



HAL
open science

Host range evolution of potyviruses: A global phylogenetic analysis.

Benoît Moury, Cecile Desbiez

► **To cite this version:**

Benoît Moury, Cecile Desbiez. Host range evolution of potyviruses: A global phylogenetic analysis.. Viruses, 2020, 12 (1), pp.111. 10.3390/v12010111 . hal-02622450

HAL Id: hal-02622450

<https://hal.inrae.fr/hal-02622450>

Submitted on 26 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Article

Host Range Evolution of Potyviruses: A Global Phylogenetic Analysis

Benoît Moury * and Cécile Desbiez

INRAE, Pathologie Végétale, F-84140 Montfavet, France; cecile.desbiez@inra.fr

* Correspondence: benoit.moury@inra.fr; Tel.: +33-(0)4-3272-2816

Received: 12 December 2019; Accepted: 10 January 2020; Published: 16 January 2020



Abstract: Virus host range, i.e., the number and diversity of host species of viruses, is an important determinant of disease emergence and of the efficiency of disease control strategies. However, for plant viruses, little is known about the genetic or ecological factors involved in the evolution of host range. Using available genome sequences and host range data, we performed a phylogenetic analysis of host range evolution in the genus *Potyvirus*, a large group of plant RNA viruses that has undergone a radiative evolution circa 7000 years ago, contemporaneously with agriculture intensification in mid Holocene. Maximum likelihood inference based on a set of 59 potyviruses and 38 plant species showed frequent host range changes during potyvirus evolution, with 4.6 changes per plant species on average, including 3.1 host gains and 1.5 host loss. These changes were quite recent, 74% of them being inferred on the terminal branches of the potyvirus tree. The most striking result was the high frequency of correlated host gains occurring repeatedly in different branches of the potyvirus tree, which raises the question of the dependence of the molecular and/or ecological mechanisms involved in adaptation to different plant species.

Keywords: potyvirus; potyviridae; ancestral reconstruction; discrete character; host jump; host shift; host range expansion

1. Introduction

Host range of a parasite is defined as the set of organisms where the parasite can perform, at least partly, its infection cycle. The number and diversity of host species are frequently used to measure a parasites' host range. Host range is an important notion in parasitology, since it is linked to risks of disease emergence [1] and has important consequences on the implementation and efficiency of disease control strategies including epidemiological surveillance, quarantine measures, vaccination, and deployment of parasite-resistant crop varieties in agricultural landscapes [2–4].

In phytopathology, parasites' host range has been used for decades for taxonomic purposes, both at parasite inter- and intraspecies levels. Host range is determined both by factors intrinsic to the parasite that determine its fitness in different hosts and by extrinsic factors related to parasite ecology and epidemiology [5]. Overall, there is little knowledge on determinants of the host species range of plant parasites by contrast with determinants of infection at the plant intraspecific level that have been the subject of numerous studies, particularly about the ability of a parasite to infect plant genotypes with specific resistance genes [5]. For plant viruses, relatively few studies are available on host range genetic determinants or on host range changes that have occurred during virus evolutionary history [6–16].

We used a phylogenetic approach to unravel the evolution of the host species range in the genus *Potyvirus*. Given its large number of species (175 member species and 31 tentative species), this genus is favorable for this kind of approach [17]. Collectively, its members have a broad host range that encompass 57 plant families, mostly in the division Magnoliophyta (i.e., angiosperms or flower

plants) [18]. However, each potyvirus species usually has a narrow host range, with a few exceptions like *Bean yellow mosaic virus*, *Turnip mosaic virus*, and *Watermelon mosaic virus* (WMV) which infect hosts in at least 12 plant families each [18]. Phylogenetic analyses showed that the diversification of the genus *Potyvirus* is quite recent, since it began circa 7000 years ago, probably in the greater region around the Fertile Crescent of the Middle East, where agriculture emerged about 9500 years ago [19]. This suggests that potyviruses have undergone numerous host range changes in a relatively short amount of time and that crop plants have contributed substantially to the diversification of potyviruses.

Our objectives were to unravel the nature and frequency of host range changes that occurred during the evolution of potyviruses and the distribution of these changes among the potyvirus and plant diversity.

2. Materials and Methods

2.1. Collating Host Range Data

Host range data of potyvirus species were obtained from two bibliographic sources, which have established quite exhaustive lists of host and non-host plant species for each potyvirus species [18,20]. In these sources, non-hosts had been determined by laboratory inoculation experiments, whereas hosts corresponded either to naturally or experimentally infected species. Local-lesion hosts had been considered as hosts. In case of disagreement between the sources, i.e., when a given plant species was considered to be either host or non-host of a given potyvirus species, we considered it to be a host, since this discrepancy is likely due to the use of different plant genotypes and/or virus isolates in different studies.

2.2. Phylogenetic Analyses

For each potyvirus species for which we could obtain host range data, we checked whether full genome or near full genome sequences were available in databanks. Genome data were obtained for 57 potyvirus species, to which we added a partial genome of *Iris mild mosaic virus* (IMMV; accession JF320812; excluding most of the P1 cistron) and a complete genome of *Amaranthus leaf mottle virus* (AmLMV) obtained in our laboratory (accession number MN709786). Sequences of three members of the genus *Rymovirus*, the sister group of the genus *Potyvirus* in the family Potyviridae, were added as outgroups to root the phylogenetic tree and to help reconstruct host range changes. These were *Hordeum mosaic virus* (HoMV), *Agropyron mosaic virus* (AgMV), and *Ryegrass mosaic virus* (RgMV).

Nucleotide or amino acid sequences of potyviruses and rymoviruses were aligned using MAFFT (Multiple Alignment using Fast Fourier Transform) version 7 [21]. Then, the best substitution model was selected by maximum likelihood using the MEGA software version 7 [22]. Finally, phylogenetic analyses were conducted using the selected substitution model and the maximum likelihood algorithm implemented in MEGA version 7. One hundred bootstraps were performed to evaluate the reliability of the tree topology.

2.3. Determination of Ancestral Host/Non-Host States in Potyvirus Phylogenetic Tree

For each plant species, by combining the obtained virus phylogeny (topology and branch lengths) and the host and non-host data, we assigned to each internal branch i of the virus phylogenetic tree the probability P_i that the given plant species was a host of the corresponding potyvirus ancestor. Hence, the probability that the given plant species was non-host of that potyvirus ancestor is $(1 - P_i)$. For this, we compared the likelihood of two models using the 'ace' function [23] in the 'ape' package [24] of the software R version 3.5.3 [25]. The equal-rates (ER) model considers that the evolution from host to non-host and from non-host to host are equally probable, whereas the all-rates-different (ARD) model considers two different evolutionary probabilities, one for the host-to-non-host transition and the other one for the non-host-to-host transition. Since there are only two possible modalities among our data (host and non-host), the ER and ARD models differ by one parameter only and their likelihoods can

be compared using a Chi-squared test with one degree of freedom. Then, the obtained probabilities P_i associated to the different branches of the phylogenetic tree allowed identifying the branches corresponding to host range changes. When changes corresponded to internal branches in a clade containing host/non-host data for a single virus species and no data was available for the other viruses in the same clade, the change was attributed to the terminal branch corresponding to that virus.

To evaluate the respective impacts of (i) the total numbers of viruses for which a given plant species is host or non-host and (ii) their distribution in the potyvirus phylogenetic tree on the host or non-host state of that plant species for the most recent common ancestor (MRCA) of potyviruses, we performed permutation tests. For these, host range data were randomly permuted among the virus species before re-inferring the probability that the plant species was a host of the potyvirus MRCA. Ten thousand permutations were performed to obtain probability values.

The Monte Carlo method was used to test if there was a significantly low or high frequency of host gains, host losses, or total host changes in the potyvirus phylogenetic tree depending on the plant family or the potyvirus clade. The Monte Carlo method was also used to test if, compared to distributions expected at random, there was significantly lower or higher frequencies of (i) host change co-occurrences (i.e., change of the host status of several plant species inferred on the same branch of the potyvirus phylogenetic tree) and (ii) repeated host change co-occurrences, i.e., the same host change co-occurrence found on several branches of the potyvirus phylogenetic tree. In all cases, the probability distribution used for simulations was proportional to the branch lengths or to the sum of branch lengths when groups of plant or virus species were analyzed. When sums of branch lengths were taken into account, we excluded the branches corresponding to viruses for which no host range data was available.

3. Results

3.1. Building a Potyvirus and Rymovirus Phylogenetic Tree

To identify host range changes during the evolution of potyviruses, we first established the phylogenetic relationships between 62 virus species for which genome and host range data were available: 59 *Potyvirus* species and three *Rymovirus* species used as outgroups. Reliable alignments of the polyprotein amino acid sequences were obtained, except for the P1 region. In contrast, alignments were uncertain for several genome regions using nucleotide sequences. The best-suited substitution model, i.e., the LG (Le-Gascuel) model [26] with gamma-distributed rates among sites, invariant sites, and including amino acid frequencies, was used to infer virus phylogeny with the amino acid alignments (Figure 1). In spite of the use of a different set of potyvirus species, the tree topology was consistent with that obtained by Gibbs and Ohshima [19] and the same potyvirus groups were monophyletic in both studies (Figure 1).

3.2. Frequent Host Range Changes during Potyvirus Evolution

For each plant species, we used the topology and branch lengths of the phylogenetic tree together with the host/non-host data to infer the ancestral states in the virus tree. We showed that less than 2% of random simulations showed a probability above 0.65 that a given plant species was a host or a non-host for the potyvirus MRCA corresponding to branch number 64 (Figure 1) (i.e., $P_{64} > 0.65$ or $(1 - P_{64}) > 0.65$) (Appendix A). Hence, we considered that the host or non-host status of the potyvirus MRCA was reliable if $P_{64} > 0.65$ or $(1 - P_{64}) > 0.65$. According to this criterion, no reliable inference could be obtained for plant species in which the host/non-host status is known for less than 12 of the 62 viruses (data not shown). Among the 66 remaining plant species (Supplementary Table S1), a reliable inference of the state of potyvirus MRCA was obtained for 39 plant species (Table 1).

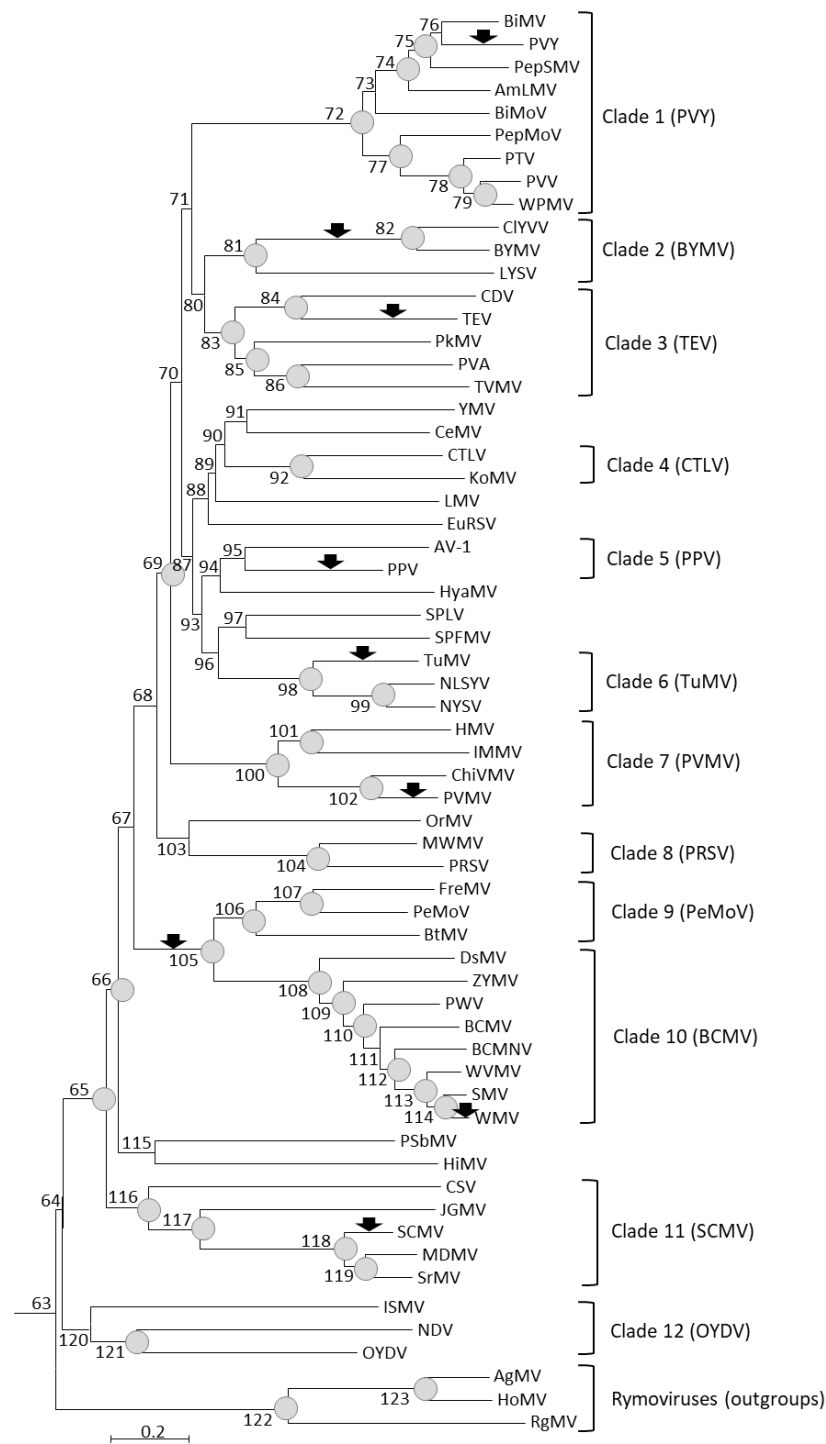


Figure 1. Phylogenetic tree of potyviruses and rymoviruses based on alignment of amino acid sequences of full-length genomes with MAFFT and the LG model with gamma-distributed rates among sites, invariant sites, and amino acid frequencies. Internal branches are numbered from 63 to 123. Terminal branches are named after the acronym of the corresponding virus species. Branches with bootstrap values above 0.7 are indicated by gray circles. Arrows indicate branches with a significantly higher number of host gains than expected by chance (Supplementary Table S3). Virus species corresponding to acronyms and GenBank accession numbers are as follows: *Agropyron mosaic virus* (AgMV; accession NC_005903), *Amaranthus leaf mottle virus* (AmLMV; MN709786), *Asparagus virus 1* (AV-1; NC_025821), *Bean common mosaic necrosis virus* (BCMNV; NC_004047), *Bean common mosaic virus* (BCMV; MK069988), *Bidens mottle virus* (BiMoV; NC_014325), *Bidens mosaic virus* (BiMV; NC_023014), *Beet mosaic virus* (BtMV;

NC_005304), Bean yellow mosaic virus (BYMV; NC_003492), Colombian datura virus (CDV; NC_020072), Celery mosaic virus (CeMV; NC_015393), Chilli veinal mottle virus (ChiVMV; NC_005778), Clover yellow vein virus (CIYVV; NC_003536), Cocksfoot streak virus (CSV; NC_003742), Carrot thin leaf virus (CTLV; NC_025254), Dasheen mosaic virus (DsMV; MG602234), Euphorbia ringspot virus (EuRSV; NC_031339), Freesia mosaic virus (FreMV; NC_014064), Hippeastrum mosaic virus (HiMV; NC_017967), Henbane mosaic virus (HMV; MH779476), Hordeum mosaic virus (HoMV; NC_005904), Hyacinth mosaic virus (HyaMV; NC_037051), Iris mild mosaic virus (IMMV; JF320812), Iris severe mosaic virus (ISMV; NC_029076), Johnsongrass mosaic virus (JGMV; NC_003606), Konjac mosaic virus (KoMV; NC_007913), Lettuce mosaic virus (LMV; NC_003605), Leek yellow stripe virus (LYSV; NC_004011), Maize dwarf mosaic virus (MDMV; NC_003377), Moroccan watermelon mosaic virus (MWMV; NC_009995), Narcissus degeneration virus (NDV; NC_008824), Narcissus late season yellows virus (NLSYV; NC_023628), Narcissus yellow stripe virus (NYSV; NC_011541), Ornithogalum mosaic virus (OrMV; NC_019409), Onion yellow dwarf virus (OYDV; NC_005029), Peanut mottle virus (PeMoV; NC_002600), Pepper mottle virus (PepMoV; NC_001517), Pepper severe mosaic virus (PepSMV; NC_008393), Pokeweed mosaic virus (PkMV; NC_018872), Plum pox virus (PPV; NC_001445), Papaya ringspot virus (PRSV; NC_001785), Pea seed-borne mosaic virus (PSbMV; NC_001671), Peru tomato mosaic virus (PTV; NC_004573), Potato virus A (PVA; NC_004039), Pepper veinal mottle virus (PVMV; NC_011918), Potato virus V (PVV; NC_004010), Potato virus Y (PVY; NC_001616), Passionfruit woodiness virus (PWV; NC_014790), Ryegrass mosaic virus (RgMV; NC_001814), Sugarcane mosaic virus (SCMV; NC_003398), Soybean mosaic virus (SMV; NC_002634), Sweet potato feathery mottle virus (SPFMV; NC_001841), Sweet potato latent virus (SPLV; NC_020896), Sorghum mosaic virus (SrMV; NC_004035), Tobacco etch virus (TEV; NC_001555), Turnip mosaic virus (TuMV; KF595121), Tobacco vein mottling virus (TVMV; NC_001768), Watermelon mosaic virus (WMV; NC_006262), Wild potato mosaic virus (WPMV; NC_004426), Wisteria vein mosaic virus (WVMV; NC_007216), Yam mosaic virus (YMV; NC_004752), and Zucchini yellow mosaic virus (ZYMV; NC_003224).

There were large differences in the plant resistance spectrum and in the virus host range breadth in the remaining dataset of 39 plant species. The resistance spectrum varied from 8% to 95% of the potyviruses (mean of 62%) across plant species. Taking into account the potyviruses for which host/non-host data were available for at least 12 plant species, the host range breadth varied from 30% to 93% of the plant species (mean of 62%) across potyviruses. For most (35 of 39) plant species, the ER model, which assumes the same rate for host-to-non-host and non-host-to-host transitions, was preferred over the ARD model, which considers two different rates for these two kinds of transitions, according to likelihood ratio tests (LRTs), or was the only feasible model because the ARD model did not converge (Table 1). For the four remaining plant species, the ARD model fitted the data slightly better than the ER model according to LRTs, with probability values close to the 5% threshold (3.4% to 4.7%; Table 1). However, under the ARD model, the state probabilities P_i of being host or non-host associated to many internal branches in the virus tree were close to 50%, which precluded reconstruction of host range changes. Consequently, since the LRTs were close to the 5% threshold, we also used the host/non-host state inferences obtained with the ER model for these four plant species.

Using the state probabilities associated to each branch in the tree, we could reliably infer the host range changes and identify the corresponding branches in the potyvirus phylogeny for 31 of the 39 plant species. The state of many internal branches in the potyvirus tree could not be inferred reliably for *Nicotiana tabacum*, with P_i values close to 50%, probably because there were too many state changes (at least 15) for this plant species. Consequently, *N. tabacum* was discarded from following analyses. For seven other plant species (*Avena sativa*, *Capsicum annuum*, *Chenopodium amaranticolor*, *Nicotiana benthamiana*, *N. megalosiphon*, *Phaseolus vulgaris*, and *Vigna unguiculata*), there was an uncertainty for the number and sometimes the direction of host/non-host state transitions and for the identity of the corresponding branches in a particular part of the phylogenetic tree.

Table 1. Summary results of phylogenetic inference of host range change.

Plant Species	Viruses with Host/Non-Host Status	LRT (p Value) ¹	Potyvirus Ancestral State Inference ²	Permutation Test (p Value) ³	Non-Host to Host Changes ⁴	Host to Non-Host Changes ⁴	Number of Changes ⁴
<i>Allium cepa</i>	12	0.293	0.17	0.9	2	0	2
<i>Apium graveolens</i>	16	0.132	0.05	0.82	2	0	2
<i>Avena sativa</i>	14	0.824	0.36	0.56	2*	0.5*	2.5
<i>Beta vulgaris</i>	31	0.224	0.04	0.048	8	0	8
<i>Brassica campestris</i>	17	0.202	0.03	0.62	2	0	2
<i>Brassica oleracea</i>	23	0.303	0	0.56	2	0	2
<i>Capsicum annuum</i>	34	0.034	0.01	0.13	3.5*	2.5*	6
<i>Capsicum frutescens</i>	19	0.343	0.10	0.022	3	1	4
<i>Catharanthus roseus</i>	17	NA	0.16	0.93	4	0	4
<i>Chenopodium amaranticolor</i>	50	0.894	0.95	0.47	0	7.5*	7.5
<i>Chenopodium quinoa</i>	51	0.38	0.91	0.70	0	9	9
<i>Citrullus lanatus</i>	15	0.77	0.20	0.017	3	0	3
<i>Cucumis melo</i>	19	0.48	0.04	0.15	3	0	3
<i>Cucumis sativus</i>	47	0.043	0.04	0.13	9	1	10
<i>Datura stramonium</i>	48	0.090	0.01	0.3	6	0	6
<i>Glycine max</i>	28	0.431	0.13	0.016	4	1	5
<i>Hordeum vulgare</i>	15	0.074	0.06	0.58	2	0	2
<i>Lactuca sativa</i>	22	0.047	0.08	0.83	4	0	4
<i>Lathyrus odoratus</i>	14	0.189	0.68	0.61	0	3	3
<i>Lupinus albus</i>	13	0.363	0.96	0.94	0	1	1
<i>Medicago sativa</i>	18	0.179	0.10	0.25	3	0	3
<i>Nicotiana benthamiana</i>	31	0.864	0.70	0.96	0	6.5*	6.5
<i>Nicotiana clevelandii</i>	48	0.857	0.82	0.88	0	11	11
<i>Nicotiana glutinosa</i>	51	0.146	0.04	0.003	11	2	13
<i>Nicotiana megalosiphon</i>	22	0.705	0.33	0	2	1.5*	3.5
<i>Nicotiana sylvestris</i>	13	0.164	0.93	0.81	0	2	2
<i>Nicotiana tabacum</i>	53	0.167	0.33	NA	NA	NA	NA
<i>Phaseolus vulgaris</i>	49	0.413	0.15	0.11	7.5*	2	9.5
<i>Raphanus sativus</i>	20	0.2	0	0.42	1	0	1
<i>Solanum lycopersicum</i>	43	0.201	0.09	0.039	8	0	8
<i>Solanum melongena</i>	20	0.040	0.02	0.43	3	0	3
<i>Solanum tuberosum</i>	20	NA	0.19	0.29	5	0	5
<i>Trifolium pratense</i>	22	0.082	0	0.013	3	0	3
<i>Trifolium repens</i>	20	0.452	0.04	0.6	2	0	2
<i>Trigonella foenum-graecum</i>	14	0.186	0.92	0.93	0	2	2
<i>Triticum aestivum</i>	15	0.657	0.19	0.59	1	0	1
<i>Vicia faba</i>	33	0.385	0.09	0.004	5	2	7
<i>Vigna unguiculata</i>	44	0.131	0.06	0.26	6.5*	1	7.5
<i>Zea mays</i>	25	0.742	0	0.046	1	0	1
Average per plant species	-	-	-	-	3.12	1.49	4.61

¹ LRT: Likelihood ratio test of models all-rates-different (ARD) and equal-rates (ER); $p < 0.05$ means that model ARD is preferred over model ER. ² Probability P_{64} that the plant species was host of the potyvirus most recent common ancestor (MRCA; branch number 64 in Figure 1). For *Avena sativa*, P_{64} was above the 0.35 threshold (0.36), but the status of next branch in the tree was inferred with more confidence ($P_{65} = 0.24$) allowing reliable host change reconstruction. ³ The p value indicates the frequency of random permutations of host and non-host data among the potyvirus species providing a probability that the plant species was host of the potyvirus MRCA lower than the inferred one (P_{64} in column 4). ⁴ Number of changes among potyviruses (i.e., excluding rymoviruses). * Reconstruction of host range evolution was uncertain in some parts of the virus phylogenetic tree. Consequently, two alternative scenarios of host range evolution were considered, and the indicated number of changes corresponds to the average of these two scenarios. NA: Not available (for LRT, it means that model ARD did not converge).

Host status changes that could not be attributed reliably to a precise branch of the potyvirus phylogenetic tree were excluded from further analyses, except in the comparison of the number of host status changes among plant families. In that latter case, the mean number of host status changes in the different plausible evolutionary scenarios was considered (Table 1).

Among the 38 plant species for which a reliable inference of host or non-host status of the potyvirus MRCA was obtained, most (34/38; 89%) belonged to only six botanical families: Amaranthaceae, Poaceae, Brassicaceae, Solanaceae, Cucurbitaceae, and Fabaceae (Table 2). Most plant species (30/38; 79%) were inferred to be non-hosts for the potyvirus MRCA. The eight other plant species, which are inferred to be hosts for the potyvirus MRCA, include wild plant species frequently used as laboratory plants (three *Nicotiana* spp. and two *Chenopodium* spp.) and three members of the Fabaceae (*Lathyrus odoratus*, *Lupinus albus* and *Trigonella foenum-graecum*). Permutation tests showed that the host/non-host state of the potyvirus MRCA was primarily influenced by the total number of viruses for which the plant species had the same state (Table 1; column 5 'Permutation test'). However, for ten plant species inferred to be non-host of the potyvirus MRCA, the total number of viruses for which the plant species

is non-host was not sufficient to explain this inference and the distribution of viruses for which the plant species is host or non-host on the tree was also important.

Table 2. Distribution of the analyzed plant species among botanical families and crop or wild plant categories.

Family	Crop	Wild	Total
Amaranthaceae	2	1	3
Poaceae	4	0	4
Brassicaceae	3	0	3
Solanaceae	6	6	12
Cucurbitaceae	3	0	3
Fabaceae	10	0	10
Others	3	1	4

On average, for each plant species, 4.61 host/non-host state changes were inferred with our set of potyviruses. Non-host-to-host transitions were about twice more frequent than the opposite transitions (3.12 versus 1.49 per plant species on average) (Table 1). For nine (or 10 depending on the host status reconstruction for *Avena sativa*) plant species, both kinds of transitions were inferred to have occurred during potyvirus evolution. However, for most plant species, a single type of transition was inferred. Consequently, for plant species inferred to be non-hosts of the potyvirus MRCA, means of 3.95 host gains and 0.48 host loss were predicted, whereas means of zero host gain and 5.25 host losses were predicted for plant species inferred to be hosts of the potyvirus MRCA.

3.3. Distribution of Host Range Changes across Plant Families and Potyvirus Clades

In spite of some differences in the numbers of inferred host gains, host losses, or total host changes across plant families or potyvirus clades (Supplementary Figure S1), these differences were almost never significant statistically when considering change frequencies, i.e., numbers of changes per total branch length. The single exception was the significantly higher frequency of host changes in the Amaranthaceae than in the Poaceae (Kruskal–Wallis multiple test; p -value < 0.05) (Supplementary Figure S1). Accordingly, the frequencies of host changes, host gains, or host losses inferred in the six main analyzed plant families (Table 2) were within the 95% confidence intervals (CIs) of random Monte Carlo simulations. Monte Carlo simulations showed also only two significant departures from random expectations for host gains, host losses, or total host changes inferred among the 12 main potyvirus clades (Figure 1). First, no host gain was inferred in clade 4 (including CTLV and KoMV) whereas from 1 to 9 (95% CI) gains would be expected according to Monte Carlo simulations. Second, 11 host gains were inferred in clade six (including *Turnip mosaic virus* (TuMV), *Narcissus yellow stripe virus* (NYSV), and *Narcissus late season yellows virus* (NLSYV)) whereas from one to nine (95% CI) gains would be expected according to Monte Carlo simulations. In both cases, the 99% CIs of Monte Carlo simulations included the inferred value.

According to their position on the virus phylogenetic tree, most host range changes are recent. Indeed, about 3/4 of the changes (115/156; 74%) were placed on terminal branches, whereas 26% were placed on internal branches (Supplementary Table S2). Similar imbalances were also observed for host gains (73 and 35 on terminal and internal branches, respectively) and host losses (42 and six on terminal and internal branches, respectively). Distributions of host changes, gains, or losses in terminal and internal branches were however not significantly different from what could be expected at random if they occurred proportionally to the branch lengths (Monte Carlo simulations; $p > 0.05$). Taking into account the host range changes placed on terminal branches only, 1.24 host gain and 0.71 host loss per potyvirus species were inferred on average. The majority of the 35 host gains inferred onto internal branches of the potyvirus phylogenetic tree (21/35; 60%) occurred on only four branches. These were branch numbers 82 (grouping *Bean yellow mosaic virus* (BYMV) and *Clover yellow vein virus* (CIYVV); 11 inferred host gains), 100 (clade seven or *Pepper veinal mottle virus* (PVMV); three host gains), 104 (clade

eight or *Papaya ringspot virus* (PRSV); three host gains), and 105 (grouping clade nine or *Peanut mottle virus* (PeMoV) and clade 10 or *Bean common mosaic virus* (BCMV); four host gains) (Supplementary Table S3).

We also analyzed the co-occurrence of host changes, host gains, or host losses. Compared to the 95% CI of Monte Carlo simulations, there was a significant excess of co-occurrence of host gains (or of total host changes) for seven terminal branches and two internal branches of the potyvirus tree (Figure 1 and Supplementary Table S3).

Finally, we observed a large number of repeated (parallel) host gain co-occurrences (Table 3). Considering all possible pairs of plant species, a total of 119 host gain co-occurrences were observed on at least two different branches of the potyvirus phylogenetic tree. This is significantly higher than what would be expected by chance. Indeed, among 10,000 simulations where the host gains were randomly distributed among the tree branches proportionally to their length, a maximum of 23 (mean of eight) repeated co-occurrences of host gains were obtained. Among the 119 repeated co-occurrences of host gains, 43 (36%) corresponded to plant species of the same family and 55 (46%) corresponded to plant species sharing the same continent of origin, America, Africa, or Eurasia/Mediterranean Basin (Table 3). These frequencies are significantly higher than expected by chance considering the plant species distribution among families (p -value = 0.0047) or among continents of origin (p -value = 0.0476) by comparison with 10,000 Monte Carlo simulations. By contrast, repeated co-occurrence of host losses, or of one host gain and one host loss were much less frequent with nine and two cases, respectively (Supplementary Table S4).

Table 3. Host gain co-occurrence (i.e., host gains inferred on the same branch of the potyvirus tree) found repeatedly (i.e., on at least two different branches for a given pair of plant species).

Plant Species 1	Plant Species 2	Branch Name or Number (Figure 1)	Independent Host Gain Co-Occurrences	Plant Species 1 and 2 Belong to the Same Family	Plant Species 1 and 2 Belong to the Same Continent of Origin ¹
<i>Beta vulgaris</i>	<i>Brassica campestris</i>	PSbMV, TuMV	2	NO	YES (EU)
<i>Beta vulgaris</i>	<i>Brassica oleracea</i>	PSbMV, TuMV	2	NO	YES (EU)
<i>Beta vulgaris</i>	<i>Cucumis sativus</i>	PVY, TuMV, 82	3	NO	YES (EU)
<i>Beta vulgaris</i>	<i>Lactuca sativa</i>	LMV, TuMV	2	NO	YES (EU)
<i>Beta vulgaris</i>	<i>Medicago sativa</i>	PSbMV, 82	2	NO	YES (EU)
<i>Beta vulgaris</i>	<i>Nicotiana glutinosa</i>	BtMV, TEV, TuMV, 82	4	NO	NO
<i>Beta vulgaris</i>	<i>Phaseolus vulgaris</i>	LMV, 82	2	NO	NO
<i>Beta vulgaris</i>	<i>Solanum lycopersicum</i>	PVY, TEV	2	NO	NO
<i>Beta vulgaris</i>	<i>Solanum melongena</i>	PVY, TEV	2	NO	YES (EU)
<i>Beta vulgaris</i>	<i>Solanum tuberosum</i>	PVY, TEV	2	NO	NO
<i>Beta vulgaris</i>	<i>Vicia faba</i>	PSbMV, TuMV, 82	3	NO	YES (EU)
<i>Beta vulgaris</i>	<i>Vigna unguiculata</i>	PVY, 82	2	NO	NO
<i>Brassica campestris</i>	<i>Brassica oleracea</i>	PSbMV, TuMV	2	YES	YES (EU)
<i>Brassica campestris</i>	<i>Vicia faba</i>	PSbMV, TuMV	2	NO	YES (EU)
<i>Brassica oleracea</i>	<i>Vicia faba</i>	PSbMV, TuMV	2	NO	YES (EU)
<i>Capsicum annuum</i>	<i>Solanum lycopersicum</i>	TEV, 100	2	YES	YES (AM)
<i>Capsicum annuum</i>	<i>Solanum tuberosum</i>	TEV, 100	2	YES	YES (AM)
<i>Capsicum frutescens</i>	<i>Nicotiana glutinosa</i>	TEV, 72	2	YES	YES (AM)
<i>Catharanthus roseus</i>	<i>Datura stramonium</i>	PVMV, WMV	2	NO	NO
<i>Catharanthus roseus</i>	<i>Nicotiana glutinosa</i>	PVMV, WMV	2	NO	NO
<i>Citrullus lanatus</i>	<i>Cucumis sativus</i>	82, 104	2	YES	NO
<i>Citrullus lanatus</i>	<i>Glycine max</i>	82, 105	2	NO	NO
<i>Citrullus lanatus</i>	<i>Phaseolus vulgaris</i>	82, 105	2	NO	NO
<i>Citrullus lanatus</i>	<i>Vigna unguiculata</i>	82, 105	2	NO	YES (AF)
<i>Cucumis melo</i>	<i>Medicago sativa</i>	BYMV, 108	2	NO	NO
<i>Cucumis sativus</i>	<i>Datura stramonium</i>	HiMV, TuMV, WMV	3	NO	NO
<i>Cucumis sativus</i>	<i>Nicotiana glutinosa</i>	HiMV, PPV, TuMV, WMV, 82	5	NO	NO
<i>Cucumis sativus</i>	<i>Solanum lycopersicum</i>	HiMV, PPV, PVY	3	NO	NO
<i>Cucumis sativus</i>	<i>Trifolium pratense</i>	PPV, WMV, 82	3	NO	YES (EU)
<i>Cucumis sativus</i>	<i>Trifolium repens</i>	PPV, 82	2	NO	YES (EU)
<i>Cucumis sativus</i>	<i>Vicia faba</i>	TuMV, 82	2	NO	YES (EU)
<i>Cucumis sativus</i>	<i>Vigna unguiculata</i>	PeMoV, PPV, PVY, 82, 109	5	NO	NO
<i>Datura stramonium</i>	<i>Nicotiana glutinosa</i>	HiMV, H MV, PVMV, TuMV, WMV	5	YES	YES (AM)
<i>Glycine max</i>	<i>Phaseolus vulgaris</i>	SCMV, 82, 105	3	YES	NO
<i>Glycine max</i>	<i>Vicia faba</i>	82, 105	2	YES	YES (EU)
<i>Medicago sativa</i>	<i>Vicia faba</i>	PSbMV, 82	2	YES	YES (EU)
<i>Nicotiana glutinosa</i>	<i>Solanum lycopersicum</i>	HiMV, PPV, TEV, 86	4	YES	YES (AM)
<i>Nicotiana glutinosa</i>	<i>Solanum melongena</i>	PVMV, TEV	2	YES	NO
<i>Nicotiana glutinosa</i>	<i>Trifolium pratense</i>	PPV, WMV, 82	3	NO	NO
<i>Nicotiana glutinosa</i>	<i>Trifolium repens</i>	PPV, 82	2	NO	NO
<i>Nicotiana glutinosa</i>	<i>Vigna unguiculata</i>	PPV, 82	2	NO	NO
<i>Phaseolus vulgaris</i>	<i>Vicia faba</i>	82, 105	2	YES	NO

Table 3. Cont.

Plant Species 1	Plant Species 2	Branch Name or Number (Figure 1)	Independent Host Gain Co-Occurrences	Plant Species 1 and 2 Belong to the Same Family	Plant Species 1 and 2 Belong to the Same Continent of Origin ¹
<i>Solanum lycopersicum</i>	<i>Solanum melongena</i>	PVY, TEV	2	YES	NO
<i>Solanum lycopersicum</i>	<i>Solanum tuberosum</i>	PVY, TEV, 100	3	YES	YES (AM)
<i>Solanum lycopersicum</i>	<i>Vigna unguiculata</i>	PPV, PVY	2	NO	NO
<i>Solanum melongena</i>	<i>Solanum tuberosum</i>	PVY, TEV	2	YES	NO
<i>Trifolium pratense</i>	<i>Trifolium repens</i>	PPV, 82	2	YES	YES (EU)
<i>Trifolium pratense</i>	<i>Vigna unguiculata</i>	PPV, 82	2	YES	NO
<i>Trifolium repens</i>	<i>Vigna unguiculata</i>	PPV, 82	2	YES	NO
Total			119		

¹ EU: Eurasia and Mediterranean Basin; AF: Africa; and AM: Americas.

4. Discussion

Given the paucity of data on host range evolution in plant viruses, we conducted a global analysis to unravel its main tendencies in potyviruses, a wide group of RNA viruses. Using host range compendia, we could accurately infer, in the potyvirus phylogenetic tree, the changes in the capacity to infect 38 plant species, i.e., the changes in potential host range. In nature, potyviruses are transmitted horizontally by aphids, i.e., insect vectors from superfamily Aphidoidea. As potyviruses are easily mechanically transmissible, their host ranges can be tested in laboratory conditions without biases related to host or virus specificity of the aphid vectors. However, virus host range evolution in natural conditions can be shaped by extrinsic factors not taken into account here [5].

This study revealed the ubiquity of changes in the potential host range during potyvirus evolution. Given the recent age of the potyvirus radiation, which started about 7000 years ago and is consequently largely posterior to the speciation of the studied plant species, these host changes correspond to host jumps rather than to host–virus co-divergences. This time frame of host changes is quite similar to other virus species, genera, or families. Indeed, the ages of family Luteoviridae, genus *Sobemovirus*, and species *Cucumber mosaic virus* (family Bromoviridae), *Citrus tristeza virus* (family Closteroviridae), *Tomato yellow leaf curl virus* (family Geminiviridae), and *Turnip yellow mosaic virus* (family Tymoviridae) were estimated to be <12,000 years (reviewed in [27]). The age of genera *Begomovirus*, *Mastrevirus*, and *Tobamovirus* was more ancient, with at least partial evidence of host–virus co-divergence for the latter [8].

Eighty percent (30/38) of analyzed plant species were inferred to be non-host of the potyvirus MRCA and for these plant species 3.95 host gains and 0.48 host loss were inferred on average, among a set of 59 potyviruses. This suggests that overall potyviruses have undergone host range expansions, i.e., an increase in the number of potential hosts, rather than host shifts, i.e., concurrent host gains and losses. Taking into account all 38 plant species, averages of 3.12 host gains and 1.49 host loss were estimated. These host change frequencies are underestimated for several reasons. First, some host/non-host data are missing in our dataset. Indeed, when a virus is first described on a specific plant, plants from the same genus or family are frequently over-represented in subsequent host range studies, and very distant species (e.g. monocots for a dicot-infecting virus) are more rarely tested. Second, some plant species may be wrongly considered as non-hosts because of the choice of particular plant genotypes and/or virus isolates to perform the tests. Third, the host change inference is largely based on parsimony principles which minimize the numbers of host changes per branch length (i.e., time unit), and fourth, some plant species may have been excluded from our analyses because too large numbers of host changes have precluded their reliable inference. The number of plant species corresponding to the latter case is difficult to estimate because the lack of reliable inference may either be due to an excessive number of host changes, an excessive number of host/non-host missing data, or both.

The inferred host gains, host losses, and total host changes were fairly distributed among the main potyvirus clades and among the main plant families examined. However, it should be emphasized that

most plant species belonged to only six families corresponding to important groups of crop species, what may introduce a bias.

The majority (68%) of host gains were placed on the terminal branches of the potyvirus phylogenetic tree, meaning that they occurred after the potyvirus diversification into separate species. Among the 73 gains in potential host range that were inferred on terminal branches of the potyvirus tree, 27 cases (37%) correspond to natural host gains, i.e., the virus was found to infect the plant in field or natural conditions. This high relative frequency of host gains in terminal branches is compatible with a role of host jumps on the specialization and isolation of virus populations in new plant species environments and subsequently on potyvirus speciation [28]. Host gains inferred on internal branches of the potyvirus phylogenetic tree were scarcer and concentrated on a few branches. While the global radiation of potyviruses took place circa 7000 years ago, the diversification of some clades into species was estimated around 3600 years ago [28,29], and intraspecific diversification probably took place only a few centuries ago [30,31]. Indeed, many host–virus encounters that could lead to host jumps took place less than 500 years ago with the intensification of intercontinental exchanges of plants [30,32].

Given the small number of plant species analyzed in the present study and the bias in favor of crop species, it is unlikely to help identifying the plant species or families from which the genus *Potyvirus* originated. Accordingly, 80% (31/39) of the analyzed plant species were inferred to be non-hosts of the potyvirus MRCA (Table 1). Five of the remaining species are laboratory plants chosen in many studies because of their susceptibility to many viruses, inducing either local lesions or systemic infections (*Chenopodium amaranticolor*, *C. quinoa*, *Nicotiana benthamiana*, *N. clevelandii*, and *N. sylvestris*). The inference that they were hosts for the potyvirus MRCA may reflect the fact that they are hosts of most of the analyzed potyviruses (from 73% to 85%). The three other plant species (*Lathyrus odoratus*, *Lupinus albus*, and *Trigonella foenum-graecum*) belong to the same family Fabaceae and originate from the Near East or the Mediterranean Basin. However, the accuracy of their inference as hosts of the potyvirus MRCA should be taken with care because relatively few viruses of our dataset have been tested on these plants (13 or 14). Gibbs and Ohshima [18] assumed that the *Potyvirus* genus emerged from wild monocotyledonous plants around the Fertile Crescent of the Middle East. This assumption was based on the topology of the potyvirus phylogenetic tree. The sister group of potyviruses, genus *Rymovirus*, and the two “early-branching” clades of potyviruses (clades 11 or *Sugarcane mosaic virus* (SCMV) and 12 or *Onion yellow dwarf virus* (OYDV)) contain viruses that infect monocot plants (in families Poaceae or Liliaceae) that were domesticated in the broad geographical region spanning from Southwest Eurasia to North Africa. However, early-branching lineages do not necessarily signify ancestral traits as emphasized by Crisp and Cook [33] and intuitive assertion of ancestry from phylogenetic trees may be misleading. In our analyses, the five monocots (*Allium cepa* and four species of the Poaceae) were inferred to be non-hosts for the potyvirus MRCA but were only tested against a small subset of viruses (12 to 15, except *Zea mays*, a New World Poaceae, which was tested against 25 viruses). It remains also possible that, after jumping to crop plants, the ancestor of potyviruses had lost the capacity to infect plants belonging to its original species or family.

Molecular determinants of host jumps are largely unknown for plant viruses (but see [3] and references therein), in contrast with molecular determinants involved in the breakdown of resistance genes, i.e., host adaptation at the within-plant-species level. Perhaps the most striking result of this study is the large number of repeated correlated host gains in the potyvirus phylogenetic tree, i.e., the fact that several host gains were inferred on the same branch of the tree. It is questionable whether the molecular events involved in these different host changes are independent or not. Many of these correlated host gains occurred several times in the potyvirus phylogenetic tree and the frequency of such recurrent events largely outweighs the expectations of chance. Hence, it is probable that several of these host gains were not independent and that the acquired capacity to infect a first plant species increased the probability that the virus also acquired the capacity to infect a second plant species. Underlying molecular mechanisms could be either pleiotropy, i.e., the fact that a mutation is responsible for both host gains simultaneously, or the ‘springboard’ effect, where a mutation responsible

for a first host gain favors the acquisition of secondary mutations involved in the second host gain. Recombination is also widespread in potyviruses and may alter their genome in such a drastic way that multiple traits could be affected simultaneously by a single recombination event. Likewise, in our dataset, one potyvirus species, WMV, is the result of interspecific recombination between *Soybean mosaic virus* (SMV) and BCMV [34] and has an expanded host range compared to its parental species (37 species compared to 19 and 21 species, respectively, for SMV and BCMV; Supplementary Table S1). One would expect that these molecular events would have affected the capacity to infect plant species belonging to the same family, since in general host range barriers for plant viruses are much more frequent and/or stronger for plants belonging to different families than within families [35]. An alternative “bridgehead” hypothesis would be that acquiring the capacity to infect a novel plant species allows a virus to invade a novel ecosystem and increases its exposure and capacity to jump to other plant species of that ecosystem [5]. In that case, we would expect that recurrent co-occurrences of host gains would preferentially affect plant species from the same geographical origin. Both the molecular and bridgehead hypotheses are supported by the fact that plant species corresponding to repeated host gain co-occurrences belong more frequently to the same botanical family or share more frequently the same continent of origin than expected under random.

5. Conclusions

This study shows that the potyvirus host range followed a dynamic pattern, with frequent gains or losses during evolution and, surprisingly, frequent and repeated co-occurrences of host gains in different potyvirus clades. It also confirms the potentially important role of agriculture and crop plants on the diversification of potyviruses. Consequently, potyvirus evolution and speciation in the future could be strongly affected by future agricultural changes, such as the choice of new crop species or of growing crop species in new geographical areas. Indeed, agricultural environments are favorable for virus adaptation, specialization, and speciation, given the large number and genetic homogeneity of hosts in crops [27,28,36].

The molecular determinants involved in potyvirus host jumps, either from the virus or the plant side, are largely unknown (but see [6,14,16]). Given the multiple virus and host factors interacting during virus infection [37–39], it would be interesting to test formally the hypothesis of common genetic determinants involved in multiple host gains in potyviruses by reverse genetics approaches. Our study suggests candidate potyviruses and plant species for these approaches (Table 3).

Supplementary Materials: The following are available online at <http://www.mdpi.com/1999-4915/12/1/111/s1>; Table S1: Host range data used for phylogenetic analyses; Table S2: Inferred host changes corresponding to the different branches in the potyvirus phylogenetic tree (Figure 1). Branches with no changes and changes for which alternative scenarios exist are omitted; Table S3: Branches in the potyvirus phylogenetic tree where the number of inferred host gains significantly exceeded expectations of random distributions; Table S4. Co-occurrence of host losses (i.e., host losses inferred on the same tree branch) or host gain and loss that were found multiple times (i.e., on at least two different tree branches for pairs of plant species); Figure S1: Distribution of host gains, host losses or total host changes across potyvirus clades (Figure 1) and across plant families.

Author Contributions: Conceptualization: B.M.; methodology: B.M., C.D.; formal analysis: B.M., C.D.; writing: B.M., C.D.; writing—Review and editing: B.M., C.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We thank Emmanuel Szadkowski and Loup Rimbaud for insightful comments on a previous version of the manuscript. We thank Sylvie German-Retana and Kristiina Mäkinen for inviting us to contribute this special issue.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Use of Random Simulations to Establish a Threshold Value to Infer the Host Status of the Considered Plant Species for the Potyvirus Most Recent Common Ancestor (MRCA)

We used the topology and branch lengths of the phylogenetic tree obtained from our potyvirus and rymovirus dataset (Figure 1) to perform simulations in order to define a threshold probability value for the confidence on the host/non-host status of a given plant species for the potyvirus MRCA (branch number 64 in Figure 1).

For each batch of 10,000 simulations, an identical number N of host and non-host statuses corresponding to a fictitious plant species was fixed. N varied from six to 27 across simulation batches (hence there were 12 to 54 host/non-host data in total; Table A1), in accordance with our true datasets, where the sum of host and non-host statuses varied from 12 to 53, depending on the plant species (Table 1). Then, each simulation involved the two following steps. First, the N host and non-host statuses were randomly assigned to potyvirus species and the $(62 - 2 \times N)$ remaining potyvirus and rymovirus species were attributed “missing data” statuses. Second, for each internal branch i of the virus phylogenetic tree, we computed the probability P_i that the given plant species was a host for the corresponding potyvirus ancestor by maximum likelihood, using the equal-rates (ER) model of the ‘ace’ function in the ‘ape’ package of the software R version 3.5.3 (see Methods section for details). This allowed obtaining the probability P_{64} that the plant species was host of the potyvirus MRCA corresponding to branch 64 (Figure 1). Finally, we counted the number of simulations among 10,000 where $P_{64} > T$ or $(1 - P_{64}) > T$, T being a threshold value to be defined.

Table A1 indicates that between 0.79% and 1.67% of the 10,000 simulations had $P_{64} > 0.65$ or $(1 - P_{64}) > 0.65$, depending on N . Hence, using $T = 0.65$ as a threshold value ensures less than 2% type I errors in inferring the host or non-host status of a given plant species for the potyvirus MRCA.

It was not possible to define similarly a threshold value in the case of imbalance between the numbers of host and non-host statuses because this imbalance impacts directly the inference of the host or non-host status of the plant species for the potyvirus MRCA as shown in Table 1 (column 5, “Permutation test”).

Table A1. Summary results of simulations.

Simulation Batch (10,000 Each)	Simulated Datasets		Results of Simulations for the Host Status of the Plant Species for the Potyvirus MRCA			
	Number of Host or Non-Host Statuses (N)	Number of Missing Data ($62 - 2 \times N$)	Mean P_{64}	N_1 : Number of Simulations with $P_{64} > 0.65$	N_2 : Number of Simulations with $(1 - P_{64}) > 0.65$	$N_1 + N_2$
1	6	50	0.5	75	92	167
2	12	38	0.5	74	107	181
3	18	26	0.5	68	44	112
4	24	14	0.5	38	52	90
5	27	8	0.5	49	30	79

References

1. Woolhouse, M.E.J.; Gowtage-Sequeria, S. Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.* **2005**, *11*, 1842–1847. [[CrossRef](#)]
2. Fabre, F.; Rousseau, E.; Mailleret, L.; Moury, B. Durable strategies to deploy plant resistance in agricultural landscapes. *New Phytol.* **2012**, *193*, 1064–1075. [[CrossRef](#)] [[PubMed](#)]
3. Morris, C.E.; Moury, B. Revisiting the concept of host range of plant pathogens. *Annu. Rev. Phytopathol.* **2019**, *57*, 63–90. [[CrossRef](#)] [[PubMed](#)]
4. Rimbaud, L.; Papaix, J.; Barrett, L.G.; Burdon, J.J.; Thrall, P.H. Mosaics, mixtures, rotations or pyramiding: What is the optimal strategy to deploy major gene resistance? *Evol. Appl.* **2018**, *11*, 1791–1810. [[CrossRef](#)] [[PubMed](#)]

5. McLeish, M.; Sacristan, S.; Fraile, A.; Garcia-Arenal, F. Scale dependencies and generalism in host use shape virus prevalence. *Proc. R. Soc. B* **2017**, *284*, 20172066. [[CrossRef](#)]
6. Chen, K.C.; Chiang, C.H.; Raja, J.A.; Liu, F.L.; Tai, C.H.; Yeh, S.D. A single amino acid of NIaPro of Papaya ringspot virus determines host specificity for infection of papaya. *Mol. Plant-Microbe Interact.* **2008**, *21*, 1046–1057. [[CrossRef](#)] [[PubMed](#)]
7. Dombrovsky, A.; Tran-Nguyen, L.T.T.; Jones, R.A.C. Cucumber green mottle mosaic virus, Rapidly increasing global distribution, etiology, epidemiology, and management. *Annu. Rev. Phytopathol.* **2017**, *55*, 231–256. [[CrossRef](#)]
8. Gibbs, A.J.; Wood, J.; García-Arenal, F.; Ohshima, K.; Armstrong, J.S. Tobamoviruses have probably co-diverged with their eudicotyledonous hosts for at least 110 million years. *Virus Evol.* **2015**, *1*, vev019. [[CrossRef](#)]
9. Hilf, M.E.; Dawson, W.O. The tobamovirus capsid protein functions as a host specific determinant of long distance movement. *Virology* **1993**, *193*, 106–114. [[CrossRef](#)]
10. Nguyen, H.D.; Tomitaka, Y.; Ho, S.Y.; Duchêne, S.; Vetten, H.J.; Lesemann, D.; Walsh, J.A.; Gibbs, A.J.; Ohshima, K. Turnip mosaic potyvirus probably first spread to Eurasian brassica crops from wild orchids about 1000 years ago. *PLoS ONE* **2013**, *8*, e55336. [[CrossRef](#)]
11. Poulicard, N.; Pinel-Galzi, A.; Traoré, O.; Vignols, F.; Ghesquière, A.; Konate, G.; Hébrard, E.; Fargette, D. Historical contingencies modulate the adaptability of Rice yellow mottle virus. *PLoS Pathog.* **2012**, *8*, e1002482. [[CrossRef](#)]
12. Ryu, K.H.; Kim, C.H.; Palukaitis, P. The coat protein of cucumber mosaic virus is a host range determinant for infection of maize. *Mol. Plant-Microbe Interact.* **1998**, *11*, 351–357. [[CrossRef](#)]
13. Schoelz, J.E.; Shepherd, R.J. Host range control of cauliflower mosaic virus. *Virology* **1988**, *162*, 30–37. [[CrossRef](#)]
14. Suehiro, N.; Natsuaki, T.; Watanabe, T.; Okuda, S. An important determinant of the ability of Turnip mosaic virus to infect *Brassica* spp. and/or *Raphanus sativus* is in its P3 protein. *J. Gen. Virol.* **2004**, *85*, 2087–2098. [[CrossRef](#)]
15. Tatineni, S.; Robertson, C.J.; Garnsey, S.M.; Dawson, W.O. A plant virus evolved by acquiring multiple nonconserved genes to extend its host range. *Proc. Nat. Acad. Sci. USA* **2011**, *108*, 17366–17371. [[CrossRef](#)]
16. Vassilakos, N.; Simon, V.; Tzima, A.; Johansen, E.; Moury, B. Genetic determinism and evolutionary reconstruction of a host jump in a plant virus. *Mol. Biol. Evol.* **2016**, *33*, 541–553. [[CrossRef](#)]
17. ICTV. Available online: <https://talk.ictvonline.org/taxonomy> (accessed on 16 October 2019).
18. Brunt, A.; Crabtree, K.; Dallwitz, M.; Gibbs, A.; Watson, L. *Viruses of Plants: Descriptions and Lists from the VIDE Database*; C.A.B. International: Wallingford, UK, 1996; 1484p.
19. Gibbs, A.; Ohshima, K. Potyviruses and the digital revolution. *Annu. Rev. Phytopathol.* **2010**, *48*, 205–223. [[CrossRef](#)]
20. Edwardson, J.R.; Christie, R.G. The potyviruses. *Fla. Agric. Exp. Stat. Monogr. Ser.* **1991**.
21. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service, multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **2017**, *20*, 1160–1166. [[CrossRef](#)]
22. Kumar, S.; Stecher, G.; Tamura, K. MEGA7, Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)]
23. Pagel, M. Detecting correlated evolution on phylogenies, a general method for the comparative analysis of discrete characters. *Proc. R. Soc. B* **1994**, *255*, 37–45. [[CrossRef](#)]
24. Paradis, E.; Schliep, K. ape 5.0, an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **2018**, *35*, 526–528. [[CrossRef](#)]
25. R, A Language and Environment for Statistical Computing. Available online: <https://www.R-project.org> (accessed on 27 September 2019).
26. Le, S.Q.; Gascuel, O. LG, An improved, general amino-acid replacement matrix. *Mol. Biol. Evol.* **2008**, *25*, 1307–1320. [[CrossRef](#)]
27. Pagán, I. The diversity, evolution and epidemiology of plant viruses: A phylogenetic view. *Inf. Genet. Evol.* **2018**, *65*, 187–199. [[CrossRef](#)]
28. Gibbs, A.J.; Ohshima, K.; Phillips, M.J.; Gibbs, M.J. The prehistory of potyviruses: Their initial radiation was during the dawn of agriculture. *PLoS ONE* **2008**, *3*, e2523. [[CrossRef](#)]

29. Desbiez, C.; Wipf-Scheibel, C.; Millot, P.; Verdin, E.; Dafalla, G.; Lecoq, H. New species in the papaya ringspot virus cluster, Insights into the evolution of the PRSV lineage. *Virus Res.* **2017**, *241*, 88–94. [[CrossRef](#)]
30. Gibbs, A.J.; Fargette, D.; Garcia-Arenal, F.; Gibbs, M.J. Time—the emerging dimension of plant virus studies. *J. Gen. Virol.* **2010**, *91*, 13–22. [[CrossRef](#)]
31. Yasaka, R.; Fukagawa, H.; Ikematsu, M.; Soda, H.; Korkmaz, S.; Golnaraghi, A.; Katis, N.; Ho, S.Y.W.; Gibbs, A.J.; Ohshima, K. The timescale of emergence and spread of turnip mosaic potyvirus. *Sci. Rep.* **2017**, *7*, 4240. [[CrossRef](#)]
32. Kehoe, M.A.; Coutts, B.A.; Buirchell, B.J.; Jones, R.A.C. Hardenbergia mosaic virus: Crossing the barrier between native and introduced plant species. *Virus Res.* **2014**, *184*, 87–92. [[CrossRef](#)]
33. Crisp, M.D.; Cook, L.G. Do early branching lineages signify ancestral traits? *Trends Ecol. Evol.* **2005**, *20*, 122–128. [[CrossRef](#)]
34. Desbiez, C.; Lecoq, H. The nucleotide sequence of *Watermelon mosaic virus* (WMV, Potyvirus) reveals interspecific recombination between two related potyviruses in the 5′ part of the genome. *Arch. Virol.* **2004**, *149*, 1619–1632. [[CrossRef](#)]
35. Moury, B.; Fabre, F.; Hébrard, E.; Froissart, R. Determinants of host species range in plant viruses. *J. Gen. Virol.* **2017**, *98*, 862–873. [[CrossRef](#)]
36. Bernardo, P.; Charles-Dominique, T.; Barakat, M.; Ortet, P.; Fernandez, E.; Filloux, D.; Hartnady, P.; Rebelo, T.A.; Cousins, S.R.; Mesleard, F.; et al. Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *ISME J.* **2018**, *12*, 173–184. [[CrossRef](#)]
37. Revers, F.; Garcia, J.A. Molecular biology of potyviruses. *Adv. Virus Res.* **2015**, *92*, 101–199. [[CrossRef](#)]
38. Wang, A. Dissecting the molecular network of virus-plant interactions: The complex roles of host factors. *Annu. Rev. Phytopathol.* **2015**, *53*, 45–66. [[CrossRef](#)]
39. Wu, X.Y.; Valli, A.; Garcia, J.A.; Zhou, X.P.; Cheng, X.F. The tug-of-war between plants and viruses: Great progress and many remaining questions. *Viruses* **2019**, *11*, 25. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).