Genetic diversity and molecular evolution of *Ornithogalum mosaic virus* based on the coat protein gene sequence

Fangluan Gao, Zhenguo Du, Jianguo Shen Corresp., Hongkai Yang, Furong Liao Corresp.

1 Fujian Key Laboratory of Plant Virology, Institute of Plant Virology, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China
2 Inspection and Quarantine Technology Center, Fujian Exit-Entry, Inspection and Quarantine Bureau, Fuzhou, Fujian, China
3 Inspection and Quarantine Technology Center, Xiamen Exit-Entry Inspection and Quarantine Bureau, Xiamen, Fujian, China

Corresponding Authors: Jianguo Shen, Furong Liao
Email address: shjg8765@126.com, lfr005@163.com

*Ornithogalum mosaic virus* (OrMV) has a wide host range and affects the production of a variety of ornamentals. In this study, the coat protein (CP) gene of OrMV was used to investigate the molecular mechanisms underlying the evolution of this virus. The 36 OrMV isolates fell into two groups which have significant subpopulation differentiation with an $F_{ST}$ value of 0.470. One isolate was identified as a recombinant and the other 35 recombination-free isolates could be divided into two major clades under different evolutionary constraints with $dN/dS$ values of 0.055 and 0.028, respectively, indicating a role of purifying selection in the differentiation of OrMV. In addition, the results from analysis of molecular variance (AMOVA) indicated that the effect of host species on the genetic divergence of OrMV is greater than that of geography. Furthermore, OrMV isolates from the genera *Ornithogalum*, *Lachenalia* and *Diuri* tended to group together, indicating that OrMV diversification was maintained, in part, by host-driven adaptation.
Genetic diversity and molecular evolution of Ornithogalum mosaic virus based on the coat protein gene sequence

Fangluan Gao¹, Zhengu Du¹, Jianguo Shen², Hongkai Yang¹, Furong Liao³

¹ Fujian Key Laboratory of Plant Virology, Institute of Plant Virology, Fujian Agriculture and Forestry University, Fuzhou, P. R. China
² Inspection and Quarantine Technology Center, Fujian Exit-Entry, Inspection and Quarantine Bureau, Fuzhou 350001, P. R. China
³ Inspection and Quarantine Technology Center, Xiamen Exit-Entry Inspection and Quarantine Bureau, Xiamen, 361012, P.R. China

Corresponding author:
Jianguo Shen, shjg8765@126.com
Furong Liao, lfr005@163.com
Abstract

*Ornithogalum mosaic virus* (OrMV) has a wide host range and affects the production of a variety of ornamentals. In this study, the coat protein (CP) gene of OrMV was used to investigate the molecular mechanisms underlying the evolution of this virus. The 36 OrMV isolates fell into two groups which have significant subpopulation differentiation with an $F_{ST}$ value of 0.470. One isolate was identified as a recombinant and the other 35 recombination-free isolates could be divided into two major clades under different evolutionary constraints with $dN/dS$ values of 0.055 and 0.028, respectively, indicating a role of purifying selection in the differentiation of OrMV. In addition, the results from analysis of molecular variance (AMOVA) indicated that the effect of host species on the genetic divergence of OrMV is greater than that of geography. Furthermore, OrMV isolates from the genera *Ornithogalum*, *Lachenalia* and *Diuri* tended to group together, indicating that OrMV diversification was maintained, in part, by host-driven adaptation.
Introduction

RNA viruses, many of which threaten human health or agricultural safety, form measurably evolving populations as a result of their high mutation rate and short generation times. Molecular evolution studies are useful in understanding the molecular bases of the adaptation, geographical expansion, and process of emergence of RNA viruses, which are key to the design of management measures (Lauring & Andino 2010; Moya et al. 2000).

*Ornithogalum mosaic virus* (OrMV) is one of the most important viral pathogens of floricultural crops, causing severe leaf symptoms as well as flower deformation of the affected plants (Burger et al. 1990). Under natural conditions, OrMV has a wide host range, infecting plants of the genera *Gladiolus*, *Iris*, *Ornithogalum* and *Diuris* (Burger & von Wechmar 1989; Kaur et al. 2011; Wylie et al. 2013). In addition, OrMV can infect saffron corms (*Crocus sativus*) as described in our previous report (Liao et al. 2017). OrMV was first detected in the United States in 1940 (Smith & Brierley 1944). After that, OrMV has been reported in Netherlands (Bouwen & von der Vlugt 1989), France (Grisoni et al. 2006), South Africa (Burger & von Wechmar 1989), Israel (Zeidan et al. 1998), India (Kaur et al. 2011), South Korea (Cho et al. 2016), Japan (Fuji et al. 2003), Zealand (Wei et al. 2006), Australia (Wylie et al. 2013) and China (Chen et al. 2009).

OrMV is a member of the genus *Potyvirus*, which includes more than 100 viral species. The biology and molecular biology of OrMV have not been studied intensively. However, it is known that, similar to some well-characterized potyviruses, OrMV has a single-stranded, positive-sense RNA genome, encoding a single polyprotein which is cleaved into 10 mature proteins by three virus-specific proteases (King et al. 2011). Additionally, a short polypeptide (PIPO) is expressed by a +2 nucleotide frame shifting from the P3 crison, resulting in a P3-PIPO fusion product dedicated to movement of the virus *in planta* (Chung et al. 2008; Wei et al. 2010). Although only 5 complete genomes of OrMV have been determined, CP sequences of 36 OrMV isolates are publicly available from GenBank. In this study, CP sequences were used to investigate the genetic diversity of OrMV and investigate the evolutionary forces responsible for the diversity. Our results will improve understanding of viral genetic variation and adaptive evolution, which may be helpful in developing sustainable management strategies for control of OrMV.

Materials & Methods

**Virus isolates and sequence alignment**

CP gene sequences with known geographic locations and host origins were obtained from GenBank database using its Batch Entrez facility (Table S1). Multiple sequence alignments were performed with MUSCLE codon algorithm (Edgar 2004) implemented in MEGA5 (Tamura et al. 2011).

**Phylogenetic network and recombination analyses**

Two different approaches were used to investigate the occurrence of recombination events in CP sequences. First, the aligned CP gene sequences of 36 OrMV isolates were analysed using the Neighbor-Net method in SplitsTree 4.13.1 (Huson 1998). In contrast to traditional bifurcating phylogenetic trees, SplitsTree constructs phylogenetic networks which allows for reticulation in the evolutionary relationships among taxa. Such reticulation could highlight the presence of
recombination and if present, we followed up with a second analysis.

Second, sequences involved in the recombination and breakpoints were determined by using RDP4 suite (Martin et al. 2015), which incorporates the algorithms RDP, GENECONV, BOOTSCAN, MAXCHI, CHIMAERA, SISCAN, and 3SEQ. For each putative recombination breakpoint, a Bonferroni correction P-value (with a cutoff point at $P<0.01$) was calculated. All isolates recognized were considered probable recombinants, supported by at least four different algorithms in RDP 4 with an associated $P$-value of $<1.0 \times 10^{-4}$. Simultaneously, the recombinants were further confirmed by GARD (Kosakovsky Pond et al. 2006) implemented in the Datamonkey web interface (Delport et al. 2010). The reliability of recombination breakpoints was evaluated using a KH test. To avoid false identification, only recombination breakpoints supported both by RDP4 and GARD were considered to be recombinants.

Genetic diversity and population subdivision

To investigate the genetic variation of the CP gene of OrMV, haplotype diversity ($H_d$) and nucleotide diversity ($\pi$) were calculated using DnaSP 5.0 (Librado & Rozas 2009). Hudson’s estimates of $K_{ST}$ and $S_{in}$ were used to determine the presence of subdivision in populations (Hudson 2000; Hudson et al. 1992). Genetic differentiation among populations was also evaluated by $F_{ST}$ using Arlequin 3.5 (Excoffier & Lischer 2010). The ranges of differentiation and corresponding $F_{ST}$ values were as follows: a moderate degree of differentiation, 0.05 to 0.15; a large degree, 0.15 to 0.25; and a substantial degree, $>0.25$ (Balloux & Lugon-Moulin 2002). In addition, analysis of molecular variance (AMOVA) was conducted using Arlequin 3.5 (Excoffier & Lischer 2010), with counties and host species as grouping factors to test for the effects of country and host on the genetic diversity of OrMV. The statistical significance of $\varphi$-statistics was tested based on 1023 permutations (default).

Phylogenetic analysis

After the potential recombinants were excluded, the phylogenetic relationships were reconstructed using the Maximum Likelihood (ML) approach implemented in MEGA5 (Tamura et al. 2011). For the ML analysis, substitution saturation was measured by Xia’s test implemented in DAMBE 5.3.8 (Xia 2013). The best-fitting of nucleotide substitution model was determined using MrModeltest (Nylander 2008). ML analysis was performed under the $GTR+\Gamma_4$ model using the corrected Akaike Information Criterion and the robustness of the ML tree topology was assessed with 1,000 bootstrap replicates.

Bayesian tip-association significance testing for the geographic and host species

To determine the potential geographic and host-origin effects on OrMV CP diversification, Bayesian Tip-association significance (BaTS) testing was performed in BEAST 2.4.6 (Bouckaert et al. 2014). Three statistics of phylogeny-trait association were computed: association index ($AI$), parsimony score ($PS$) and maximum monophyletic clade ($MC$) calculated from the posterior set of trees generated by BEAST 2.4.6 (Bouckaert et al. 2014). The statistical significance against the null distribution of trees was assessed by comparing it with the randomized trees generated from 10,000 reshufflings of tip characters. All P-values $< 0.05$ from the three statistics, with low $AI$ and $PS$ scores and a high $MC$ score, were considered significant, indicating a strong phylogeny-trait association.
Test for natural selection

Two different types of analyses were performed to test for natural selection using the CODEML algorithm (Yang 2007) implemented in EasyCodeML (https://www.github.io/bioeasy/EasyCodeml). Firstly, the branch model was used to identify CP genes with a null model assuming that the entire tree has been evolving at the same rate (one-ratio model) and an alternative model allowing foreground branch to evolve under a different rate (two-ratio model). Multiple testing was corrected by applying the false discovery rate (FDR) method (Storey & Tibshirani 2003) implemented in R. The CP gene of OrMV was considered as evolving with a significantly faster rate in foreground branch if the FDR-adjusted P-value less than 0.05 and a higher $\omega$ values ($\omega=dN/dS$, synonymous to non-synonymous substitution rates) in the foreground branch than the background branches. Secondly, the site model was used to identify nucleotide sites in the CP-coding region that were likely to be involved in OrMV evolution. For the site model, six codon substitution models described as M0, M1a, M2a, M3, M7, and M8, were investigated. The M1a model assumes two categories of sites ($\omega_0<1$, $\omega_1=1$), whereas the M2a model adds a third set of sites ($\omega_2>1$) to the M1a model. The M3 model, with three categories of sites, allows $\omega$ to vary among sites by defining a set number of discrete site categories, each with its own $\omega$ value. The M7 model partitions all the sites into ten different categories with $\omega<1$ and fits a beta distribution to $\omega$. In the M8 model, an 11th category is added to the M7 model allowing $\omega$ values >1. For each nested model, the likelihood ratio test (LRT) was conducted by comparing twice the difference in log-likelihood values ($2\Delta\ln L$) against a $x^2$-distribution, with degrees of freedom equal to the difference in the number of parameters between models. Only a P-value of 0.05 or less in the all LRTs was considered to be significant. Additionally, pairwise $dN/dS$ ratios were estimated using the yn00 program of PAML (Yang & Nielsen 2000). Isolates that $dS > 2x$ the mean $dS$ estimated from all isolates, as well as isolate pairs for which $dS$ estimates approached 0, were removed as advised by Finseth et al (2014).

Results

Recombination analyses

Recombination is an important source of genetic variability in viruses. To investigate the role of recombination in the evolution of OrMV, the split-decomposition network analysis with the CP gene sequences of 36 OrMV isolates was performed. A phylogenetic network showing reticulation was obtained (Fig. 1), indicating conflicting phylogenetic signals that are possibly attributed to recombination among viral genomes. The sequences were then checked for recombination using the RDP4 package (Martin et al. 2015). Four unique recombination events were detected by at least three independent methods implemented in the RDP suite (Table S2). However, only one isolate, Glad-8, was identified as a recombinant, with a breakpoint in the nucleotide 256, as confirmed by GARD analysis with a high level of confidence (both LHS and RHS p-values < 0.01). The recombinant was excluded from the phylogenetic and selection analyses below.

Genetic diversity and population subdivision

OrMV isolates could be divided into two subgroups reflecting two different origins of OrMV or
representing two divergent OrMV populations (Fig. 1). The haplotype diversity for both subgroup 1 and subgroup 2 was 1.000, whereas the nucleotide diversity for these two subgroups was 0.106 and 0.017, respectively. Haplotype diversity and nucleotide diversity for all OrMV isolates were 1.000 and 0.156, respectively, indicating a high genetic diversity in OrMV populations and among subpopulations (Table S3a). Three independent tests of population differentiation were significant (Table S3b), indicating a great genetic differentiation between clade groups of OrMV.

To evaluate the role of geography and host specificity in shaping the population structure of OrMV, geographic regions and host genus were respectively used as a grouping factor to analyze the isolates of OrMV. When geographic regions were used as grouping factors, AMOVA tests revealed significant variation among geographic groups, making up 15.85% of the total variation, \( \Phi_{ST} = 0.159, P\text{-value} < 0.001 \) (Table 1). Similar results were obtained when host species was used as a grouping factor. Significant subpopulation differentiation was observed among groups \( \Phi_{ST} = 0.297, P\text{-value} < 0.001 \), which accounted for nearly 30% of the total variation of OrMV. Taken together, it seems that the effect of host species on the genetic variance of OrMV is greater than that of geography although both host species and geographic effects contributed to the genetic variance of OrMV.

**Phylogenetic analyses and BaTS testing**

The ML phylogenetic trees based on the CP gene sequences showed that the 35 recombination-free OrMV isolates were grouped into two distinct clades with high bootstrap supports (Fig. 2A), consistent with the results of phylogenetic network analysis. With the exception of an isolate from Australia, no significant signal for geographic structure in the diversity of the CP gene was observed when the OrMV isolates were grouped by their geographic origins \( (P_{MC} > 0.05, \text{Table 2}) \). However, when the OrMV isolates were grouped by their host origins, a signal was found with more host-specific clustering than expected by chance, particularly for *Ornithogalum*, *Lachenalia* and *Diuris* \( (P_{MC} < 0.05, \text{Table 2}) \). The BaTS results indicated that OrMV CP diversification could be maintained in part by host-driven adaptation.

**Selection pressures**

To investigate the differences in selective pressures behind the two clades (clade A and B) of OrMV, a two-ratio branch model test was performed using PAML, in which different \( \omega \) values were assigned to the two clades. A LRT indicated that the one-ratio model should be rejected \( (p < 0.05, \text{Table S4a}) \); hence, selective pressures differed between the two clades. The mean \( \omega \) values for clades A and B were 0.055 and 0.028 (Fig. 2A, Table S4a), respectively, indicating that clade B was subjected to stronger purifying selection than clade A. Furthermore, the results from pair-wise analyses showed that there are differences between the distribution of \( dN/dS \) values between clade A and clade B (Fig. 2B). In the site model, there were no codons identified as being under positive selection and purifying selection was detected at the majority of polymorphic sites in the CP gene (Table S4b). Sliding-window analysis for sites under purifying selection was plotted in Fig. S1. Although the \( dN/dS \) values were below 1.00 for both clades, the
$dN/dS$ values of clade A were generally higher than those of clade B, indicating the CP gene in clade B had a stronger purifying selection pressure than those in clade A, in agreement with previous results from the branch model analysis.

**Discussion**

Recombination plays an important role in the evolutionary history of plant viruses, including potyviruses (Moreno et al. 2004) (Gao et al. 2016a; Gao et al. 2017; Ohshima et al. 2007), luteoviruses (Pagán & Holmes 2010) and cucumoviruses (Nouri et al. 2014). The greatest numbers of recombination events have been detected for the genus *Caulimovirus* (*Cauliflower mosaic virus*) in which the rate of recombination per base exceeds that of mutation (Froissart et al. 2005). However, the genetic variation generated by recombination is limited in OrMV and only one recombinant was observed in our analysis. There are two possible explanations. One is that the CP gene is a cold spot for recombination for OrMV. Such an idea has been proposed for some other plant viruses, such as *Chilli veinal mottle virus* (Gao et al. 2016a) and *Arabis mosaic virus* (ArMV)(Gao et al. 2016b). The other is that there is a strong selective pressure against the survival of OrMV recombinants. Consistently, purifying selection was detected at the majority of the polymorphic sites by two evolutionary analyses using the CODEML algorithm (Fig. S1), suggesting that most mutations in the OrMV CP gene were deleterious and consequently eliminated by natural selection.

Utilizing statistical models of variable $\omega$ ratios among sites, evidence of diversifying selection have been found in genes of potyvirus, such as *Potato virus Y* (Moury et al. 2002) and *Tobacco etch virus* (Cuevas et al. 2015). In this study, our results indicated that most codons of the OrMV CP gene were under purifying selection and no positively-selected amino acid site was identified. Strong selective constraint on the CP protein is probably attributed to the fact that it performs many different functions in the lifecycle of the virus, such as genome encapsidation, cell-to-cell movement, and plant-to-plant transmission (King et al. 2011). Interestingly, we found a difference in the selective constraints experienced by the two lineages of OrMV (Fig. 2, Table S4). In this case, the selective agents may be habitat differences between the two clades such as differences in the host species.

Geographic subdivision and host species contribute to the evolutionary dynamics of potyviruses, such as PVY, whose CP diversification was driven by both geographic and host-driven adaptations (Cuevas et al. 2012). In this study, the ML phylogenetic analysis did not show a clear geography-specific or host species specific clustering of OrMV possibly due to the occurrence of genetic exchange, but distinct genetic differences were discovered by AMOVA (Table 1) and BaTS (Table 2) analyses. The results of BaTS analyses provided evidence for host-specific clustering of OrMV isolates from the plant genera *Ornithogalum*, *Lachenalia* and *Diuri*, and we propose that to some extent, host-driven adaptation was responsible for the OrMV CP diversification. The possible role of geographic driven adaptation was not significant based on BaTs analyses at both large scale and finer scale. Therefore, geographic driven adaptation is not a major factor affecting OrMV CP diversification. Interestingly, a similar observation has been made for ArMV, a member of the genus *Nepovirus* of the subfamily *Comovirinae* within the family...
Secoviridae (Gao et al. 2016b). This suggests that OrMV and ArMV may share similar evolutionary mechanisms and that human activity has played a role in virus evolution because the introduction of ArMV and OrMV are more strictly controlled than that for PVY.

Conclusions
In summary, this study represents the first attempt to understand the molecular evolution of OrMV. We found evidence of selective constraints in OrMV evolution and its diversification was maintained partially by host-driven adaptation. However, isolates included in this analysis were relatively limited both in geography and host species. Further studies with larger, multiple-location and multiple-host-species sampling are needed to confirm our results and generalize the findings.

Acknowledgements
We thank Dr Sally Potter (Australian National University) for her comments and suggestions about the manuscript.

Reference
Balloux F, and Lugon-Moulin N. 2002. The estimation of population differentiation with microsatellite markers. *Molecular Ecology* 11:155-165.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, and Drummond AJ. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* 10:e1003537. 10.1371/journal.pcbi.1003537

Bouwen I, and von der Vlugt RAA. 1989. Natural Infection of *Alstroemeria caryophyllea* with *Ornithogalum mosaic virus*. *Plant Disease* 84:202-202.

Burger JT, Brand RJ, and Rybicki EP. 1990. The molecular cloning and nucleotide sequencing of the 3' terminal region of Ornithogalum mosaic virus. *Journal of General Virology* 71 ( Pt 11):2527-2534. DOI: 10.1099/0022-1317-71-11-2527

Burger JT, and von Wechmar MB. 1989. Purification and some properties of South African isolates of Ornithogalum mosaic virus. *Phytopathology* 79:385-391.

Chen CC, Chang YL, Chang CA, and Tsai HT. 2009. Molecular evidences for the dual infection of *Ornithogalum mosaic virus* and *Pterostylis virus Y* in an imported *Ornithogalum thyrsoides* plant with severe mosaic symptom. *Plant Pathology Bulletin* 9:184-185.

Cho SY, Lim S, Kim H, Yi SI, and Moon JS. 2016. Complete genome sequence of Ornithogalum mosaic virus Infecting *Gladiolus* spp. in South Korea. *Genome Announcements* 4:e00816-00816. DOI: 10.1128/genomeA.00816-16

Chung BY, Miller WA, Atkins JF, and Firth AE. 2008. An overlapping essential gene in the *Potyviridae. Proceedings of the National Academy of Sciences of the United States of America* 105:5897-5902. DOI: 10.1073/pnas.0800468105

Cuevas J, Delaunay A, Rupar M, Jacquot E, and Elena SF. 2012. Molecular evolution and phylogeography of *Potato virus Y* based on the CP gene. *Journal of General Virology* 93:2496-2501. DOI:10.1099/vir.0.044347-0

Cuevas JM, Willemsen A, Hillung J, Zwart MP, and Elena SF. 2015. Temporal dynamics of intrahost molecular evolution for a plant RNA virus. *Molecular Biology and Evolution* 32:1132-1147.
Delport W, Poon AF, Frost SD, and Kosakovsky Pond SL. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26:2455-2457. DOI: 10.1093/bioinformatics/btq429

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Resource* 32:1792-1797. DOI: 10.1093/nar/gkh340

Excoffier L, and Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resource* 10:564-567. DOI: 10.1111/j.1755-0998.2010.02847.x

Finseth FR, Bondra E, and Harrison RG. 2014. Selective constraint dominates the evolution of genes expressed in a novel reproductive gland. *Molecular Biology and Evolution* 31:3266-3281. DOI: 10.1093/molbev/msu259

Froissart R, Roze D, Uzest M, Galibert L, Blanc S, and Michalakis Y. 2005. Recombination every ay: Abundant recombination in a virus during a single multi-cellular host infection. *PLOS Biology* 3:e89. DOI: 10.1371/journal.pbio.0030089

Fuji S, Inoue M, Yamamoto H, Furuya H, Naito H, and Matsumoto T. 2003. Nucleotide sequences of the coat protein gene of potyviruses infecting *Ornithogalum thyrsoides*. *Archive of Virology* 148:613-621. DOI: 10.1007/s00705-002-0961-9

Gao F, Jin J, Zou W, Liao F, and Shen J. 2016a. Geographically driven adaptation of chilli veinal mottle virus revealed by genetic diversity analysis of the coat protein gene. *Archive of Virology* 161:1329-1333. DOI: 10.1007/s00705-015-2729-z

Gao F, Lin W, Shen J, and Liao F. 2016b. Genetic diversity and molecular evolution of Arabis mosaic virus based on the CP gene sequence. *Archive of Virology* 161:1047-1051. DOI: 10.1007/s00705-015-2729-z

Gao F, Zou W, Xie L, and Zhan J. 2017. Adaptive evolution and demographic history contribute to the divergent population genetic structure of *Potato virus Y* between China and Japan. *Evolutionary Applications* 10:379-390. DOI: 10.1111/eva.12459

Grisoni M, Moles M, Farreyrol K, Rassaby L, Davis R, and Pearson M. 2006. Identification of potyviruses infecting vanilla by direct sequencing of a short RT-PCR amplicon. *Plant Pathology* 55:523-529. DOI: 10.1111/j.1365-3059.2006.01397.x

Hudson RR. 2000. A new statistic for detecting genetic differentiation. *Genetics* 155:2011-2014.

Hudson RR, Boos DD, and Kaplan NL. 1992. A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution* 9:138-151.

Huson DH. 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14:68-73.

Kaur C, Raj SK, Snehi SK, Goel AK, and Roy RK. 2011. Natural occurrence of *Ornithogalum mosaic virus* newly reported on gladiolus in India. *New Disease Reports* 24:2.

King AM, Lefkowitz E, Adams MJ, and Carstens EB. 2011. Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses. *Amsterdam: Elsevier Academic Press*.

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, and Frost SD. 2006. GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22:3096-3098. DOI: 10.1093/bioinformatics/btl474

Lauring AS, and Andino R. 2010. Quasispecies theory and the behavior of RNA viruses. *PLOS Pathogens* 6:e1001005. DOI: 10.1371/journal.ppat.1001005

Liao F, Liao W, Chen X, Chen Q, Chen H, Huang P, Fang Z, Wu Y, Shen J, and Lin S. 2017. Molecular identification and sequence analysis of *Ornithogalum mosaic virus* in saffron (*Crocus sativus*) corms. *Scientia Agricultura*
Sinica 50:4046-4054.

Librado P, and Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452. DOI: 10.1093/bioinformatics/btp187

Martin DP, Murrell B, Golden M, Khoosal A, and Muhire B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evolution 1:vev003. DOI: 10.1093/ve/vev003

Moreno IM, Malpica JM, Diaz-Pendón JA, Moriones E, Fraile A, and García-Arenal F. 2004. Variability and genetic structure of the population of watermelon mosaic virus infecting melon in Spain. Virology 318:451-460. DOI: 10.1016/j.virol.2003.10.002

Moury B, Morel C, Johansen E, and Jacquemond M. 2002. Evidence for diversifying selection in Potato virus Y and in the coat protein of other potyviruses. Journal of General Virology 83:2563-2573.

Moya A, Elena SF, Bracho A, Miralles R, and Barrio E. 2000. The evolution of RNA viruses: A population genetics view. Proceedings of the National Academy of Sciences 97:6967-6973. DOI: 10.1073/pnas.97.13.6967

Nouri S, Arevalo R, Falk BW, and Groves RL. 2014. Genetic structure and molecular variability of Cucumber mosaic virus isolates in the United States. PLOS ONE 9:e96582. DOI: 10.1371/journal.pone.0096582

Nylander JAA. 2008. MrModeltest v2.3. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Ohshima K, Tomitaka Y, Wood JT, Minematsu Y, Kajiyama H, Tomimura K, and Gibbs AJ. 2007. Patterns of recombination in turnip mosaic virus genomic sequences indicate hotspots of recombination. Journal of General Virology 88:298-315. DOI: 10.1099/vir.0.82335-0

Pagán I, and Holmes EC. 2010. Long-term evolution of the Luteoviridae: time scale and mode of virus speciation. Journal of Virology 84:6177-6187. DOI: 10.1128/jvi.02160-09

Smith FF, and Brierley P. 1944. Ornithogalum mosaic. Phytopathology 34:497-503.

Storey JD, and Tibshirani R. 2003. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences 100:9440-9445. DOI: 10.1073/pnas.1530509100

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28:2731-2739. DOI: 10.1093/molbev/msr121

Wei T, Pearson MN, and Cohen D. 2006. First report of Ornithogalum mosaic virus and Ornithogalum virus 2 in New Zealand. New Disease Reports 55:820-820.

Wei T, Zhang C, Hong J, Xiong R, Kasschau KD, Zhou X, Carrington JC, and Wang A. 2010. Formation of complexes at plasmodesmata for Potyvirus intercellular movement is mediated by the viral protein P3N-PIPO. PLOS Pathogens 6:e1000962. DOI: 10.1371/journal.ppat.1000962

Wylie SJ, Li H, Dixon KW, Richards H, and Jones MG. 2013. Exotic and indigenous viruses infect wild populations and captive collections of temperate terrestrial orchids (Diuris species) in Australia. Virus Research 171:22-32. DOI: 10.1016/j.virusres.2012.10.003

Xia X. 2013. DAMBES5: A comprehensive software package for Data Analysis in Molecular Biology and Evolution. Molecular Biology and Evolution 30:1720-1728. DOI: 10.1093/molbev/mst064

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24:1586-1591. DOI: 10.1093/molbev/msm088

Yang Z, and Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Molecular Biology and Evolution 17:32-43.
Zeidan M, Cohen J, Watad A, and Gera A. 1998. Improved purification and molecular properties of Omithogalum mosaic virus in Israel. *Annals of Applied Biology* 133:167-176. DOI: 10.1111/j.1744-7348.1998.tb05817.x

**FIGURE LEGENDS**

**Figure 1** Phylogenetic networks of the CP gene from 36 OrMV isolates from different countries (A) and hosts (B). *Hippeastrum mosaic virus* (NC_017967) served as an outgroup. OrMV isolates from different countries or hosts are indicated by a unique color. Branch lengths are proportional to the genetic distances.

**Figure 2** Evolutionary relationship of the CP gene from the 35 recombination-free OrMV isolates and comparison of dN/dS values between the two clades. (A) ML phylogenetic tree showing genetic relationship among OrMV isolates. OrMV isolates from different regions (solid circle) and host species (diamond) are indicated by a unique color. Bootstrap percentage (BP≥50%) are indicated above major branches. The distance unit is substitutions/site; (B) Boxplots showing the dN/dS ratio of clade A (pink) and clade B (light blue) for the CP gene of OrMV.

**Figure S1** Sliding window plot of dN/dS values for CP gene. Sites under neutral (dN/dS = 1) are marked in orange dotted line and each phylogenetic clade is indicated with a unique color. The window size is 13 codons, and the offset between windows is one codon.
Figure 1 (on next page)

Phylogenetic networks of the CP gene from 36 OrMV isolates from different countries (A) and hosts (B).

*Hippeastrum mosaic virus* (NC_017967) served as an outgroup. OrMV isolates from different countries or hosts are indicated by a unique color. Branch lengths are proportional to the genetic distances.
Figure 2 (on next page)

Evolutionary relationship of the CP gene from the 35 recombination-free OrMV isolates and comparison of $dN/dS$ values between the two clades.

(A) ML phylogenetic tree showing genetic relationship among OrMV isolates. OrMV isolates from different regions (solid circle) and host species (diamond) are indicated by a unique color. Bootstrap percentage (BP≥50%) are indicated above major branches. The distance unit is substitutions/site; (B) Boxplots showing the $dN/dS$ ratio of clade A (pink) and clade B (light blue) for the CP gene of OrMV.
Table 1 (on next page)

Hierarchical Analysis of Molecular Variance for the effects of geography and host species.
Table 1. Hierarchical Analysis of Molecular Variance for the effects of geography and host species.

| Grouping factors | Source of variation | d.f. | Sum of squares | Variance components | Percentage of variation | Fixation index |
|------------------|---------------------|------|----------------|----------------------|-------------------------|----------------|
| Country          | Among groups        | 9    | 752.922        | 9.593                | 15.85                   | $\Phi_{ST} = 0.159^{***}$ |
|                  | Within groups       | 26   | 1323.717       | 50.912               | 84.15                   |                |
|                  | Total               | 35   | 2076.639       | 60.505               |                         |                |
| Host             | Among groups        | 9    | 939.994        | 18.492               | 29.73                   | $\Phi_{ST} = 0.297^{***}$ |
|                  | Within groups       | 26   | 1136.644       | 43.717               | 70.27                   |                |
|                  | Total               | 35   | 2076.639       | 62.209               |                         |                |

Significance thresholds: * $0.01 < P < 0.05$; ** $0.001 < P < 0.01$; *** $P < 0.001$;
Table 2 (on next page)

Results of Bayesian Tip-association significance (BaTS) testing for the geographical and host species on the genetic diversity of OrMV
Table 2. Results of Bayesian Tip-association significance (BaTS) testing for the geographical and host species on the genetic diversity of OrMV

| Analyses | Statistic         | $n$ | Observed Mean (95% HPD) | Null Mean (95% HPD) | $P$-value |
|----------|-------------------|-----|-------------------------|---------------------|----------|
| Region   |                   |     |                         |                     |          |
|          | AI                |     | 1.920 (1.492, 2.314)    | 3.094 (2.659, 3.464) | < 0.001*** |
|          | PS                | 18.785 (18.000, 19.000) | 23.989 (22.051, 25.597) | < 0.001*** |
| Asia     |                   | 17  | 2.235 (2.000, 3.000)    | 2.485 (1.760, 4.002) | 0.860 ns  |
|          | MC (Japan)        | 6   | 1.987 (2.000, 2.000)    | 1.281 (1.000, 2.000) | 0.080 ns  |
|          | MC (Israel)       | 1   | n/a                     | n/a                 |          |
|          | MC (China)        | 1   | n/a                     | n/a                 |          |
|          | MC (India)        | 8   | 2.224 (2.000, 3.000)    | 1.485 (1.000, 2.129) | 0.180 ns  |
|          | MC (South Korea)  | 1   | n/a                     | n/a                 |          |
| Oceania  |                   | 9   | 2.264 (1.000, 3.000)    | 1.545 (1.003, 2.000) | 0.190 ns  |
|          | MC (Australia)    | 5   | 2.178 (1.000, 3.000)    | 1.154 (1.000, 1.999) | 0.050*    |
|          | MC (New Zealand)  | 4   | 1.012 (1.000, 1.000)    | 1.114 (1.000, 1.767) | 1.000 ns  |
| Africa   |                   | 2   | 1.000 (1.000, 1.000)    | 1.014 (1.000, 1.021) | 1.000 ns  |
|          | MC (South Africa) | 2   | 1.000 (1.000, 1.000)    | 1.014 (1.000, 1.021) | 1.000 ns  |
| North America |       | 4   | 1.999 (2.000, 2.000)    | 1.158 (1.000, 2.000) | 0.090 ns  |
|          | MC (USA)          | 4   | 1.999 (2.000, 2.000)    | 1.158 (1.000, 2.000) | 0.090 ns  |
| Europe   |                   | 3   | 1.000 (1.000, 1.000)    | 1.080 (1.000, 1.767) | 1.000 ns  |
|          | MC (Netherlands)  | 3   | 1.000 (1.000, 1.000)    | 1.080 (1.000, 1.767) | 1.000 ns  |
| Host     |                   |     |                         |                     |          |
|          | AI                |     | 1.262 (0.896, 1.652)    | 2.981 (2.451, 3.449) | < 0.001*** |
|          | PS                | 13.007 (13.000, 13.000) | 22.136 (20.187, 23.975) | < 0.001*** |
|          | MC (Ornithogalum) | 10  | 4.000 (4.000, 4.000)    | 1.706 (1.001, 3.000) | 0.030*    |
|          | MC (Chenopodium)  | 1   | n/a                     | n/a                 |          |
|          | MC (Iris)         | 5   | 1.493 (1.000, 2.000)    | 1.224 (1.000, 1.996) | 1.000 ns  |
|          | MC (Narcissus)    | 1   | n/a                     | n/a                 |          |
|          | MC (Vanilla)      | 2   | 1.000 (1.000, 1.000)    | 1.033 (1.000, 1.004) | 1.000 ns  |
|          | MC (Lachenalia)   | 3   | 2.000 (2.000, 2.000)    | 1.046 (1.000, 1.028) | 0.020 *   |
|          | MC (Gladiolus)    | 8   | 2.224 (2.000, 3.000)    | 1.451 (1.000, 2.244) | 0.200 ns  |
|          | MC (Diuris)       | 3   | 2.178 (1.000, 3.000)    | 1.029 (1.000, 1.046) | 0.010 **  |
|          | MC (Crocosmia)    | 1   | n/a                     | n/a                 |          |
|          | MC (Crocus)       | 1   | n/a                     | n/a                 |          |

$AI$, association index; $PS$, parsimony score; $MC$, maximum monophyletic clade; HPD, highest probability density interval; n/a: no data available because of insufficient sample size ($n < 2$).

Significance thresholds: * $0.01 < p < .05$; ** $0.001 < p < 0.01$; *** $p < 0.001$.