Author for correspondence:
Gemma G. R. Murray
e-mail: ggrm2@cam.ac.uk

Suppressors of RNAi from plant viruses are subject to episodic positive selection

Gemma G. R. Murray[1,2], Sergei L. Kosakovsky Pond[3] and Darren J. Obbard[1]

[1]Centre for Infection Immunity and Evolution, Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK
[2]Department of Genetics, University of Cambridge, Cambridge, UK
[3]Department of Medicine, University of California San Diego, San Diego, CA, USA

Viral suppressors of RNAi (VSRs) are proteins that actively inhibit the antiviral RNA interference (RNAi) immune response, providing an immune evasion route for viruses. It has been hypothesized that VSRs are engaged in a molecular 'arms race' with RNAi pathway genes. Two lines of evidence support this. First, VSRs from plant viruses display high sequence diversity, and are frequently gained and lost over evolutionary time scales. Second, *Drosophila* antiviral RNAi genes show high rates of adaptive evolution. Here, we investigate whether VSRs diversify faster than other genes and, if so, whether this is a result of positive selection, as might be expected in an arms race. By analysis of 12 plant RNA viruses, we show that the relative rate of protein evolution is higher for VSRs than for other genes, but that this is not attributable to pervasive positive selection. We argue that, because evolutionary time scales are extremely different for viruses and eukaryotes, it is improbable that viral adaptation (as measured by the ratio of non-synonymous to synonymous change) will be dominated by one-to-one coevolution with eukaryotes. Instead, for plant virus VSRs, we find strong evidence of episodic selection—diversifying selection that acts on a subset of lineages—which might be attributable to frequent shifts between different host genotypes or species.

1. Introduction

The interests of viruses and hosts often conflict: for a virus, host infection is necessary for replication, whereas for a host, infection can cause disease. This relationship exerts selective pressures on both host and virus, which may result in reciprocal adaptation and counter-adaptation in the form of an evolutionary 'arms race' [1]. At the genetic level, such arms races have been described for host and virus proteins that directly interact, and particularly in those involved in host antiviral immunity and viral evasion of host immunity [2–4]. The interaction between the RNA interference (RNAi) antiviral immune system of many eukaryotes and viral suppressors of RNAi appears to have the potential to instigate such an arms race [5,6].

RNAi-related pathways perform a range of functions in eukaryotes, but common to all is the role of short RNA molecules (approx. 20–30 nucleotides) in recognizing and manipulating complementary nucleotide sequences [7,8]. These systems have been found across eukaryotes [9], and function as an antiviral immune system in many lineages, including plants [10], *Drosophila* [11], mosquitoes [12], nematode worms [13] and fungi [14]. Antiviral RNAi pathways involve the Dicer family (Dcr) of proteins, which are members of the Ribonuclease III family of enzymes, the Argonaute family (Ago) [15], and various accessory proteins. Briefly, the pathway involves the recognition of viral dsRNA by Dcr, which dices it into short interfering RNAs (siRNAs). These are loaded into an Ago-containing effector complex, where one siRNA strand is lost and the other used to target and cleave RNA with the complementary sequence [7]. In plants [16] and in some animals [17], the small RNA signal is amplified and exported, resulting in non-cell-autonomous antiviral defence.

Many viruses express products that actively block the function of the antiviral RNAi pathway, termed viral suppressors of RNAi (VSRs), or RNA silencing suppressors (RSSs) [7,18]. VSRs are thought to be ubiquitous in viral genera. They have been found in RNA and DNA viruses, with both plant and animal hosts [18]. Suppression of the antiviral RNAi pathway by a VSR may often be a key stage of viral infection [7], and some viruses even encode multiple VSRs (e.g. potyviruses; P1 and HcPro) [19,20]. VSRs may inhibit the viRNAi pathway at various stages. Some bind dsRNA and sequester siRNAs away from the RNAi pathway. These include P10 of vitiviruses [21], NS3 of tenuiviruses [22], the NSs of tosposviruses [23], and the joint function of HcPro and P1 from potyviruses [24]. The 2b protein of cucumoviruses binds to Ago, preventing the RNA-induced silencing complex (RISC) from cleaving target RNA [25]. The P0 of poleroviruses induces the degradation of Ago [26]. Others inhibit cell-to-cell signalling of immunity, for example, the P30 of tobamoviruses [27] and 16k protein of tobraviruses [28]. A number of VSRs interfere with the pathway in multiple ways. For instance, HcPro inhibits both immunity in the infected cell and cell-to-cell signalling [29], and the P25 of potexviruses has been found to both prevent long-distance signalling [30] and induce the degradation of Ago [31]. On the other hand, it has been observed that the P1 of sobemoviruses inhibits the viRNAi pathway in the infected cell by removing siRNAs from the cell, but enhances the signalling of cell-to-cell immunity [32].

If the genes mediating antiviral RNAi pathways were engaged in a classical one-to-one arms race with VSRs, both host and virus genes might be expected to undergo rapid diversifying evolution under the force of strong positive selection. Consistent with this scenario, three key proteins in the antiviral RNAi pathway of *Drosophila* (Dcr-2, Ago-2 and R2D2) are among the most rapidly evolving genes in the *Drosophila* genome, and population-genetic analysis suggests that this is due to positive selection rather than relaxed constraint [33,34]. In addition, signatures of recent and recurrent selective sweeps can be found in Ago2 and Dcr2 across many *Drosophila* species [35,36].

If the genes controlling antiviral pathways are evolving rapidly and adaptively as the consequence of arms race selection, then VSRs are good candidates for the source of the antagonistic selection that drives this. There is some anecdotal evidence of rapid evolution in VSRs in viruses that infect plants. First, VSRs found in different viral families have no detectable sequence homology, even when their functions are similar, suggesting rapid evolution or multiple independent acquisitions [7,37]. Second, some VSRs appear to have arisen recently, perhaps as the result of adaptation to a host, suggesting the existence of selective pressure on VSR function [18]. Third, some VSRs in plant viruses show high protein sequence diversity within viral species relative to other genes (e.g. HcPro in potyviruses [20]), which is consistent with rapid evolution.

If the VSRs of plant viruses were engaged in an arms race with their host, this might be detectable as an elevated rate of non-synonymous substitutions (dN) relative to the rate of synonymous substitutions (dS), and thus a higher dN/dS ratio for VSRs than for other viral genes. However, while an elevated dN/dS might be suggestive of adaptive arms-race-driven evolution, it may also result from relaxed constraint. To test specifically for adaptive evolution one can compare the model fit for models of sequence evolution in which some codons evolve adaptively (dN/d$S > 1$) with those in which all codons are constrained to evolve neutrally or under selective constraint (dN/d$S \leq 1$) [38,39].

While a conventional arms race scenario implies constant reciprocal adaptation in both host and virus, in reality viral host-shifts can be frequent relative to the time scale of host evolution, so that selective pressures on the virus may vary across viral lineages. Therefore, in addition to testing for pervasive positive (diversifying) selection, we also took advantage of recent advances in the modelling of sequence evolution to test for episodic diversifying selection.

We performed these tests on all the known coding regions of the genomes of 12 plant viruses with described VSRs. These were selected because they have well-characterized VRSs and substantial publicly available genetic data. We compared the rates of protein evolution of VSRs with other genes and found that although VSRs did show elevated rates of non-synonymous to synonymous substitution, there was no evidence of ubiquitous positive selection, as might have been expected from a simplistic one-to-one arms race. Instead, we found strong evidence of episodic adaptation, consistent with coevolutionary dynamics that involve strong, but intermittent, positive selection.

2. Material and methods

(a) Sequence data

We searched the literature for publicly available data from single-stranded RNA viruses of plants with known VSRs. We chose not to include animal viruses as there are relatively few with well-characterized VSRs, and none of these has substantial population-genetic data. We identified 41 such viruses (see electronic supplementary material, table S1), but 29 of these had fewer than five alignable non-identical isolates in GenBank, making them unsuitable for phylogenetic analysis of adaptive sequence evolution because of the low power of such analyses on small alignments [40]. Our dataset, therefore, comprised the remaining 12 viruses, spanning 10 distinct viral genera (table 1). Some include more than one 'named' viral taxon, although all are predominantly from the species named, and all have divergence in a suitable range for our analyses. We have chosen to treat the P1 protein of potyviruses as a VSR, as it enhances the VSR activity of HcPro and, in the absence of HcPro, has evolved to act as a suppressor in its own right [46]. The datasets contained an average of 57 non-identical isolates (range 5–100). The within-species diversity varied substantially between genes and viruses: the average tree length for the viruses was 3.5 expected substitutions per codon (range 0.5–11.9), average gene length was 486 codons (range 17–2920), and average non-recombinant gene segment length was 336 codons (range 9–1711; see electronic supplementary material, table S2). Coding sequences for each viral gene were aligned using CLUSTALW in Bioedit [47] and adjusted by eye (alignments are available in the electronic supplementary material).

(b) Recombination and phylogenetic reconstruction

Since recombination can mislead phylogenetic analyses [48], we tested each gene alignment for evidence of recombinants using the GARD analysis implemented in datamonkey.org [49]. Genes in which recombination was detected were divided at the inferred break-points prior to the construction of phylogenetic trees [48]. For phylogenetic analysis by maximum likelihood (PAML) [39], trees were constructed for each non-recombinant gene segment by MRBAYES [50], using a partitioned

Table 1. Viruses analysed, their VSRs and which part of the viRNAi pathway they are thought to target.

family, genus	species	VSR	function	references
Alphaflexiviridae, *Potexvirus* (ssRNA+)	potato virus X (PVX)	P25	Argonaute and signal	[30,31]
Bunyaviridae, *Tospovirus* (ssRNA−)	tomato-spotted wilt virus (TSWV)	NSs	siRNA	[23]
Betaflexiviridae, *Vitivirus* (ssRNA+)	grapevine virus A (GVA)	P10	siRNA	[21]
Bromoviridae, *Cucumovirus* (ssRNA+)	cucumber mosaic virus (CMV)	2b	Argonaute	[41,42]
Luteoviridae, *Polerovirus* (ssRNA+)	sugarcane yellow leaf virus (SYLV)	P0	Argonaute	[43]
Potyviridae, *Potyvirus* (ssRNA+)	turnip mosaic virus (TurMV)	P1 and HcPro	siRNA and signal	[24,44,45]
	plum pox virus (PPV)			
	potato virus Y (PVY)			
Sobemovirus, *Sobemovirus* (ssRNA+)	rice yellow mottle virus (RYMV)	P1	siRNA	[32]
Tenuivirus, *Tenuivirus* (ssRNA+)	rice stripe virus (RSV)	NS3	siRNA	[22]
Unknown, *Tobamovirus* (ssRNA+)	tobacco mosaic virus (ToMV)	P30	signal	[27]
Unknown, *Tobravirus* (ssRNA+)	tobacco rattle virus (TRV)	16K	signal	[28]

(site-specific) rate model in which each codon position is ascribed a different rate. Run length ranged from 10 000 to 500 000 MCMC iterations, and chain convergence was determined by comparing two parallel runs and ensuring that variance in split frequencies dropped below 0.05. Burn-in length was determined by visual inspection of changes in log-likelihood over the MCMC, and maximum clade-credibility trees were used in downstream analysis. For the analysis using HYPHY [38], trees were constructed for each gene as part of the GARD analysis after using a codon model selector to determine the optimal model.

(c) Analysis of sequence evolution

Estimates of relative rates of protein evolution and tests for positively selected sites and classes of site (i.e. with $dN > dS$) were obtained through a phylogenetic approach implemented with two software packages: PAML v. 4 and HYPHY. An 'evolutionary fingerprint' analysis, which quantifies the pattern of constant positive selection and constraint across codons, a clustering analysis on these 'fingerprints' and a test for episodic selection were applied with HYPHY only, since comparable tests were unavailable in PAML [51,52]. Where possible, both packages were used to guard against our results being an artefact of a particular methodology or a set of assumptions.

Codeml (PAML) fits a codon substitution model to an alignment conditional on a phylogenetic tree using maximum likelihood [53]. Codeml model M0 was used to estimate a single best-fit dN/dS (ω) for each non-recombining gene segment, with confidence intervals calculated by the curvature method implemented in codeml [54], and two pairs of other models (codeml models M1a versus M2a and M8a versus M8) were used to test for site-specific positive selection through likelihood ratio tests (LRTs) [39]. In addition to fitting the rate class parameters to the data, codeml implements a 'Bayes empirical Bayes' approach that infers posterior probabilities of site classes for each site [55]. These probabilities were used to identify sites under positive selection.

The HYPHY package [38] provides three different pre-configured tests for the detection of site-specific positive selection, all of which were applied to the data: SLAC (single likelihood ancestor counting), REL (random effects likelihood) and FEL (fixed effects likelihood). REL was also used to provide an estimate of the average relative rate of protein evolution for each gene, expressed as $dN − dS$ to avoid numerical issues when dS is zero, which is possible because HYPHY permits synonymous substitution rates to vary

from site to site, while PAML fixes dS across sites [56]. PARRIS (a PARtitioning approach for Robust Inference of Selection), which allows site-variable dS, was used to provide a comparison with the LRTs in codeml. These HYPHY analyses were performed using the online interface www.datamonkey.org [57].

We also performed the 'evolutionary fingerprinting' and clustering analysis in HYPHY, which fits a general discrete bivariate model of evolutionary rates across a gene [51], with the number of rate classes in the model determined by the data. The 'evolutionary fingerprint' describes the joint distribution of synonymous and non-synonymous rates across codons within each gene, and the similarity between two fingerprints is quantified by a distance metric (termed the evolutionary selection distance, ESD). Calculating a distance matrix for a set of genes, allows us to compare their fingerprints.

Finally, we applied a recently developed mixed-effects model of evolution (MEME) test for site-specific episodic selection in HYPHY [52]. In general, tests for positive selection are relatively insensitive to brief periods of selection, as subsequent constraints can obscure a brief elevation in dN. However, MEME tests whether a non-zero proportion of branches is evolving with $dN > dS$ at each site, thereby gaining power to detect selection. The key difference between MEME and other methods is that the former require the mean dN/dS at a site to be greater than 1 when averaged over time (termed 'pervasive' or 'ubiquitous' positive selection), while MEME also detects bursts of selection followed by conservation that often yield mean $dN/dS < 1$, which would be missed by conventional approaches (termed 'episodic' positive selection). Simulation suggests that MEME is considerably more powerful than the other approaches, but equally accurate, often discovering 3–4 times the number of sites subject to episodic selection than are subject to pervasive selection [52].

(d) Statistical analysis of dN/dS and $dN − dS$

A meta-analysis of the gene-wise dN/dS estimates was performed to test for a difference between VSRs and other genes. We applied a variance (assuming variance~mean) stabilizing transformation (log) to the dN/dS point estimates. The delta method was applied to determine the variances of the log-transformed estimates [58]. We failed to normalize the distribution of average $dN − dS$ for each gene, and therefore only non-parametric methods were used on this measure. A model of the transformed point estimates of dN/dS estimates was fitted using the restricted maximum-likelihood software package ASREML [59]. In the model, gene class

(i.e. VSR or non-VSR) was treated as a fixed effect, and gene and viral family were treated as random effects. Numerical variance estimates obtained from PAML were taken into account by weighting the estimates inversely by the transformed variances.

Two non-parametric tests were also applied to test for a difference in dN/dS (or $dN-dS$) between VSRs and other genes. A Mann–Whitney U-test was performed on both the dN/dS and $dN-dS$ estimates, though this test fails to account for the effect of virus species on the rate of evolution. In addition, the probability of the observed rankings of VSRs (when ordered by dN/dS or $dN - dS$) within each virus was calculated through use of Fisher's method of combining the p-values for each individual virus (i.e. the probability that a VSR has the observed rank or higher, given the number of genes in that virus). This test treats dN/dS (or $dN-dS$) as a factor nested within species.

(e) Statistical analysis of tests for site-specific selection

Few positively selected sites were discovered using the tests for pervasive positive selection. However, such tests may suffer from deficiencies in power. Despite this, if VSRs are under an unusually strong selective pressure, which may be expected under the reciprocal selection (arms race) scenario, we might expect them to be more frequently identified as containing a class of positively selected sites than other types of genes. This hypothesis was tested using Fisher's exact tests (FETs) on the numbers of VSRs and non-VSR genes where positively selected sites were or were not detected. To evaluate potential bias in these tests, the statistics that jointly determine their power were also tested, and no significant difference was found between VSRs and other genes through Mann–Whitney U-tests on sample size ($p = 0.22$), tree length ($p = 0.21$) and gene length ($p = 0.60$).

Many positively selected sites were discovered using the MEME test for episodic selection. Therefore, for this analysis, we performed tests on the proportion of sites detected by MEME at $p \leq 0.05$ as having a non-zero fraction of branches with $dN > dS$. We calculated the probability of the cumulative observed ranking of VSRs when ordered by proportion of branches under episodic selection (as done with the dN/dS and $dN-dS$ estimates) and performed a Mann–Whitney U-test on the proportions of sites detected. Despite the increased sensitivity of MEME compared with other approaches, power is finite and there will be unknown false-negatives.

(f) Statistical analysis of evolutionary fingerprint analysis

The significance of VSR clustering in the evolutionary fingerprint analysis was tested using a permutation test, allowing the comparison of the null distribution of ESDs between VSRs (estimated by permuting distances to calculate a null distribution) and the observed average ESD between VSRs to be compared.

3. Results

(a) Mean dN/dS is higher for viral suppressors of RNAi than for other genes

The meta-analysis suggests that VSRs evolve with a significantly higher mean dN/dS ratio than other classes of viral genes (Wald test: $p < 0.001$; see figure 1a,b; electronic supplementary material, table S3), although the effect is small (dN/dS effect size = 0.04). This was true whether or not the relative rate of protein evolution was modelled as a function of viral species. This result was supported by a statistically significant Mann–Whitney U-test performed on the $dN-dS$
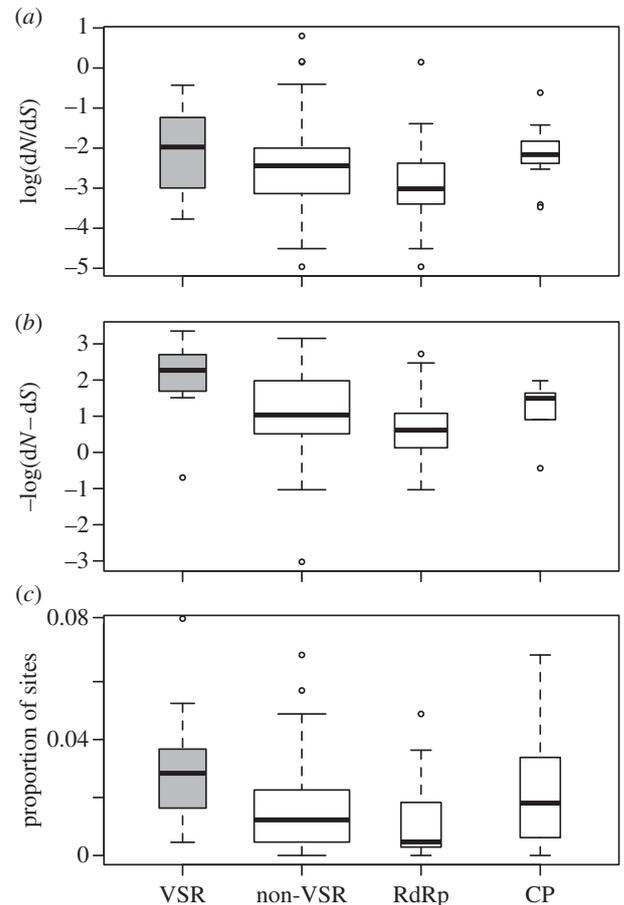

Figure 1. Boxplots of (a) log(dN/dS) estimates from PAML, (b) $-$log(d$N-$dS) estimates from REL and (c) MEME estimates of the proportion of sites under episodic selection. (a), (b) and (c) categorize these estimates into VSR genes (grey) and all other (non-VSR) genes, including coat proteins and RNA-polymerases, and also show the coat protein (CP) and RNA-dependent RNA polymerase (RdRp) genes separately. Widths of boxes reflect number of genes.

estimates from REL (HYPHY; $p = 0.044$), although not by the dN/dS estimates from *codeml* (PAML; $p = 0.089$). In addition, the VSR has the highest average dN/dS of any gene in six out of the 12 viruses we tested using the *codeml* (PAML) estimates and five out of 12 using the REL (HYPHY) estimates. The high ranking of the VSRs is unlikely to be by chance (PAML dN/dS ranking $p = 0.025$ and REL $dN-dS$ ranking $p = 0.011$, using Fisher's method for combining p-values).

However, it is known that certain types of viral genes are subject to significantly higher constraint than others. Therefore, the observation that VSRs evolve faster than other viral genes might not result from positive selection on VSRs, but rather from reduced constraint relative to other gene families. Tests for positive selection are required to determine this.

(b) Viral suppressors of RNAi do not show evidence of ubiquitous diversifying selection

LRTs for site-specific positive selection acting on VSRs did not provide evidence of consistent positive selection across VSRs (table 2; electronic supplementary material, table S4). Moreover, it was found that VSRs are no more likely to test positive than non-VSRs (PAML: $p = 0.54$ and PARRIS: $p = 0.57$, FET; but note that the power to detect selection will differ between genes). Similarly, site-specific analyses using REL (HYPHY) and *codeml* M8 (PAML) neither consistently

Table 2. Number of genes within a gene class (VSR; coat protein, CP; RNA-polymerase, RdRp; other; and non-VSR total) that showed significant evidence of positive selection ($p < 0.05$) and numbers of genes that did not ($p > 0.05$) through LRTs in PAML (M8a versus M8) and PARRIS (HyPhy).

gene class	PAML (M8a versus M8)		PARRIS	
	$p < 0.05$	$p > 0.05$	$p < 0.05$	$p > 0.05$
VSR	5	10	1	14
CP	6	6	0	12
RdRp	1	10	0	11
other	10	33	3	40
non-VSR total	17	49	3	63

detected positively selected codons in VSRs nor detected them significantly more frequently in VSRs than in other genes (FET, REL: $p = 1$, M8: $p = 1$, SLAC: $p = 0.21$, FEL: $p = 1$; table 3; electronic supplementary material, table S4). The results from these methods were not consistent, but this is not surprising; the tests use different criteria and vary in power and accuracy in detecting different patterns of selection in different datasets. Nevertheless, no VSR shows consistent positive results across all tests.

(c) Viral suppressors of RNAi evolutionary fingerprints do not cluster together

Evolutionary 'fingerprints' (the bivariate discrete distributions of dN/dS) were found to be no more similar between VSRs than would be expected by chance ($p = 0.31$ by permutation test), indicating that pervasive selective pressures shaping their evolution do not set them apart from other viral genes (example fingerprints in figure 2; all given in electronic supplementary material, figure S1; clustering diagram in figure 3). The same was found for coat proteins ($p = 0.14$) and polymerases ($p = 0.21$), which have been identified in almost all of the viruses in our dataset. Consistent with the tests we performed for pervasive positive selection, VSR fingerprints do not consistently have a class of sites with $\omega > 1$.

(d) Viral suppressors of RNAi are subject to unusually extensive episodic selection

In seven of the 12 viruses tested, a VSR is the gene with the greatest proportion of sites evolving under episodic selection as identified by MEME (see electronic supplementary material, table S4). The p-value for the likelihood of the observed ranking of VSRs within viruses under the null of random rankings (with respect to VSR) is 0.0008 (using Fisher's method for combining p-values). Under a Mann–Whitney U-test of the ranking of the VSRs with respect to other genes (not accounting for variation between viruses) it is found that VSRs have higher proportions of sites under episodic selection than other genes ($p = 0.024$; figure 1c). Furthermore, the pattern we observed in mean dN/dS and d$N-$dS across different types of genes (VSRs, non-VSRs, RdRps and CPs) is broadly reflected in the patterns in the proportion of sites found to be evolving under episodic selection (figure 1).

4. Discussion

We were unable to identify a ubiquitous elevated rate of adaptive evolution in VSRs when compared with other genes, as might have been expected under a conventional one-to-one arms race scenario. Although VSRs did show significantly higher mean rates of protein evolution (quantified in different analyses by dN/dS and d$N-$dS), few of them showed significant evidence of ubiquitous adaptive evolution, and this was not significantly different to the rate of 'positive' tests for genes with other functions. Thus, the slightly elevated rate of protein evolution in VSRs might be due to reduced constraint compared with other genes, such as polymerases, which are known to be highly conserved. However, VSRs do display a strikingly high proportion of sites evolving under episodic selection as identified by MEME. While MEME is capable of detecting both episodic and pervasive selection, the test detected a much larger number of sites under selection than tests that are insensitive to episodic selection (see electronic supplementary material, table S4); thus we conclude that episodic, rather than pervasive selection, is the driving force behind the MEME results.

Below, we suggest that the null result for conventional one-to-one arms race selection is unlikely to be due to low power, and that it instead reflects a potential asymmetry in selective responses between the host and the virus. This asymmetry means that, while host evolution can certainly drive adaptive substitutions in the virus (and vice versa), it is unlikely to result in a significantly elevated dN/dS ratio in the virus. Additionally, the high frequency of episodic selection in VSRs may suggest that the dominant positive selective pressures on viruses, at least over observable time scales, results from variation between host immune systems (genotype-to-genotype or species-to-species) rather than host adaptation.

(a) Power to detect selection is high

The power and accuracy of the tests for positively selected sites are dependent on sample size, tree length, gene length, and the pattern and strength of selection. Although the inferences that can be drawn from power comparisons with simulated datasets are limited, such studies suggest our power to detect moderate-to-strong persistent selection should have been substantial. For example, Wong $et\ al.$ [60] tested simulated data consisting of sequences of 500 codons with a tree length of three and 30 isolates. They found that when 10 per cent of sites are evolving with $\omega = 5$, 45 per cent with $\omega = 1$ and 45 per cent with $\omega = 0$, an LRT results in 76 per cent true-positives and no false-positives. Similarly, simulations by Kosakovsky Pond $et\ al.$ [38] using 250 codons, tree length 3 and 32 isolates suggest that the HyPhy REL analysis should provide a very powerful and moderately accurate test when one-fifth of sites are under positive selection with ω between 2 and 4 (nominal Bayes factor of 50; REL yields approx. 95% true-positives and 20% false-positives). In our dataset, the mean sample size was 57 sequences and the mean total tree length was 3.5 substitutions per codon; average gene length was 486 codons (see electronic supplementary material, table S2). These results suggest that our $codeml$ (PAML) M8a/M8 and REL analyses should provide substantial power to test for strong selection, and provide a less powerful but valid test for weak selection.
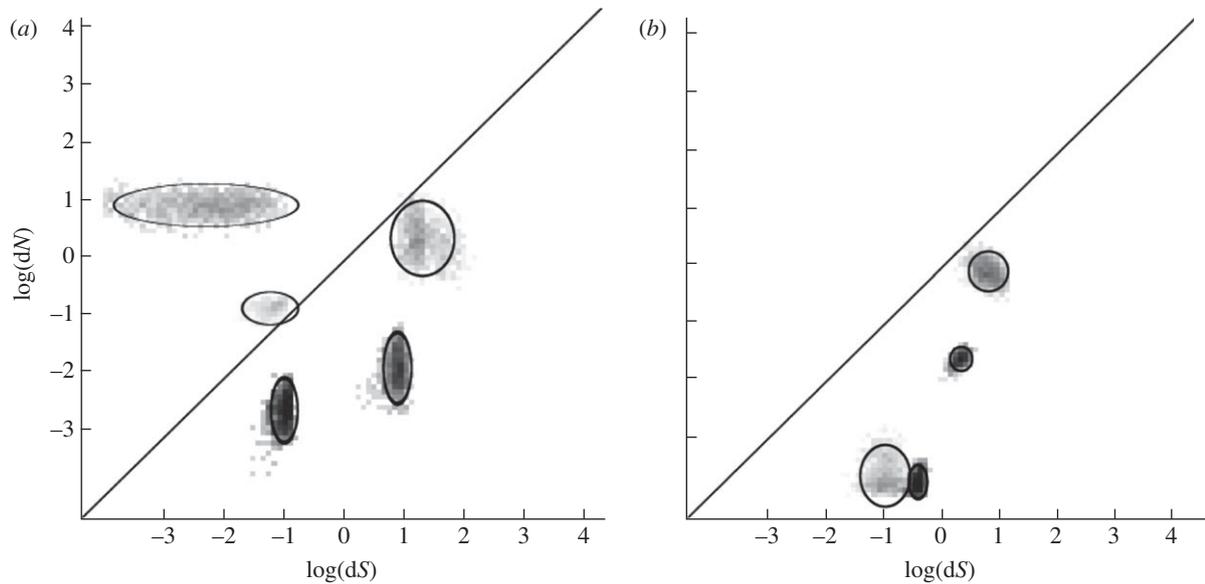
Figure 2. VSR fingerprints for (a) sugarcane yellow leaf VSR and (b) tobacco mosaic VSR. These plots describe the rate classes that have been inferred from the data: log(dN) against log(dS). The depth of colour represents the weight of a given estimate of the point ω value for that rate class. The ellipses are centred on approximate sampling means. The diagonal line represents a neutral rate (dN = dS). Rate classes evolving under positive selection are above the line, and ones evolving under constraint are below the line.

Table 3. Numbers of genes within a gene class (VSR; coat protein, CP; RNA-polymerase, RdRp; other; and non-VSR total) that did and did not have sites that were inferred to be evolving under positive selection (with $\omega > 1$) by REL (HᴙPʜʏ), M8 (codeml PAML), SLAC (HʏPʜʏ) and FEL (HʏPʜʏ).

gene class	HʏPʜʏ REL		PAML M8		HʏPʜʏ SLAC		HʏPʜʏ FEL	
sites with $\omega > 1$	present	not present	present	not present	present	not present	present	not present
VSR	9	6	5	10	7	8	10	5
CP	10	2	7	5	5	7	9	3
RdRp	5	6	4	7	3	8	9	2
other	25	16	12	29	10	31	23	18
non-VSR total	40	24	23	41	18	46	41	23

Thus, it seems unlikely that the apparent lack of strong selection acting on VSRs reflects low power alone.

(b) Ubiquitous positive selection acts only on a subset of viral suppressors of RNAi

We find that some VSRs show evidence of persistent positive selection, but that others do not, and this may be a true reflection of the evolutionary process. Even though VSRs as a group do not show evidence of pervasive diversifying selection, it is possible that the participation in an arms race is not uniform across VSRs, but rather that some VSRs are responding to selective pressures by rapid evolution and others are not. For example, in our analyses, 2b (CMV) shows evidence of adaptive evolution in all but the PARRIS analysis, which appears to be a conservative test, while NSs (TSWV) and P0 (SYLV) both had positive results in all but two tests. Nevertheless, we were unable to identify any mechanistic basis for differences in rate. VSRs can be categorized by how they suppress RNAi. NSs (TSWV), P10 (GVA), P1 (RYMV), NS3 (RSV), and HcPro and P1 (potyviruses) are thought to inhibit the accumulation of siRNAs; 2b (CMV),

P0 (SYLV) and P25 (PVX) are thought to act on Argonaute; and P25 (PVX), HcPro and P1 (potyviruses), P30 (ToMV) and 16K (TRV) are thought to inhibit cell-to-cell signalling. However, none of these groups consistently showed evidence of persistent positive selection across different analyses.

Host range is a further factor that could influence patterns of pervasive positive selection across viral species. Host range data on 11 of the viruses was downloaded from the Plant Viruses Online database [61] (see electronic supplementary material, table S5). The detection of positive selection in VSRs, the rate of protein evolution across all genes and the proportion of sites detected to be under episodic selection were found to be uncorrelated with any measure of host range—namely, number of known susceptible species (S_s), number of known susceptible families (F_s), proportion of tested species susceptible (S_t) and proportion of tested families susceptible (F_t). However, among VSRs, the rate of protein evolution (rather than the probability of testing 'positive') and the proportion of sites found to be under episodic selection appear to be weakly correlated with some measures of host range. Specifically, dN/dS was positively correlated with S_s (linear model, no correction for multiple testing,
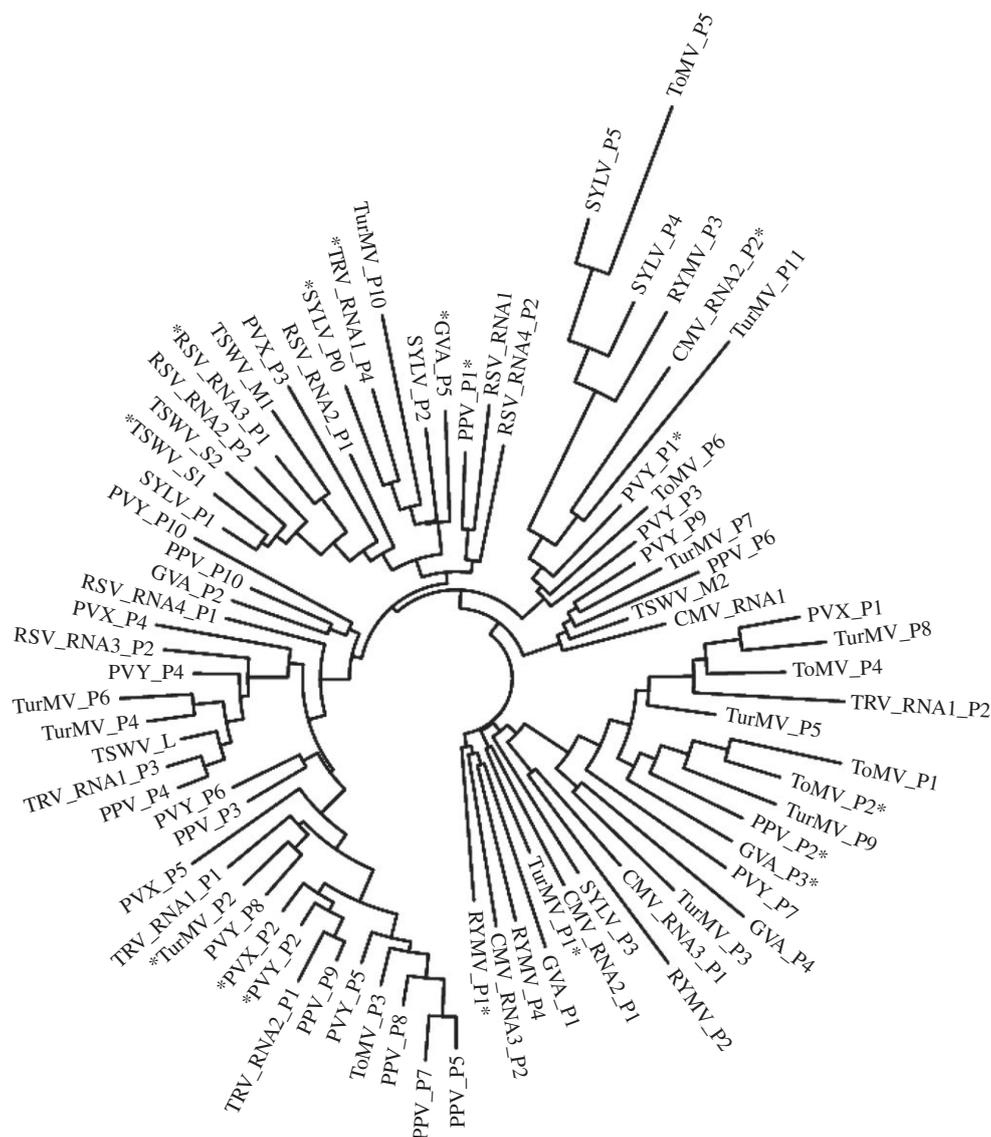
Figure 3. Clustering diagram of evolutionary distances between fingerprints. VSRs are asterisked. The length of the branches indicates the evolutionary selection distance (ESD) between genes.

$p = 0.001$; Spearman's rank correlation coefficient, $r_s = 0.37$), F_s ($p = 0.018$; $r_s = 0.35$), S_t ($p = 0.016$; $r_s = 0.55$), but not F_t ($p = 0.25$ $r_s = 0.38$). Similarly, the proportion of sites under episodic selection correlated with some measures of host range ($p = 0.001$, $r_s = 0.49$ for S_s; $p = 0.033$, $r_s = 0.44$ for F_s; $p = 0.13$, $r_s = 0.42$ for S_t; $p = 0.70$, $r_s = 0.19$ for F_t). This may suggest that host range plays a role in determining the strength or frequency of episodic selection on VSRs. However, experimental host range may be poorly known in many groups, and may not reflect host range in the wild. For this dataset, while S_s and F_s are correlated ($p = 0.00014$), and S_t and F_t are correlated ($p = 0.0010$), S_s and F_s do not correlate with S_t and F_t (for species: $p = 0.18$; for families: $p = 0.40$). This suggests that sampling strategies may have varied across viruses, and thus that this dataset is not ideal for such an analysis. This therefore warrants further study when the data allow.

(c) The separation of evolutionary time scales may make reciprocal coevolution hard to detect

Even if an arms race does occur, and it is reciprocal in the sense that adaptation in each party selects for counter-adaptation in the other, the elevated rate of adaptive evolution might be undetectable in the virus because of the different time scales over which evolution occurs in eukaryotes and viruses. Imagine a hypothetical scenario in which every amino acid substitution across the host genome was driven by one-to-one reciprocal coevolution with a single virus, and every amino acid substitution across the whole viral genome was similarly driven by one-to-one reciprocal coevolution with that host. Data from *Drosophila* suggest this may be on the order of one adaptive amino acid substitution every 50 years [62] for a multicellular eukaryotic host with large effective population size, short generation time and relatively compact genome, and it is unlikely to be substantially higher for most plants [63]. By the assumption of one-to-one reciprocity, this would drive one adaptive amino acid fixation every 50 years in the virus. Assuming synonymous substitutions are neutral, given eukaryotic mutation rates on the order of 1×10^{-8} site^{-1} yr^{-1} and viral mutation rates of 1×10^{-3} site^{-1} yr^{-1}, and respective genome sizes of 15 000 two-kb protein-coding genes and 10 one-kb protein-coding genes, this would amount to genome-wide dN/dS $\sim$ 0.1 for the host, but a dN/dS that was 50–100-fold lower for the virus. Even more extreme scenarios, such as a 1 : 10 host :

virus ratio of substitution, would still be difficult to detect, and the presumption that every single host substitution (regardless of gene) would mediate novel selection of the virus is unrealistically favourable to the detection of selection. Thus, the high mutation rate in RNA viruses may make such coevolutionary selection very hard to detect. This is in sharp contrast to the rapid adaptive evolution seen in viruses that infect vertebrates, whose evolution is not primarily driven by coevolution with the host, but by antagonism with an acquired host response that adapts plastically over the same time scale that governs viral evolution [64,65].

(d) Episodic selection and a one-sided arms race

There are at least two other reasons why we might not observe ubiquitous positive selection in these viruses. First, it has been suggested [4] that, while capable of rapid change, viruses are also under very high constraint owing to the necessity of successful interaction with the host for replication [66]. Second, viruses are able to move between hosts and host populations. High constraint may make it easier for a virus to move between hosts with varying immune systems than to adapt to a particular host immune system [4]. If this is the case, the selective pressure that drives evolution in RNAi genes might not result from rapid evolutionary change within VSRs, but result instead from changes in the composition of the viral community infecting particular host species. If the viruses that move between host species have sufficiently divergent VSRs, this process could also drive rapid evolution in the host. Although unknown, it seems plausible that this viral community changes rapidly over evolutionary time, given the rate of evolution in vertically transmitted genomic parasites such as transposable elements [67,68]. If this is the case, then depending on the frequency with which viruses shift between host species, or between individuals within a host species that display substantially divergent immune responses, we might expect selection acting on viruses to be episodic rather than ubiquitous, consistent with our results for VSRs. In the future, it would be very interesting to ask whether the shift between hosts is associated with transiently elevated dN/dS ratio. However, the extremely wide potential host range of some plant viruses, combined with the relatively poor sampling of viral lineages from noncrop plants and the small samples sizes available, precludes this analysis at present.

5. Conclusions

Although VSRs are predicted to be a focus of antagonistic host–virus interaction [5,6], we found little evidence for ubiquitous positive selection acting on the VSRs of plant viruses. Since our analyses are likely to have good power, we believe this is a robust result. However, VSRs do show slightly elevated rates of non-synonymous to synonymous substitution, and this appears to be associated with elevated rates of episodic selection and possibly with broad host range. Given the different time scales of host and pathogen evolution, ubiquitous selection driven by reciprocal arms races will be difficult to detect between viruses and eukaryotes, and therefore our results do not rule out the possibility of arms races having occurred between VSRs and antiviral RNAi genes, but instead demonstrate that these dynamics do not dominate the recent, observable evolution of the virus. The selective forces that we do detect are consistent with the type of selection that could be imposed by frequent shifts between selective environments, such as host shifts or local adaptation to host genotypes.

References

1. Dawkins R, Krebs JR. 1979 Arms races between and within species. *Proc. R. Soc. Lond. B* **205**, 489–511. (doi:10.1098/rspb.1979.0081)

2. Sawyer SL, Wu LI, Emerman M, Malik HS. 2005 Positive selection of primate *TRIM5α* identifies a critical species-specific retroviral restriction domain. *Proc. Natl Acad. Sci. USA* **102**, 2832–2837. (doi:10.1073/pnas.0409853102)

3. Elde NC, Malik HS. 2009 The evolutionary conundrum of pathogen mimicry. *Nat. Rev. Microbiol.* **7**, 787–797. (doi:10.1038/nrmicro2222)

4. Meyerson NR, Sawyer SL. 2011 Two-stepping through time: mammals and viruses. *Trends Microbiol.* **19**, 286–294. (doi:10.1016/j.tim.2011.03.006)

5. Marques JT, Carthew RW. 2007 A call to arms: coevolution of animal viruses and host innate immune responses. *Trends Genet.* **23**, 359–364. (doi:10.1016/j.tig.2007.04.004)

6. Obbard D, Gordon K, Buck A, Jiggins F. 2009 The evolution of RNAi as a defence against viruses and transposable elements. *Phil. Trans. R. Soc. B* **364**, 99–115. (doi:10.1098/rstb.2008.0168)

7. Ding S-W, Voinnet O. 2007 Antiviral immunity directed by small RNAs. *Cell* **130**, 413–426. (doi:10.1016/j.cell.2007.07.039)

8. Okamura K. 2011 Diversity of animal small RNA pathways and their biological utility. *Wiley Interdiscip. Rev. RNA.* **3**, 351–368. (doi:10.1002/wrna.113)

9. Cerutti H, Casas-Mollano JA. 2006 On the origin and functions of RNA-mediated silencing: from protists to man. *Curr. Genet.* **50**, 81–99. (doi:10.1007/s00294-006-0078-x)

10. Covey SN, Al-Kaff NS, Lángara A, Turner DS. 1997 Plants combat infection by gene silencing. *Nature* **385**, 781–782. (doi:10.1038/385781a0)

11. Li H. 2002 Induction and suppression of RNA silencing by an animal virus. *Science* **296**, 1319–1321. (doi:10.1126/science.1070948)

12. Keene KM. 2004 From the cover: RNA interference acts as a natural antiviral response to O'nyong-nyong virus (Alphavirus; Togaviridae) infection of *Anopheles gambiae*. *Proc. Natl Acad. Sci. USA* **101**, 17 240–17 245. (doi:10.1073/pnas.0406983101)

13. Felix MA *et al*. 2011 Natural and experimental infection of *Caenorhabditis* nematodes by novel viruses related to nodaviruses. *PLoS Biol.* **9**, e1000586. (doi:10.1371/journal.pbio.1000586)

14. Segers GC, Zhang X, Deng F, Sun Q, Nuss DL. 2007 Evidence that RNA silencing functions as an antiviral defense mechanism in fungi. *Proc. Natl Acad. Sci. USA* **104**, 12 902–12 906. (doi:10.1073/pnas.0702500104)

15. Cenik ES, Zamore PD. 2011 Argonaute proteins. *Curr. Biol.* **21**, R446–R449. (doi:10.1016/j.cub.2011.05.020)

16. Chapman EJ, Carrington JC. 2007 Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.* **8**, 884–896. (doi:10.1038/nrg2179)

17. Attarzadeh-Yazdi G *et al*. 2009 Cell-to-cell spread of the RNA interference response suppresses Semliki Forest virus (SFV) infection of mosquito cell cultures and cannot be antagonized by SFV. *J. Virol.* **83**, 5735–5748. (doi:10.1128/JVI.02440-08)

18. Li F, Ding S-W. 2006 Virus counterdefense: diverse strategies for evading the RNA-silencing immunity. *Annu. Rev. Microbiol.* **60**, 503–531. (doi:10.1146/annurev.micro.60.080805.142205)

19. Lu R, Folimonov A, Shintaku M, Li WX, Falk BW, Dawson WO, Ding SW. 2004 Three distinct suppressors of RNA silencing encoded by a 20-kb viral RNA genome. *Proc. Natl Acad. Sci. USA* **101**, 15 742–15 747. (doi:10.1073/pnas.0404940101)

20. Valli A, Lopez-Moya JJ, Garcia JA. 2007 Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *J. Gen. Virol.* **88**, 1016–1028. (doi:10.1099/vir.0.82402-0)

21. Zhou ZS. 2006 Identification of an RNA-silencing suppressor in the genome of Grapevine virus A. *J. Gen. Virol.* **87**, 2387–2395. (doi:10.1099/vir.0.81893-0)

22. Xiong R, Wu J, Zhou Y, Zhou X. 2009 Characterization and subcellular localization of an RNA silencing suppressor encoded by Rice stripe tenuivirus. *Virology* **387**, 29–40. (doi:10.1016/j.virol.2009.01.045)

23. Bucher E, Sijen T, de Haan P, Goldbach R, Prins M. 2003 Negative-strand tospoviruses and tenuiviruses carry a gene for a suppressor of gene silencing at analogous genomic positions. *J. Virol.* **77**, 1329–1336. (doi:10.1128/JVI.77.2.1329-1336.2003)

24. Lakatos L *et al*. 2006 Small RNA binding is a common strategy to suppress RNA silencing by several viral suppressors. *EMBO J.* **25**, 2768–2780. (doi:10.1038/sj.emboj.7601164)

25. Zhang X, Yuan YR, Pei Y, Lin SS, Tuschl T, Patel DJ, Chua NH. 2006 *Cucumber mosaic virus*-encoded 2b suppressor inhibits *Arabidopsis* Argonaute1 cleavage activity to counter plant defense. *Genes Dev.* **20**, 3255–3268. (doi:10.1101/gad.1495506)

26. Baumberger N, Tsai C-H, Lie M, Havecker E, Baulcombe David C. 2007 The Polerovirus silencing suppressor P0 targets ARGONAUTE proteins for degradation. *Curr. Biol.* **17**, 1609–1614. (doi:10.1016/j.cub.2007.08.039)

27. Kubota K, Tsuda S, Tamai A, Meshi T. 2003 Tomato mosaic virus replication protein suppresses virus-targeted posttranscriptional gene silencing. *J. Virol.* **77**, 11 016–11 026. (doi:10.1128/JVI.77.20.11016-11026.2003)

28. Martin-Hernandez AM, Baulcombe DC. 2008 Tobacco rattle virus 16-kilodalton protein encodes a suppressor of RNA silencing that allows transient viral entry in meristems. *J. Virol.* **82**, 4064–4071. (doi:10.1128/JVI.02438-07)

29. Mallory AC, Mlotshwa S, Bowman LH, Vance VB. 2003 The capacity of transgenic tobacco to send a systemic RNA silencing signal depends on the nature of the inducing transgene locus. *Plant J.* **35**, 82–92. (doi:10.1046/j.1365-313X.2003.01785.x)

30. Bayne EH, Rakitina DV, Morozov SY, Baulcombe DC. 2005 Cell-to-cell movement of Potato Potexvirus X is dependent on suppression of RNA silencing. *Plant J.* **44**, 471–482. (doi:10.1111/j.1365-313X.2005.02539.x)

31. Chiu MH, Chen IH, Baulcombe DC, Tsai CH. 2010 The silencing suppressor P25 of Potato virus X interacts with Argonaute1 and mediates its degradation through the proteasome pathway. *Mol. Plant Pathol.* **11**, 641–649. (doi:10.1111/j.1364-3703.2010.00634.x)

32. Lacombe S, Bangratz M, Vignols F, Brugidou C. 2010 The rice yellow mottle virus P1 protein exhibits dual functions to suppress and activate gene silencing. *Plant J.* **61**, 371–382. (doi:10.1111/j.1365-313X.2009.04062.x)

33. Obbard DJ, Jiggins FM, Halligan DL, Little TJ. 2006 Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr. Biol.* **16**, 580–585. (doi:10.1016/j.cub.2006.01.065)

34. Obbard D, Welch J, Kim K, Jiggins F. 2009 Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet.* **5**, e1000698. (doi:10.1371/journal.pgen.1000698)

35. Kolaczkowski B, Hupalo DN, Kern AD. 2011 Recurrent adaptation in RNA interference genes across the *Drosophila* phylogeny. *Mol. Biol. Evol.* **28**, 1033–1042. (doi:10.1093/molbev/msq284)

36. Obbard DJ, Jiggins FM, Bradshaw NJ, Little TJ. 2011 Recent and recurrent selective sweeps of the antiviral RNAi gene Argonaute-2 in three species of *Drosophila*. *Mol. Biol. Evol.* **28**, 1043–1056. (doi:10.1093/molbev/msq280)

37. Qu F, Morris T. 2005 Suppressors of RNA silencing encoded by plant viruses and their role in viral infections. *FEBS Lett.* **579**, 5958–5964. (doi:10.1016/j.febslet.2005.08.041)

38. Kosakovsky Pond SL, Frost SDW, Muse SV. 2005 HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679. (doi:10.1093/bioinformatics/bti079)

39. Yang Z. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)

40. Kosakovsky Pond SL. 2005 Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22**, 1208–1222. (doi:10.1093/molbev/msi105)

41. Brigneti G. 1998 Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. *EMBO J.* **17**, 6739–6746. (doi:10.1093/emboj/17.22.6739)

42. Canto T, Palukaitis P. 2001 A cucumber mosaic virus (CMV) RNA 1 transgene mediates suppression of the homologous viral RNA 1 constitutively and prevents CMV entry into the phloem. *J. Virol.* **75**, 9114–9120. (doi:10.1128/JVI.75.19.9114-9120.2001)

43. Mangwende T, Wang M-L, Borth W, Hu J, Moore PH, Mirkov TE, Albert HH. 2009 The P0 gene of sugarcane yellow leaf virus encodes an RNA silencing suppressor with unique activities. *Virology* **384**, 38–50. (doi:10.1016/j.virol.2008.10.034)

44. Kasschau KD, Xie Z, Allen E, Llave C, Chapman EJ, Krizan KA, Carrington JC. 2003 P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. *Dev. Cell* **4**, 205–217. (doi:10.1016/S1534-5807(03) 00025-X)

45. Simon-Mateo C, Garcia JA. 2006 MicroRNA-guided processing impairs Plum Pox virus replication, but the virus readily evolves to escape this silencing mechanism. *J. Virol.* **80**, 2429–2436. (doi:10.1128/JVI.80.5.2429-2436.2006)

46. Valli A, Martin-Hernandez AM, Lopez-Moya JJ, Garcia JA. 2006 RNA silencing suppression by a second copy of the P1 serine protease of Cucumber vein yellowing ipomovirus, a member of the family Potyviridae that lacks the cysteine protease HCPro. *J. Virol.* **80**, 10 055–10 063. (doi:10.1128/JVI.00985-06)

47. Larkin MA *et al*. 2007 Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948. (doi:10.1093/bioinformatics/btm404)

48. Anisimova M, Nielsen R, Yang Z. 2003 Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164**, 1229–1236.

49. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006 Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* **23**, 1891–1901. (doi:10.1093/molbev/msl051)

50. Ronquist F, Huelsenbeck JP. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574. (doi:10.1093/bioinformatics/btg180)

51. Kosakovsky Pond SL, Scheffler K, Gravenor MB, Poon AFY, Frost SDW. 2009 Evolutionary fingerprinting of genes. *Mol. Biol. Evol.* **27**, 520–536. (doi:10.1093/molbev/msp260)

52. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012 Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764. (doi:10.1371/journal.pgen.1002764)

53. Goldman N, Yang Z. 1994 A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736.

54. Schrago CG. 2006 An empirical examination of the standard errors of maximum likelihood phylogenetic parameters under the molecular clock via bootstrapping. *Gen. Mol. Res.* **5**, 233–241.

55. Yang Z. 2005 Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118. (doi:10.1093/molbev/msi097)

56. Kosakovsky Pond S, Muse SV. 2005 Site-to-site variation of synonymous substitution rates. *Mol. Biol. Evol.* **22**, 2375–2385. (doi:10.1093/molbev/msi232)

57. Delport W, Poon AF, Frost SD, Kosakovsky Pond SL. 2010 Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**, 2455–2457. (doi:10.1093/bioinformatics/btq429)

58. Davison AC. 2003 *Statistical models*. Cambridge, UK: Cambridge University Press.

59. Gilmour AR, Gogel BJ, Cullis BR, Thompson R. 2006 *ASReml user guide release 2.0*. Hemel Hempstead, UK: VSN International.

60. Wong WSW, Yang Z, Goldman N, Nielsen R. 2004 Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**, 1041–1051. (doi:10.1534/genetics.104.031153)

61. Brunt AA, Crabtree K, Dallwitz MJ, Gibbs AJ, Watson L, Zurcher EJ. (eds) 1996 Plant viruses online: descriptions and lists from the VIDE database. Version: 20th August 1996. See http://biology.anu.edu.au/Groups/MES/vide/.

62. Smith NG, Eyre-Walker A. 2002 Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022–1024. (doi:10.1038/4151022a)

63. Wright SI, Andolfatto P. 2008 The impact of natural selection on the genome: emerging patterns in *Drosophila* and *Arabidopsis*. *Annu. Rev. Ecol. Evol. Syst.* **39**, 193–213. (doi:10.1146/annurev.ecolsys.39.110707.173342)

64. Bhatt S, Holmes EC, Pybus OG. 2011 The genomic rate of molecular adaptation of the human influenza A virus. *Mol. Biol. Evol.* **28**, 2443–2451. (doi:10.1093/molbev/msr044)

65. Lemey P, Rambaut A, Pybus OG. 2006 HIV evolutionary dynamics within and among hosts. *AIDS Rev.* **8**, 125–140.

66. Belshaw R, Gardner A, RarnbaUt A, Pybus OG. 2008 Pacing a small cage: mutation and RNA viruses. *Trends Ecol. Evol.* **23**, 188–193. (doi:10.1016/j.tree.2007.11.010)

67. Bergman CM, Bensasson D. 2007 Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **104**, 11 340–11 345. (doi:10.1073/pnas.0702552104)

68. Hu TT *et al*. 2011 The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* **43**, 476. (doi:10.1038/ng.807)