per nucleotide [4, 9]. Besides, the mismatch error rates of the RT and PCR steps should be added to the mismatch error rate of Illumina MiSeq. The error rate of *Avian myeloblastosis* virus reverse transcriptase (0.0027% per nucleotide [8]) appears negligible compared to those of Illumina MiSeq and of the *Taq* polymerase used for PCR (0.27 to $0.85 \cdot 10^{-4}$ error per base pair and per cycle) [3]. Over the 47 PCR cycles carried out in our experiments, this *Taq*

polymerase would yield 0.13 to 0.40% of chances of cumulating at least one error per base pair. Taking into account the maximum error rate of Illumina MiSeq sequencing (0.46%) and of PCR (0.40%), we expect a total maximum error rate of 0.86% per nucleotide in our experiment. The vast majority (99.87%) of the sum of frequencies of substitutions at each nucleotide position recorded fall below this estimated error rate. Hence, given this threshold, most substitutions recorded probably result from *in vitro* errors during RT-PCR or Illumina MiSeq sequencing rather than being real substitutions occurring during virus replication.

Globally, the sum of the frequencies of *de novo* substitutions per nucleotide position varied between 0 and 71.52% in plant samples, while they reached at most 0.63% in the inoculum. Hence, in some plants, mutations have appeared and largely spread in the virus population during infection. Nevertheless, those cases are rare as only seven observations show a sum of frequencies of substitutions at a nucleotide position above 5% (Table 2), and six observations with a frequency of a single substitution (not the sum) above 5%. Details about those six latter cases are reported in Table 3.

Interestingly, three out of the six cases showed the same substitution at codon position 120 (Table 3). Three amino-acid substitutions at this position (serine to cysteine, isoleucine or threonine) have already been reported as determining adaptation to the *pvr2*^{3}-mediated resistance in several studies [1, 2, 5] and unpublished data, but not the serine (S) to arginine (R) substitution observed here. It is therefore highly likely that the $S_{120}R$ substitution was also positively selected in the *pvr2*^{3} host environment. Let us also note that one *de novo* substitution was observed at one of the three SNP positions of interest, corresponding to codon position 115 but yielded a different amino acid (arginine) than the one characterizing variant K (lysine) or SON41p reference clone (threonine). Substitution to arginine at codon 115 was shown previously to confer PVY adaptation to the *pvr2*^{3} resistance [1]. Interestingly, among the DH lines concerned by the highest frequencies of *de novo* substitutions, lines 2256 and 2400 are part of the ones inducing strongest genetic drift on virus populations, and DH line 2264 induces intermediate genetic drift (Fig 5). Hence, it is possible that the new mutations observed benefited from population expansion after a bottleneck step during plant infection, and this founder effect could have allowed these mutations to reach rapidly high frequencies in the virus populations. As a precaution, those six plants were removed for the estimation of effective population sizes (N_e) and selection coefficients (s).

We kept all remaining plant samples, showing frequencies of substitution below 5% (Table 2) and only the five variants mixed in the inoculum (G, N, K, GK and KN) for N_e and s estimations. In a numerical experiment, we tested the impact of a sixth, not accounted for

Table 3: **The six most frequent *de novo* nucleotide substitutions observed (> 5%) in the PVY populations. Indicated are the doubled-haploid (DH) line concerned, day of sampling (days post-inoculation, dpi), plant number, codon position, codon and its reference (Ref.) from the original sequence, corresponding amino-acid (aa) and its reference, and mutation frequency in the sample.**

DH line	Dpi	Plant	Codon position	Codon (Ref.)	aa (Ref.)	%
2256	34	1	120	AGA (AGT)	R (S)	71.32
2400	27	3	95	GAG (GAT)	E (D)	25.94
2256	14	3	110	AAT (GAT)	N (D)	14.07
2400	27	2	115	AGG (ACG)	R (T)	6.20
2400	6	4	120	AGA (AGT)	R (S)	5.91
2264	10	6	120	AGA (AGT)	R (S)	5.08

variant starting at a frequency of 3% in the inoculum, being neutral, and still present at the last sampling date (Text S2). The mean frequency of this sixth variant at all sampling dates and in all plants was 7% (5% quantile = 1%, median = 4%, 95% quantile = 24%). Overall, the accuracy of N_e and s estimations was not significantly impacted by the presence of this sixth variant (see Table 2 in main text). Hence, we can also trust our biological estimations based on the frequencies of substitutions recorded and neglecting *de novo* substitutions or potential recombinants present on average at a frequency < 7%.

References

1. V. Ayme, S. Souche, C. Caranta, M. Jacquemond, J. Chadœuf, A. Palloix, and B. Moury. Different Mutations in the Genome-Linked Protein VPg of *Potato virus Y* Confer Virulence on the *pvr*2^3 Resistance in Pepper. *MPMI*, 19(5):557–563, 2006.

2. V. Ayme, J. Petit-Pierre, S. Souche, A. Palloix, and B. Moury. Molecular dissection of the potato virus Y VPg Virulence factor reveals complex adaptations to the *pvr2* resistance allelic series in pepper. *Journal of General Virology*, 88:1594–1601, 2007.

3. M. A. Bracho, A. Moya, and E. Barrio. Contribution of *Taq* polymerase-induced errors to the estimation of RNA virus diversity. *Journal of General Virology*, 79:2921–2928, 1998.

4. D. Laehnemann, A. Borkhardt, and A. C. McHardy. Denoising DNA deep sequencing data— high-throughput sequencing errors and their correction. *Briefings in Bioinformatics*, 17(1):154–179, 2016.

5. J. Montarry, J. Doumayrou, V. Simon, and B. Moury. Genetic background matters: a plant-virus gene-for-gene interaction is strongly influenced by genetic contexts. *Mol. Plant Pathol.*, 12(9):911–920, 2011.

6. M. Negroni and H. Buc. Mechanisms of retroviral recombination. *Annu. Rev. Genet.*, 35:275–302, 2001.

7. V. Potapov and J. L. Ong. Examining sources of error in PCR by single-molecule sequencing. *PLoS ONE*, 12(1):e0169774, 2017.

8. J. D. Roberts, B. D. Preston, L. A. Johnston, A. Soni, L. A. Loeb, and T. A. Kunkel. Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis in vitro. *Molecular and Cell Biology*, 9(2):469–476, 1989.

9. M. G. Ross, C. Russ, M. Costello, A. Hollinger, N. J. Lennon, R. Hegarty, C. Nusbaum, and D. B. Baffe. Characterizing and measuring bias in sequence data. *Genome Biology*, 14:R51, 2013.

Supplementary Information – Text S2: Numerical experiment

We performed several batches of simulations to assess the ability of the estimation method to infer effective population sizes and selection coefficients accurately. Two numerical experiments were performed.

Experiment 1. In experiment 1, we simulated datasets involving a population of 5 virus variants, initially in equimolar mixture and all detected by the high-throughput sequencing (HTS) method. A dataset was obtained with the following three steps: (1) random sampling of selection and genetic drift parameters, (2) generation of population demogenetic dynamics given these known parameters with a Wright-Fisher model including selection and genetic drift and (3) building of numerical datasets with a structure similar to our real-life experiment, by sampling the composition of several populations at various time-points.

Step 1: Random sampling of selection and genetic drift parameters. Parameters $\boldsymbol{\theta}_{true} = (\boldsymbol{r}_{true}, \boldsymbol{N}_{true})$ are independently drawn from dedicated distributions that encompass a large diversity of selection and genetic drift scenarios. The relative fitness of the 5 virus variants $\boldsymbol{r}_{true}$ is obtained by independently drawing 5 values in uniform distribution ($\sim$ Unif$[0.85, 1.15]$) and then dividing them by their mean in order to have mean$(\boldsymbol{r}_{true}) = 1$. The scenario of genetic drift is obtained as follows. Firstly, the effective population size in the inoculated organ (N_e^{IO}) is drawn in a log-uniform distribution ($\sim$ Log-unif$[10, 2000]$). This stage is lasting 6 generations. Secondly, the effective population size at the onset of system infection ($N_e^{S_1}$) is drawn in a log-uniform distribution ($\sim$ Log-unif$[10, 2000]$). This stage is lasting 5 generations. Then, 5 more effective population sizes are drawn, corresponding to systemic infection stages. In order to avoid unrealistic trajectories ([1]), we set that the ratio of population sizes between two consecutive stages could not exceed 10. In practice, we iteratively computed $\log10(N_e^{S_{i+1}}) = \max(\min(\log10(N_e^{S_i}) + \alpha_i, \log10(2000)), \log10(10))$, with α_i sampled uniformly between -1 and 1 and $1 \leq i \leq 5$. Each stage is lasting 5 generations, except the last one lasting 3 generations. We thus obtained a vector $\boldsymbol{N}_{true}$ lasting 34 generations as follows:

$$\boldsymbol{N}_{true} = (\underbrace{N_e^{IO}, N_e^{IO}, \ldots}_{\text{6 generations}}, \underbrace{N_e^{S_1}, N_e^{S_1}, \ldots}_{\text{5 generations}}, \underbrace{N_e^{S_2}, \ldots}_{\text{5 generations}}, \ldots, \underbrace{N_e^{S_6}, \ldots}_{\text{3 generations}}) \quad (1)$$

Step 2: Generation of population demogenetic dynamics. We generated 48 independent Wright-Fisher simulations (using equations (2) and (3) of the main text) corresponding to the dynamics of populations of the 5 virus variants in 48 different plants of the same plant genotype, given $\boldsymbol{\theta}_{true} = (\boldsymbol{r}_{true}, \boldsymbol{N}_{true})$ and $\boldsymbol{\lambda}^{inoc} = (0.2, 0.2, 0.2, 0.2, 0.2)$.

Step 3: Building numerical datasets. We then carried out virtual observations for eight individual plants on each of the measurement dates corresponding to those used in the biological experiment, $\boldsymbol{T}^{obs}$ = (6, 10, 14, 20, 27, 34) days post-inoculation (dpi). We accounted for the HTS process, by sampling variant frequencies from multinomial distributions of size 3000 and with frequencies from Wright-Fisher simulations (in the same way as for step 2 of the ABC algorithm described in the main text). Importantly, in experiment 1, HTS analysis provides samples of the true frequencies of virus variants in the simulated Wright-Fisher populations. Finally, the dataset generated was accepted according to the following criteria, considered to be satisfied for all real-life experiment datasets (with the exception of one plant in 720). At each measurement date, at least two variants had to be present at a minimum frequency of 1% each in at least 50% of the populations. In addition, at least two variants had to be present in all populations at a minimum frequency of 1% each at the first measurement date (6 dpi). This criterion is hence largely permissive regarding the diversity of virus populations retained.

Step 4: Estimation of parameters. Using the three previous steps, we generated 750 datasets under as many selection and genetic drift regimes defined by the corresponding 750 values of $\boldsymbol{\theta}_{true}$. In order to assess the ability of the estimation method to infer effective population sizes and selection coefficients accurately, we estimated for each dataset $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{r}}, \hat{\boldsymbol{\eta}_e})$ using the more general model $\mathfrak{M}_4$ (*i.e.* with $\boldsymbol{\eta}_e = \left(\eta_e^{IO}, \eta_e^{S_1}, \eta_e^{S_2}, \eta_e^{S_3}\right)$). We assessed the accuracy of the estimates by comparing directly the estimated values of intrinsic rates of increase $\hat{\boldsymbol{r}}$ with their true values $\boldsymbol{r}_{true}$. For the effective population sizes, we compared the true harmonic mean of effective population sizes assessed from $\boldsymbol{N}_{true}$ at each measurement date $\boldsymbol{T}^{obs}$ and the harmonic mean of effective population sizes assessed from the piecewise function $N_e(t)$ (equation (1) of the main text) parameterized by $\hat{\boldsymbol{\eta}_e}$.

Experiment 2. In experiment 2, we tested the sensitivity of the estimation method to the presence of a sixth undetected virus variant. This sixth variant was selectively neutral (its selection coefficient is null), present in the inoculum at a frequency of 3% and still present at

the last sampling date (34 dpi) in all plants analyzed at frequencies ranging from 1% to 6%. It impacts the dynamics of the 5 variants of interest in all plants but was not detected, meaning that variant frequencies measured by HTS are noisy with respect to their true values. In all, 350 simulated datasets were analyzed for this second test. The mean relative change between the true frequencies of the 5 variants of interest in the simulated population and their measured frequencies by HTS is 0.08 (5% quantile = 0.01, median = 0.05, 95% quantile = 0.29).

In practice, a dataset in experiment 2 was obtained with the previous three steps modified as follows. In step 1 (Random sampling of selection and genetic drift parameters), the only difference is to add a sixth relative fitness equal to 1 to the vector $\boldsymbol{r}_{true}$. In step 2 (Generation of population demogenetic dynamics), as many as necessary independent Wright-Fisher simulations given $\boldsymbol{\theta}_{true} = (\boldsymbol{r}_{true}, \boldsymbol{N}_{true})$ and $\boldsymbol{\lambda}^{inoc} = (0.194, 0.194, 0.194, 0.194, 0.194, 0.03)$ are realized until having 48 independent simulations where the sixth variant is still present at the last sampling date (34 dpi) at frequencies ranging from 1% to 6%. Then, the dynamics of the sixth variant is erased from the 48 independent simulations retained. Step 3 (Building numerical datasets) is the same as previously. However, due to the deletion of the dynamics of the sixth variant, the frequencies of the 5 virus variants of interest used to mimic HTS though multinomial sampling are no more the true frequencies of the variants in the virus population but noisy values. Step 4 is also the same as in experiment 1. In particular, the inference was realized assuming that the inoculum was an equimolar mixture of the 5 variants of interest.

References

1. Boitard S, Rodríguez W, Jay F, Mona S, Austerlitz F. Inferring Population Size History from Large Samples of Genome-Wide Molecular Data - An Approximate Bayesian Computation Approach. PLOS Genetics. 2016;e1005877.

Supplementary Information – Text S3: Heritability of the intensity of selection and genetic drift exerted by plants on virus populations

Heritabilities were estimated by creating two dataset replicates of 24 randomly chosen plants for each DH line, each one including 4 plants at each sampling date. This random split was repeated 8 times in order to obtain 8 heritability estimates per variable of interest. Heritabilities were assessed for (i) the intrinsic rates of increase r_i of each virus variant i, (ii) the virus effective population size in the inoculated organ η_e^{IO} and (iii) the virus effective population size during systemic infection η_e^S. With 24 plants in each dataset, a simpler $N_e(t_g)$ piecewise function with only two parameters was used: $N_e(t) = \eta_e^{IO}$ when $t \in [1, 6]$) and (ii) $N_e(t) = \eta_e^S$ when $t \in [7, 34]$. The table below provides the mean heritabilities over the 8 repetitions of r_i (averaged over the 5 virus variants), η_e^{IO} and η_e^S. Additionally, the last column provides the range of variation of the heritabilities over the 8 repetitions.

Table 1: **Heritability estimations.**

Variable	**Mean heritability**	**Range of variation**
r_i	0.94	[0.76; 0.98]
η_e^{IO}	0.64	[0.36; 0.89]
η_e^S	0.63	[0.47; 0.85]

Supplementary Information:

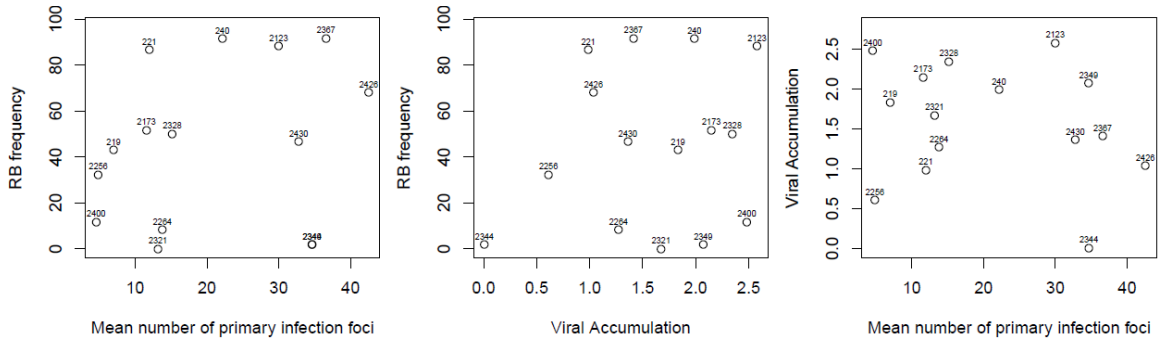


Fig S1. Resistance-breakdown (RB) frequency, viral accumulation and mean number of primary infection foci for the 15 DH lines studied. Pepper genotypes are represented as points, with their nomenclature (DH line number) given above each point. We estimated the mean number of primary infection foci for the 15 DH lines with the *Potato virus Y* (PVY, genus Potyvirus) variant K, carrying a green fluorescent marker (green fluorescent protein, GFP). The resistance-breakdown (RB) frequency and the relative viral accumulation were estimated by Quenouille et al. [12]. The RB frequency corresponds to the percentage of infected plants when inoculated with an avirulent variant regarding the allele of resistance *pvr2*^{3}, carried by all DH lines. The relative viral accumulation, or relative viral concentration, was measured by double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA).

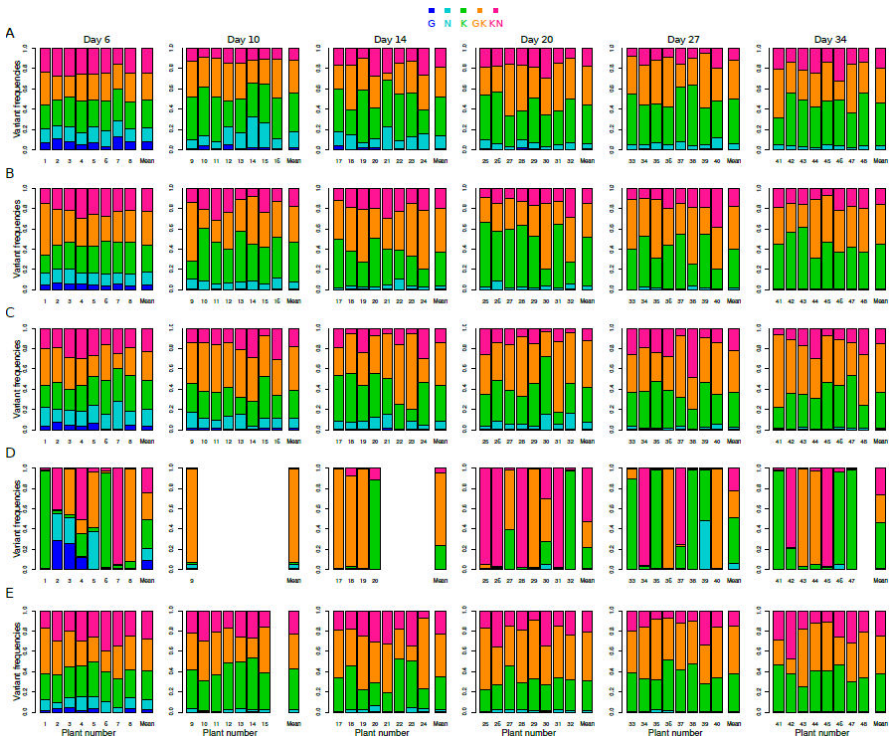


Fig S2. Five datasets obtained by high-throughput sequencing in the biological experiment. Each line of bar plots represents the dynamics of virus variants in a single DH line over time: (A) 221, (B) 2123, (C) 2173, (D) 2256 and (E) 2264. Within each bar plot, the frequencies of the five variants (see top of the figure for the color code) in each infected plant sample are represented by single bars (labeled from 1 to 48). The missing bars correspond to plant samples for which no viruses were detected. The last bar indicates the mean viral composition in the infected plants. Each individual bar plot corresponds to a single sampling date, indicated at the top of each column of barplots.

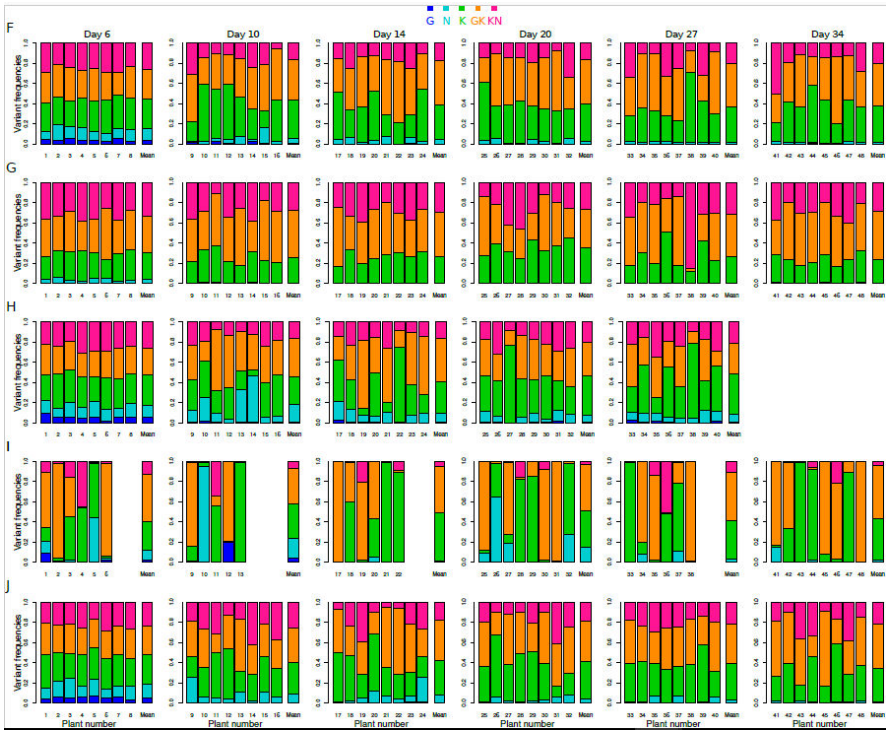


Fig S3. Five datasets obtained by high-throughput sequencing in the biological experiment. Each line of bar plots represents the dynamics of virus variants in a single DH line over time: (F) 2328, (G) 2349, (H) 2367, (I) 2400 and (J) 2426. Within each bar plot, the frequencies of the five variants (see top of the figure for the color code) in each infected plant sample are represented by single bars (labeled from 1 to 48). The missing bars correspond to plant samples for which no viruses were detected. The last bar indicates the mean viral composition in the infected plants. Each individual bar plot corresponds to a single sampling date, indicated at the top of each column of barplots.

Table S1. Tag sequences used to distinguish each plant sample after pooling and MiSeq Illumina high-throughput sequencing. The forward (Fwd.) primer sequence was the same for all amplifications and was bound to the sequence tag, just after it. Its binding site corresponds to positions 5971 to 5990 of PVY isolate SON41p (accession number AJ439544). The binding site of the reverse (Rev.) primer sequence corresponds to positions 6095 to 6114 of PVY isolate SON41p. RT-PCR amplifications were done according to the following profile: 1h at 42°C, 10 min at 95°C, 35 times the following sequence (45s at 95°C, 30s at 50°C and 20s at 72°C) and finally 10 min at 72°C.

	Name	Polarity	Sequence for binding libraries	Tag sequence	Primer sequence
Fwd.	Tag 1	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	ACGAGTGCGT	AAGAGAATGTCTATGCTGAC-3'
	Tag 2	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	CATAGTAGTG	AAGAGAATGTCTATGCTGAC-3'
	Tag 3	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	GGTCTAGTAC	AAGAGAATGTCTATGCTGAC-3'
	Tag 4	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	TCATACGCGT	AAGAGAATGTCTATGCTGAC-3'
	Tag 5	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	CTACGCTCTA	AAGAGAATGTCTATGCTGAC-3'
	Tag 6	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	ATCGATAGAC	AAGAGAATGTCTATGCTGAC-3'
	Tag 7	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	GAGGCTCTAC	AAGAGAATGTCTATGCTGAC-3'
	Tag 8	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	TGCTGATATC	AAGAGAATGTCTATGCTGAC-3'
Rev.	REV	-	5'-GGAGTTCAGACGTGTGCTCTTCCGATCT		CAGACCAATCTTTCCTGAAG-3'

Table S3. Estimations of the relative intrinsic rates of increase of the virus variants for the 15 plant genotypes. Virus variants are indexed as follow: (i) r_1 virus variant G, (ii) r_2 virus variant N, (iii) r_3 virus variant K, (iv) r_4 virus variant GK, (v) r_5 virus variant KN. The 90% confidence intervals are calculated as $\hat{r}_i \pm 1.645.\hat{\sigma}_i$.

HD line	r_1 (q-5%)	r_1 (mean)	r_1 (q-95%)	r_2 (q-5%)	r_2 (mean)	r_2 (q-95%)	r_3 (q-5%)	r_3 (mean)	r_3 (q-95%)	r_4 (q-5%)	r_4 (mean)	r_4 (q-95%)	r_5 (q-5%)	r_5 (mean)	r_5 (q-95%)
2123	0.7386	0.7425	0.7463	0.9630	0.9645	0.966	1.1113	1.1124	1.1135	1.1145	1.1156	1.1167	1.0636	1.0647	1.0659
2173	0.7529	0.7568	0.7608	0.9943	0.9956	0.9969	1.0872	1.0883	1.0894	1.1072	1.1083	1.1094	1.0496	1.0507	1.0518
219	0.9019	0.9032	0.9045	1.0358	1.0363	1.0368	1.0357	1.0362	1.0367	1.049	1.0495	1.0500	0.9738	0.9745	0.9752
221	0.8306	0.8329	0.8351	0.9945	0.9953	0.9962	1.0707	1.0713	1.072	1.0728	1.0735	1.0742	1.026	1.0267	1.0274
2256	0.8238	0.8276	0.8314	0.9508	0.9524	0.9539	1.0832	1.0843	1.0854	1.072	1.0731	1.0742	1.0613	1.0624	1.0635
2264	0.6868	0.6922	0.6977	0.9413	0.9434	0.9455	1.1220	1.1235	1.1251	1.142	1.1436	1.1451	1.0955	1.097	1.0985
2321	0.4171	0.4350	0.4529	0.8756	0.8814	0.8873	1.2045	1.2095	1.2144	1.2367	1.2418	1.2469	1.227	1.232	1.2371
2328	0.7484	0.7526	0.7567	0.9657	0.9673	0.9689	1.0963	1.0975	1.0986	1.1139	1.1151	1.1163	1.0662	1.0674	1.0686
2344	0.5155	0.5278	0.5400	0.9818	0.9850	0.9883	1.148	1.1513	1.1545	1.1919	1.1952	1.1985	1.1373	1.1405	1.1437
2349	0.4241	0.4403	0.4564	0.8759	0.8809	0.8858	1.2063	1.2107	1.2152	1.2463	1.2509	1.2554	1.2126	1.2171	1.2215
2367	0.8088	0.8115	0.8141	0.9989	1.0005	1.0011	1.0698	1.0706	1.0715	1.0828	1.0837	1.0845	1.0331	1.034	1.0349
240	0.8471	0.8493	0.8516	0.9966	0.9976	0.9986	1.0565	1.0573	1.0581	1.0752	1.0759	1.0767	1.0188	1.0196	1.0205
2400	0.7628	0.7673	0.7719	1.0091	1.0105	1.0119	1.1011	1.1024	1.1036	1.1205	1.1218	1.1231	0.9964	0.9977	0.9991
2426	0.7460	0.7495	0.7531	0.9832	0.9844	0.9857	1.0913	1.0923	1.0933	1.1079	1.109	1.11	1.0635	1.0646	1.0656
2430	0.7541	0.7580	0.7618	0.9620	0.9635	0.9651	1.105	1.1062	1.1073	1.1028	1.104	1.1051	1.067	1.0681	1.0692

Table S4. Model selection and estimations of the effective population sizes for the 15 plant genotypes. The posterior probabilities of the four models considered for the piecewise function describing the temporal variation of the effective population sizes during the time course of the experiment (models $\mathfrak{M}_1$, $\mathfrak{M}_2$, $\mathfrak{M}_3$ and $\mathfrak{M}_4$) are first indicated. The bold value corresponds to the model that is best supported by the data. The next columns indicate the estimation of the effective population sizes of the model selected and the extent of a 90% credibility intervals.

	Posterior probabilities				days 1 to 6			days 7 to 10			days 11 to 14			days 15 to 34		
HD line	model 1	model 2	model 3	model 4	Ne (q-5%)	Ne (mean)	Ne (q-95%)	Ne (q-5%)	Ne (mean)	Ne (q-95%)	Ne (q-5%)	Ne (mean)	Ne (q-95%)	Ne (q-5%)	Ne (mean)	Ne (q-95%)
2123	**0.44**	0.17	0.22	0.16	216	349	543	216	349	543	216	349	543	216	349	543
2173	0.02	0	0.02	**0.96**	96	248	620	98	459	1951	55	225	1432	403	1095	2310
219	0.14	0.26	**0.32**	0.28	10	13	19	11	42	505	31	276	1909	31	276	1909
221	0.24	0.05	**0.43**	0.28	237	552	1213	188	461	1286	616	1368	2366	616	1368	2366
2256	0.07	0.03	**0.52**	0.38	10	13	21	11	43	440	11	61	1146	11	61	1146
2264	**0.69**	0.07	0.08	0.17	252	393	597	252	393	597	252	393	597	252	393	597
2321	0.00	0.21	0.04	**0.74**	776	1329	2135	11	20	44	17	148	1687	64	400	2025
2328	0.00	0.14	0.10	**0.76**	430	866	1679	60	135	315	135	590	2139	181	534	1711
2344	0.19	0.29	0.11	**0.41**	107	421	1440	51	201	1283	72	412	2009	278	908	2269
2349	**0.35**	0.14	0.30	0.21	276	462	749	276	462	749	276	462	749	276	462	749
2367	0.00	0.35	0.31	**0.35**	526	883	1451	43	80	156	92	441	1989	254	858	2204
240	0.00	0.21	**0.52**	0.27	978	1515	2260	127	213	353	307	959	2275	307	959	2275
2400	**0.84**	0.07	0.05	0.04	11	18	31	11	18	31	11	18	31	11	18	31
2426	0.01	**0.50**	0.24	0.26	383	805	1593	122	235	448	122	235	448	122	235	448
2430	0.07	0.23	**0.47**	0.23	61	1130	1971	165	308	601	420	1077	2269	420	1077	2269

Adaptation des populations virales aux résistances variétales et exploitation des ressources génétiques des plantes pour contrôler cette adaptation

L'utilisation de variétés de plantes porteuses de gènes majeurs de résistance a longtemps été une solution privilégiée pour lutter contre les maladies des plantes. Cependant, la capacité des agents pathogènes à s'adapter à ces variétés après seulement quelques années de culture rend nécessaire la recherche de résistances à la fois efficaces et durables. Les objectifs de cette thèse étaient (i) d'identifier chez la plante des régions génomiques contraignant l'évolution des agents pathogènes en induisant des effets de dérive génétique et (ii) d'étudier l'impact des forces évolutives induites par la plante sur la capacité d'adaptation des pathogènes aux résistances variétales, l'ambition étant par la suite d'employer au mieux ces forces pour limiter l'évolution des pathogènes. Le pathosystème piment (*Capsicum annuum*) – PVY (*Potato virus Y*) a été principalement utilisé pour mener ces travaux de recherche. Afin de répondre au premier objectif, une cartographie de QTL (quantitative trait loci) sur une population biparentale de piment et une étude de génétique d'association sur une core-collection de piments ont été réalisées. Ces deux approches ont permis de mettre en évidence des régions génomiques sur les chromosomes 6, 7 et 12 impliquées dans le contrôle de la taille efficace des populations virales lors de l'étape d'inoculation du virus dans la plante. Certains de ces QTL ont montré une action vis-à-vis du PVY et du CMV (*Cucumber mosaic virus*) tandis que d'autres se sont révélés être spécifiques d'une seule espèce virale. Par ailleurs, le QTL détecté sur le chromosome 6 co-localise avec un QTL précédemment identifié comme contrôlant l'accumulation virale et interagissant avec un QTL affectant la fréquence de contournement d'un gène majeur de résistance. Pour répondre au second objectif, une analyse de la corrélation entre l'intensité des forces évolutives induites par la plante et une estimation expérimentale de la durabilité du gène majeur a été réalisée. De l'évolution expérimentale de populations de PVY sur des plantes induisant des effets de dérive génétique, de sélection et d'accumulation virale contrastés a également été effectuée. Ces deux études ont démontré qu'une plante induisant une forte dérive génétique associée à une réduction de l'accumulation virale permettait de contraindre l'évolution des populations virales, voire d'entraîner leur extinction. Ces résultats ouvrent de nouvelles perspectives pour le déploiement de déterminants génétiques de la plante qui influenceraient directement le potentiel évolutif du pathogène et permettraient de préserver la durabilité des gènes majeurs de résistance.

Mots clés : dérive génétique - quantitative trait locus (QTL) - durabilité des résistances - évolution expérimentale - *Capsicum annuum* - *Potato virus Y* - *Cucumber mosaic virus*

Adaptation of viral populations to plant resistance and exploitation of plant genetic resources to control this adaptation

Plants carrying major resistance genes have been widely used to fight against diseases. However, the pathogens ability to overcome the resistance after a few years of usage requires the search for efficient and durable resistances. The objectives of this thesis were (i) to identify plant genomic regions limiting pathogen evolution by inducing genetic drift effects and (ii) to study the impact of the evolutionary forces imposed by the plant on the pathogen ability to adapt to resistance, the goal being to further use these forces to limit pathogen evolution. The pepper (*Capsicum annuum*) – PVY (*Potato virus Y*) pathosystem has been mainly used to conduct these researches. Regarding the first objective, quantitative trait loci (QTL) were mapped on a biparental pepper population and through genome-wide association on a pepper core-collection. These approaches have allowed the detection of genomic regions on chromosomes 6, 7 and 12 controlling viral effective population size during the inoculation step. Some of these QTLs were common to PVY and CMV (*Cucumber mosaic virus*) while other were virus-specific. Moreover, the QTL detected on chromosome 6 colocalizes with a previously identified QTL controlling PVY accumulation and interacting with a QTL affecting the breakdown frequency of a major resistance gene. Regarding the second objective, a correlation analysis between the evolutionary forces imposed by the plant and an experimental estimation of the durability of a major resistance gene has been done. Experimental evolution of PVY populations on plants contrasted for the levels of genetic drift, selection and virus accumulation they imposed has also been performed. Both studies demonstrated that a plant inducing a strong genetic drift combined to a reduction in virus accumulation limits virus evolution and could even lead to the extinction of the virus population. These results open new perspectives to deploy plant genetic factors directly controlling pathogen evolutionary potential and could help to preserve the durability of major resistance genes.

Keywords: genetic drift - quantitative trait locus (QTL) - resistance durability - experimental evolution - *Capsicum annuum* - *Potato virus Y* - *Cucumber mosaic virus*