Estimation of genetic drift and selection from next generation sequencing time-sampled data

Elsa Rousseau, Frédéric Fabre, Jérôme Coville, Ludovic Mailleret, Alain Palloix, Rachid R. Senoussi, Frédéric Grognard, Benoît Moury

To cite this version:

Elsa Rousseau, Frédéric Fabre, Jérôme Coville, Ludovic Mailleret, Alain Palloix, et al.. Estimation of genetic drift and selection from next generation sequencing time-sampled data. Modelling biological evolution 2015: linking mathematical theories with empirical realities, Apr 2015, Leicester, United Kingdom. , 2015. hal-02738798

HAL Id: hal-02738798

https://hal.inrae.fr/hal-02738798v1

Submitted on 2 Jun 2020

Estimation of Genetic Drift and Selection from Next Generation Sequencing Time-sampled Data

Elsa Rousseau[1,2,3], Frédéric Fabre[4*], Jérôme Coville[5], Ludovic Mailleret[1,2], Alain Palloix[6], Rachid Senoussi[5], Frédéric Grognard[1] & Benoît Moury[3]

[1] INRIA, Biocore Team, Sophia Antipolis, France, [2] INRA, UMR ISA, Sophia Antipolis, France, [3] INRA, UR Pathologie Végétale, Montfavet, France, [4] INRA, UMR Santé et Agroécologie du Vignoble, Villenave d'Ornon cedex, France, [5] INRA, UR BioSp, Avignon, France, [6] INRA, UR Génétique et Amélioration des Fruits et Légumes, Montfavet, France, *frederic.fabre@bordeaux.inra.fr

❶ Introduction & objectives

♣ Uncovering how natural selection and genetic drift shape the evolutionary dynamics of pathogen populations within their hosts is an important issue (*e.g*. virus emergence). Methods exist to estimate the intensity of drift when neutral genetic markers are available. However few methods allow to jointly estimate the intensities of genetic drift and selection without neutral markers (Foll et al., 2014, *Plos Genet.*).

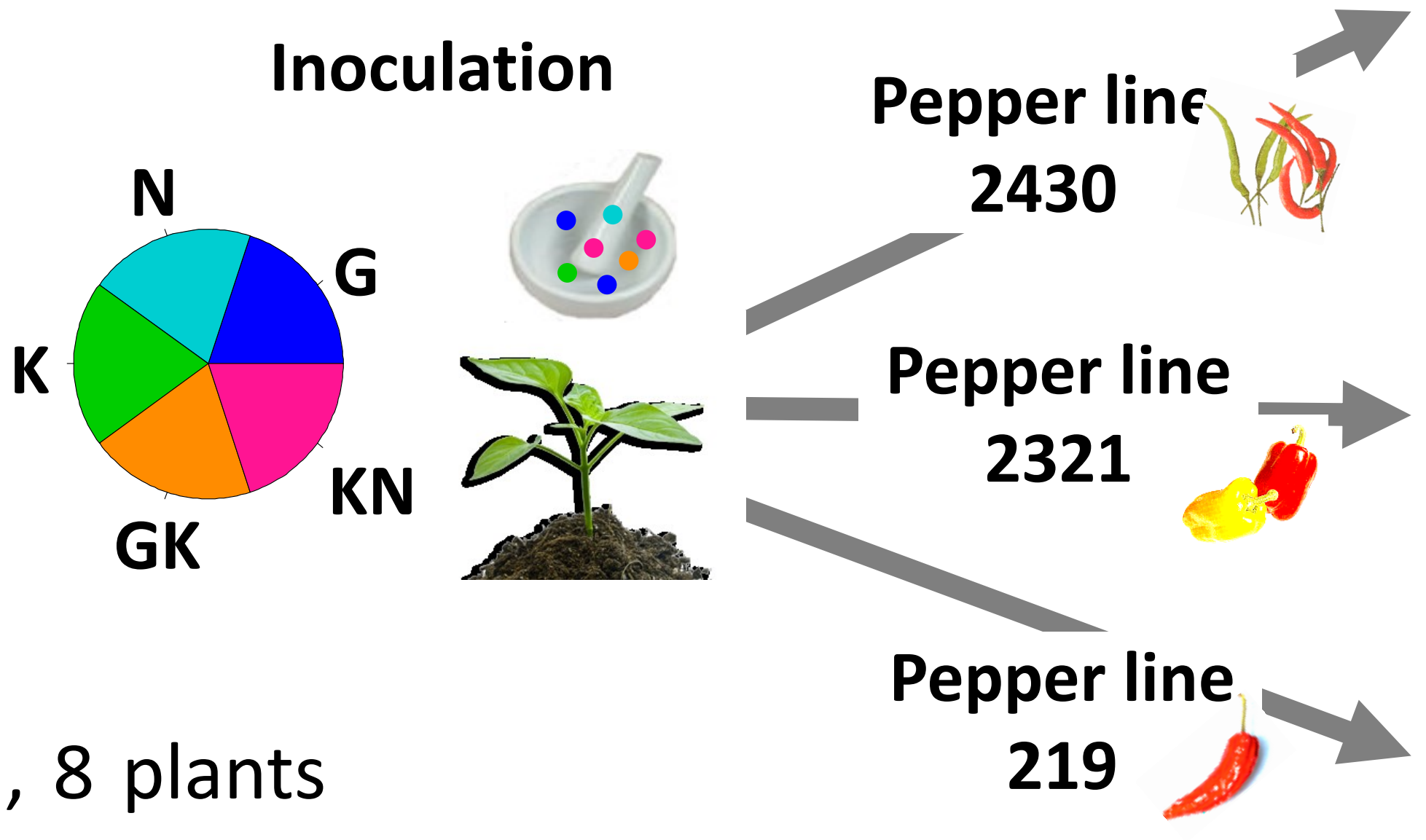
► **Aim 1**: ***Developing a method to jointly estimate effective population sizes (Ne), a measure of genetic drift, and selection coefficients (s) on time-sampled data representing temporal changes in allele frequencies.***

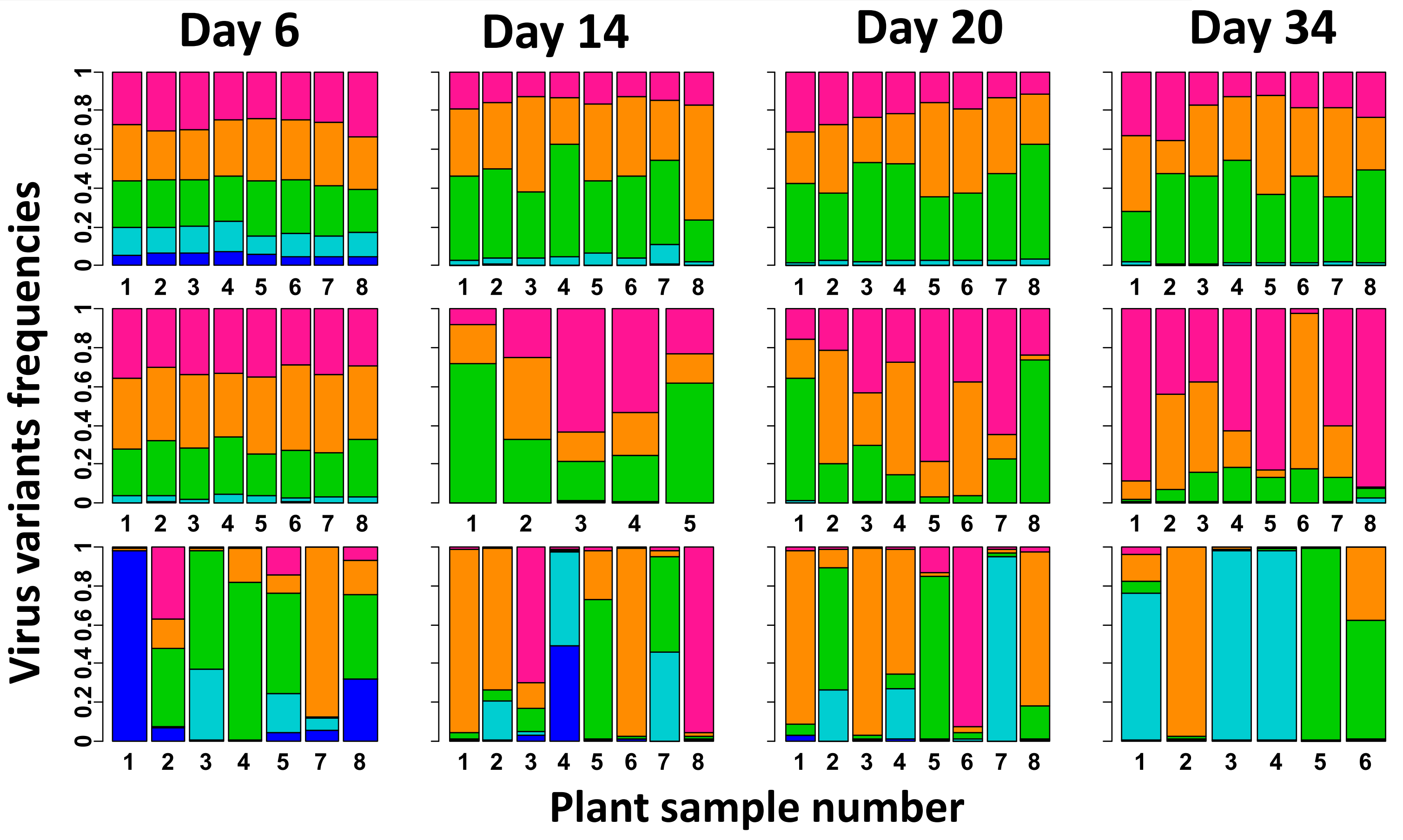
► **Aim 2**: ***Identifying genetic background of host plants increasing genetic drift acting on virus populations…in order to delay the emergence of resistance-breaking virus.***

❷ Experiments: Demo-genetic dynamics of viruses in their host plant

♣ Inoculation of 15 pepper genotypes (carrying the same resistance gene *pvr2³* but distinct Quantitative Trait Loci combinations) with a mixture of 5 *Potato virus Y* variants .

Virus variant		Adaptation to *pvr2³*
Letters indicate the amino acid present at positions 101 (G), 119 (K) and 121 (N) of the VPg cistron	G	Weak
	N	Weak
	K	Strong
	GK	Strong
	KN	Strong





♣ At 6 dates (6, 10, 14, 20, 27 and 34 dpi), 8 plants per genotype were sampled and the frequency of each variant in each plant (n=6*8*16) determined with Miseq sequencing of the VPg cistron.

► ***Contrasted patterns of genetic drift and selection between hosts***

❸ A mechanistic-statistical model to jointly estimate *Ne* and *s*

♣ A model for the dynamics of the mean density of variant i $V_i(t)$

$$\frac{dV_i(t)}{dt} = r_i V_i \left[1 - \frac{1}{K}\left(V_i + \sum_{j \neq i} \frac{r_j}{r_i} V_i\right)\right] + \sum_{j=1}^{5} \mu_{ij}(V_j - V_i)$$

with r_i ,intrinsic rate of increase of variant *i*, $\mu_{i,j}$,mutation rate from variant *j* to variant *i* (fixed) and *K* the carrying capacity (fixed)

♣ ….coupled to a model for the variance of frequencies between hosts:

$$\boldsymbol{X}(p,d) \sim Dir\left(\boldsymbol{f}(d) = \frac{(V_1(d),\ldots,V_5(d))}{K}, \alpha(d)\right)$$

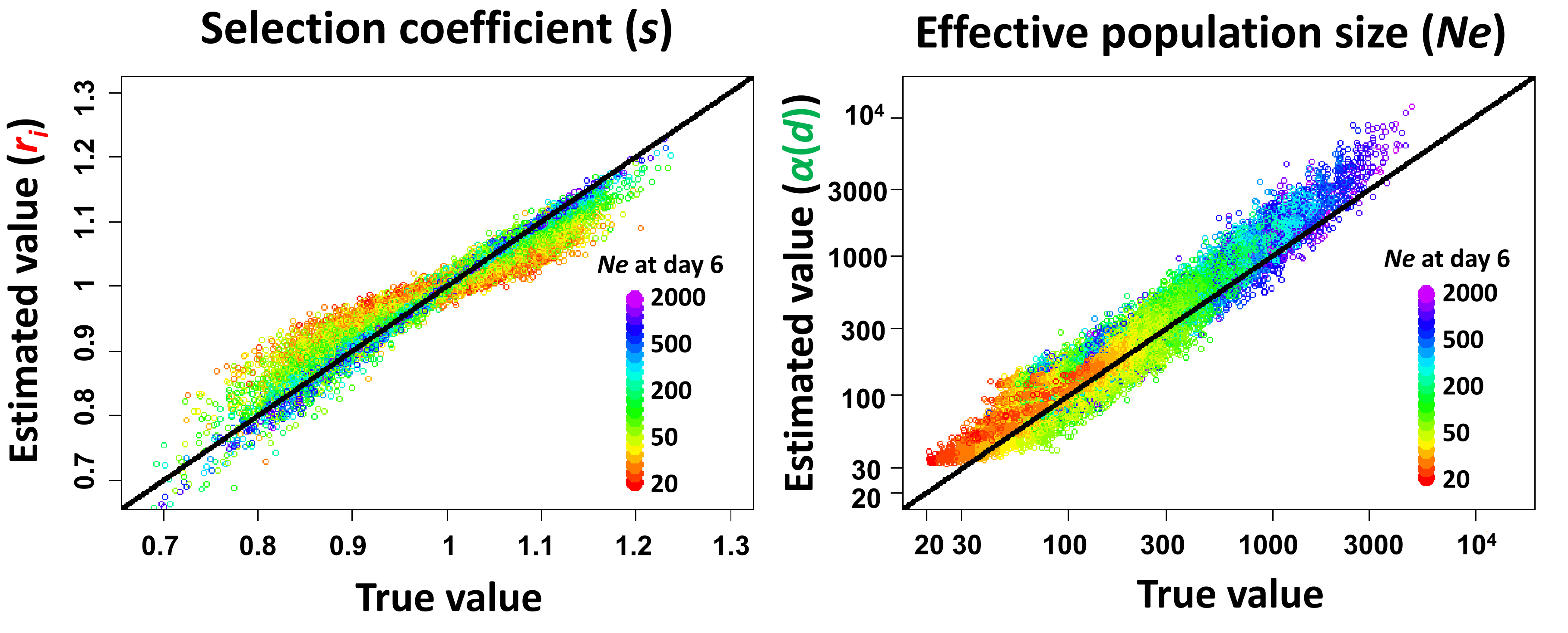
with $\boldsymbol{X}(p,d)$ *the observed frequencies* in plant *p* at time *d*, $\boldsymbol{f}(d)=(f_1(d),\ldots,f_5(d))$ the mean frequencies of virus variants at time *d* *and* $\alpha(d)$ a scale parameter at time *d.*

♣ $\theta=(r_1,\ldots,r_5, \alpha(6),\ldots \alpha(34))$ can be estimated by maximum likelihood given that mean(r_i)=1 (Fabre et al., 2012, *Plos Path*.).

► ***Mean trajectories of virus variants contain the evidence of selection and between-hosts variances contain the evidence of genetic drift.***

♣ Simulation of datasets with varying *s* and *Ne* using a Wright-Fisher model with selection, mutation and genetic drift in a haploid context (FFPopSim ; Zanin & Neher, 2012, *Bioinformatics*)

♣ For each simualted dataset, the vector θ *of* model parameters is estimated.

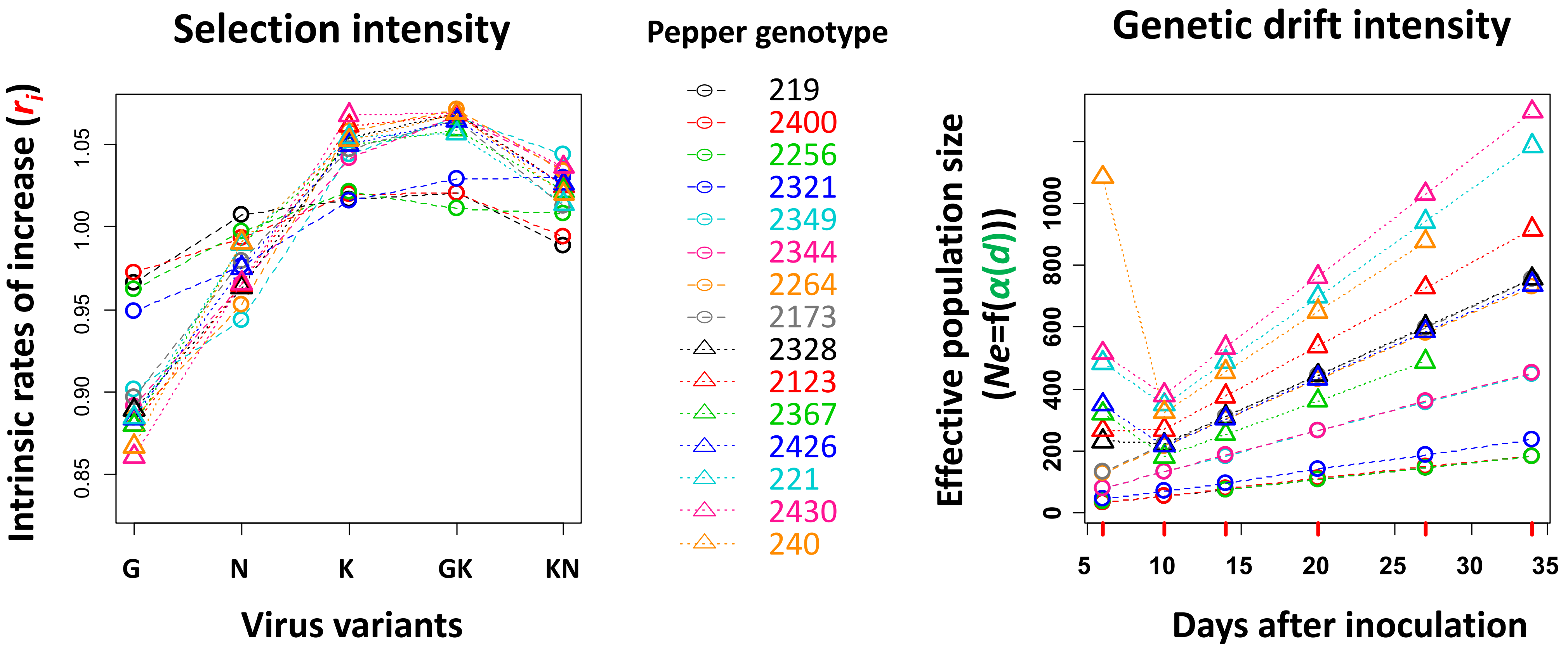


► ***Validation of a method to jointly estimate both s and Ne without neutral loci using NGS method and a time-sampling scheme of independent hosts.***

❹ Results: *Ne* is varying from 35 to 1200 depending on host genotype

► ***The fitness landscapes of the virus population have roughly the same shape whatever the pepper genotype is.***

► ***Notably, no situation where "a variant selected in one host is counter-selected in another host environment" is reported.***



► ***Large variability of Ne between pepper genotypes mainly determined by infection process within the inoculated leaf (days 1 to 6).***

► ***Some Quantitative Trait Loci are likely to control the Ne of virus populations in pepper plants.***

Modelling Biological Evolution 2015, Leicester (UK), April-28 – May 1, 2015