A peer-reviewed version of this preprint was published in PeerJ on 27 March 2018.

View the peer-reviewed version (peerj.com/articles/4550), which is the preferred citable publication unless you specifically need to cite this preprint.

Genetic diversity analysis of the coat protein gene revealed strong evolutionary constraints on *Ornithogalum mosaic virus*

Fangluan Gao 1, Zhenguo Du 1, Jianguo Shen Corresp. 2, Hongkai Yang 1, Furong Liao Corresp. 3

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 a 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 $\omega$-values of 0.055 and 0.028, respectively, indicating a role of purifying selection in the differentiation of OrMV. In addition, the results from molecular variance of analysis (AMOVA) indicated that the effect of host species on the genetic divergence of OrMV is greater than that of geography. In BaTS analysis, OrMV isolates from the genera *Ornithogalum, Lachenalia, Diuri* tended to group together, indicating that OrMV diversification was maintained, in part, by host-driven adaptation. Furthermore, age calculations suggested that the first divergence event of the OrMV isolates analyzed might take place around 1068 BC.
Genetic diversity analysis of the coat protein gene revealed strong evolutionary constraints on *Ornithogalum mosaic virus*

Fangluan Gao¹, Zhenguo 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 a 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 $\omega$-values of 0.055 and 0.028, respectively, indicating a role of purifying selection in the differentiation of OrMV. In addition, the results from molecular variance of analysis (AMOVA) indicated that the effect of host species on the genetic divergence of OrMV is greater than that of geography. In BaTS analysis, OrMV isolates from the genera Ornithogalum, Lachenalia, Diuri tended to group together, indicating that OrMV diversification was maintained, in part, by host-driven adaptation. Furthermore, age calculations suggested that the first divergence event of the OrMV isolates analyzed might take place around 1068 BC.

Key words: Ornithogalum mosaic virus; phylogenetic analysis; selective constraints; 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 their management measures (Lauring & Andino 2010; Moya et al. 2000).

Ornithogalum mosaic virus (OrMV) is one of the most important 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 & van 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 with a characterized single-stranded, positive-sense RNA genome, encoding a single polypeptide 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 publically available from GenBank. In this study, these CP gene sequences were used to investigate the genetic diversity of OrMV and investigate the evolutionary forces responsible for the diversity. Moreover, the divergence time of the CP gene were analyzed to understand the process of emergence 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. 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 inferred using the Neighbor-Net method in SplitsTree 4.13.1 (Huson 1998). In contrast to traditional bifurcating phylogenetic trees, SplitsTree constructs recombination networks, illustrating the evolutionary relationships among taxa in the presence of recombination.

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 $S_{mt}$ were used to determine the presence of subdivision in populations (Hudson 2000; Hudson et al. 1992). Genetic differentiation among populations 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, 0.05 to 0.15; a large degree, 0.15 to 0.25; and a great degree, >0.25 (Balloux & Lugon-Moulin 2002). Besides, molecular variance of analysis (AMOVA) was conducted 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 $\phi$-statistics was tested based on 1023 permutations (default).

**Phylogenetic analysis and divergence time estimates**

After the potential recombinants were excluded, the phylogenetic relationships were reconstructed using the Maximum Likelihood (ML) 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 and the robustness of the ML tree topology was assessed with 1,000 bootstrap replicates.

To estimate OrMV divergence time, Bayesian coalescent analysis was performed using BEAST 2.4.6 (Bouckaert et al. 2014) with the same substitution model as described above. A Bayes factors (BF) test indicated that the relaxed uncorrelated lognormal was a better fit to the sequence data than the strict clock model. Therefore, the strict clock model was selected to estimate the molecular clock of CP gene. Owing to a lack of sufficient time stamp, a previous estimation of a mutation rate for the CP genes in potyviruses ($1.15 \times 10^{-4}$ nucleotide substitutions/site/year), was used to calibrate the analysis and a narrow prior on the clock rate was specified (Gibbs et al. 2008). The constant demographic model was considered as my tree prior compared with the exponential model. The MCMC was run for $5 \times 10^8$ generations to ensure convergence of all parameters. Acceptable sampling from the posterior and convergence to the stationary (ESS >200) were checked using Tracer 1.6. The trees were summarized into a maximum clade credibility tree (MCC) using TreeAnnotator 2.4.6, discarding the first 25% of samples as burn-in.

**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 used to compute three statistics of
phylogeny-trait association with the traits, 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 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 by means of 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 null model assuming that the entire tree has been evolving at the same rate (one-ratio model) and the 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 ω values (ω=dN/dS, synonymous to non-synonymous substitution rates) in the foreground branch than the background branches. Secondly, the site model was used identify nucleotide sites in CP-coding region that were likely to be involved in OrMV evolution. For the site model, seven codon substitution models described as M0, M1a, M2a, M3, M7, M8 and M8a, were investigated. The M1a model assumes two categories of sites (ω₀<1, ω₁=1), whereas the M2a model adds a third set of sites (ω₂>1) to M1a model. The M3 model, with three categories of sites, allows ω to vary among sites by defining a set number of discrete site categories, each with its own ω value. The M7 model partitions all the sites into ten different categories with ω<1 and fits a beta distribution to ω. In the M8 model, an 11th category is added to the M7 model allowing ω values >1, but ω is fixed to 1 for the 11th category of sites in the M8a model. For each nested model, the likelihood ratio test (LRT) was conducted by comparing twice the difference in log-likelihood values (2ΔLnL) against a χ²-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 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 > 2times 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, particularly for RNA 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 reticulated structure of phylogenetic network 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, which was higher than 0.500 and 0.005, indicating a high genetic diversity in OrMV populations and among subpopulations (Table 1a). Three independent tests of population differentiation were significant (Table 1b), 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 2). 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 divergence of OrMV based on the CP gene**

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 \( (P_{MC} > 0.05, \text{Table 3}) \). However, when the OrMV isolates were grouped by their host origins, the stronger signal was found since more host-specific clustering than expected by chance alone, particularly *Ornithogalum*, *Lachenalia* and *Diuris* \( (P_{MC} < 0.05, \text{Table 3}) \). The BaTS results indicated that OrMV CP diversification could be maintained in part by host-driven adaptation.

As inferred from the MCC trees (Fig. S1), the first divergence event of OrMV start at 3,081 (95%CI 494-8189) years ago, suggesting that the divergence time of the clade A and clade B OrMV isolates might take place around 1,068 BC when the last sampling date (known as 23 Nov, 2013) was calibrated. The last common ancestor of clade A and that of clade B were 1,087 (95% HPD: 855–1167) and 281 (95% HPD: 281–4495.78) years of 2013, respectively.
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 ($p < 0.05$, Table S3); hence, selective pressures differed between the two clades. The mean $\omega$ values for clades A and B were 0.055 and 0.028 (Table S3), respectively, indicating that clade B was subjected to stronger purifying selection than clade A. Furthermore, the results from pairwise analyses showed that there are differences between the distribution of $dN/dS$ values between clade A and clade B. Besides, sliding-windows analysis for sites under purifying selection detected in the site model was plotted in Fig.3. Although the $dN/dS$ values were blow 1.00 for both clades, the $dN/dS$ values of clade A were generally higher than those of clade B, indicating 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.

Discussion

RNA viruses tend to display high mutation rates compared to DNA viruses. However, quantifying their rates has proved difficult because they leave no fossil records. Alternatively, sampling times of the sequences that used to calibrate the molecular clock were not available in this study (Ho et al. 2011). Consequently, the estimated evolutionary rate of OrMV was calibrated by the rates reported previously in potyviruses (Gibbs et al. 2008). Our studies indicate that OrMV evolves at a rate of $2.61 \times 10^{-4}$ nucleotide substitutions/site/year (95%HPD ranging from $7.37 \times 10^{-5}$ to $4.42 \times 10^{-4}$), slightly higher than that of other potyviruses including Rice yellow mottle virus, which was estimated to be evolving at $1.40 \times 10^{-4}$ nucleotide substitutions/site/year, and Papaya ringspot virus, which has an estimated evolutionary rate of $1.25 \times 10^{-4}$ nucleotide substitutions/site/year (Gibbs et al. 2008).

Recombination plays an important role in the evolutionary 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 others 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. Unsurprisingly, purifying selection was detected at the majority of the polymorphic sites by two evolutionary analyses using the CODEML algorithm (Fig. 3), suggesting that most mutations in 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 observation 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 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 S3). In this case, the selective agents may be habitat differences between the two clades such as differences in the host species and geographic origins.

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, however, the diversification of OrMV isolates was affected, in part, by host-driven adaptation. The results from AMOVA and BaTS analyses (Table 2, Table 3) also provided evidence that the genetic variation of OrMV CP may be maintained partly by host-driven adaption. 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 share similar evolutionary mechanisms 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 samples are needed to confirm the results and generalize the findings.

Reference
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


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

**FIGURE LEGENDS**

**Figure 1** Phylogenetic networks of the CP gene from 36 OrMV isolates. *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** ML phylogeny of the CP gene of the OrMV isolates, information of which is given as isolate name/ GenBank accession. OrMV isolates from different regions and host species are indicated by a unique color. Bootstrap percentage (BP≥50%) are indicated above major branches. The distance unit is substitutions/site.

**Figure 3** Sliding window plot of $dN/dS$ values for CP gene. Sites under neutral ($dN/dS = 1$) are
marked in red 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 S1 Bayesian Maximum clade credibility (MCC) chronogram inferred from the CP gene of the recombination-free isolates of OrMV. The tree is scaled to time generated under the relaxed clock and constant demographic models. Posterior probabilities are shown at the major nodes.
Figure 1

Phylogenetic networks of the CP gene from 36 OrMV isolates. *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.
ML phylogeny of the CP gene of the OrMV isolates, information of which is given as isolate name/ GenBank accession. OrMV isolates from different regions and host species are indicated by a unique color. Bootstrap percentage (BP≥50%) are indicated above major branches. The distance unit is substitutions/site.
Figure 2

A. Phylogenetic tree showing the relationship between different BMV strains and their host origin. The tree is divided into two clades, Clade A and Clade B, with bootstrap support values shown for each node. The tree also includes information on the country and host origin of the strains.

B. gsi plot indicating the gsi values for Clade A and Clade B. The gsi values are shown with error bars, and the gsi values are compared to a reference gsi value of 0.0.

Country origin:
- New Zealand
- India
- Israel
- Netherlands
- South Africa
- South Korea
- USA
- Australia
- Japan
- China

Host origin:
- Ornithogalum
- Chenopodium
- Iris
- Narcissus
- Vanilla
- Lachenalia
- Gladiolus
- Diuris
- Crocosmia
- Crocus

Clade A: gsi = 0.055
Clade B: gsi = 0.028

**Note:** The gsi values indicate the degree of similarity between the BMV strains and their hosts, with higher gsi values indicating a greater degree of similarity. The gsi values are compared to a reference gsi value of 0.0 to determine the degree of similarity.
Figure 3

Sliding window plot of $dN/dS$ values for CP gene. Sites under neutral ($dN/dS = 1$) are marked in red dotted line and each phylogenic clade is indicated with a unique color. The window size is 13 codons, and the offset between windows is one codon.
Figure 3

dN/dS = 1

Site (Codon)

Table 1 (on next page)

Table 1
Table 1 (a) Genetic diversity parameters estimated for the CP gene of OrMV. (b) Summary of test statistics for population differentiation

<table>
<thead>
<tr>
<th>Phylogenetic Group</th>
<th>Sample size</th>
<th>Haplotypes</th>
<th>Haplotype diversity</th>
<th>Nucleotide diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>23</td>
<td>23</td>
<td>1.000 (±0.013)</td>
<td>0.106 (±0.014)</td>
</tr>
<tr>
<td>Group 2</td>
<td>13</td>
<td>13</td>
<td>1.000 (±0.030)</td>
<td>0.117 (±0.015)</td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>30</td>
<td>1.000 (±0.007)</td>
<td>0.156 (±0.010)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phylogenetic Group</th>
<th>$K_{ST}$</th>
<th>$K_S$</th>
<th>P-value</th>
<th>$S_{nn}$</th>
<th>P-value</th>
<th>$F_{ST}$</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1 vs. Group 2</td>
<td>0.296</td>
<td>83.545</td>
<td>&lt; 0.001***</td>
<td>1.00</td>
<td>&lt; 0.001***</td>
<td>0.470</td>
<td>&lt; 0.001***</td>
</tr>
</tbody>
</table>

Significance thresholds: *, 0.01 < P-value < 0.05; **, 0.001 < P-value < 0.01; ***, P-value < 0.001
Table 2

Hierarchical Analysis of Molecular Variance for the effects of geography and host species.
**Table 2.** Hierarchical Analysis of Molecular Variance for the effects of geography and host species.

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

Significance thresholds: * $0.01 < P < 0.05$; ** $0.001 < P < 0.01$; *** $P < 0.001$;
Table 3. Results of Bayesian Tip-association significance (BaTS) testing for the geographical and host species on the genetic diversity of OrMV

<table>
<thead>
<tr>
<th>Analyses</th>
<th>Statistic</th>
<th>Observed Mean (95% HPD)</th>
<th>Null Mean (95% HPD)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Country</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI</td>
<td>1.920 (1.492, 2.314)</td>
<td>3.094 (2.659, 3.464)</td>
<td>&lt; 0.001***</td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td>18.785 (18.000, 19.000)</td>
<td>23.989 (22.051, 25.597)</td>
<td>&lt; 0.001***</td>
<td></td>
</tr>
<tr>
<td>MC (Japan)</td>
<td>1.987 (2.000, 2.000)</td>
<td>1.281 (1.000, 2.000)</td>
<td>0.080 ns</td>
<td></td>
</tr>
<tr>
<td>MC (Israel)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>MC (Australia)</td>
<td>2.178 (1.000, 3.000)</td>
<td>1.154 (1.000, 1.999)</td>
<td>0.050*</td>
<td></td>
</tr>
<tr>
<td>MC (China)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>MC (India)</td>
<td>2.224 (2.000, 3.000)</td>
<td>1.485 (1.000, 2.129)</td>
<td>0.180 ns</td>
<td></td>
</tr>
<tr>
<td>MC (New Zealand)</td>
<td>1.012 (1.000, 1.000)</td>
<td>1.114 (1.000, 1.767)</td>
<td>1.000 ns</td>
<td></td>
</tr>
<tr>
<td>MC (South Africa)</td>
<td>1.000 (1.000, 1.000)</td>
<td>1.014 (1.000, 1.021)</td>
<td>1.000 ns</td>
<td></td>
</tr>
<tr>
<td>MC (USA)</td>
<td>1.999 (2.000, 2.000)</td>
<td>1.158 (1.000, 2.000)</td>
<td>0.090 ns</td>
<td></td>
</tr>
<tr>
<td>MC (Netherlands)</td>
<td>1.000 (1.000, 1.000)</td>
<td>1.080 (1.000, 1.767)</td>
<td>1.000 ns</td>
<td></td>
</tr>
</tbody>
</table>

| 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) | 4.000 (4.000, 4.000) | 1.706 (1.001, 3.000) | 0.030* |
| MC (Chenopodium) | n/a | n/a | n/a |
| MC (Iris) | 1.493 (1.000, 2.000) | 1.224 (1.000, 1.996) | 1.000 ns |
| MC (Narcissus) | n/a | n/a | n/a |
| MC (Vanilla) | 1.000 (1.000, 1.000) | 1.033 (1.000, 1.004) | 1.000 ns |
| MC (Lachenalia) | 2.000 (2.000, 2.000) | 1.046 (1.000, 1.028) | 0.020* |
| MC (Gladiolus) | 2.224 (2.000, 3.000) | 1.451 (1.000, 2.244) | 0.200 ns |
| MC (Diuris) | 2.178 (1.000, 3.000) | 1.029 (1.000, 1.046) | 0.010** |
| MC (Crocosmia) | n/a | n/a | n/a |
| MC (Crocus) | n/a | 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
Table 3

Table 3

Results of Bayesian Tip-association significance (BaTS) testing for the geographical and host species on the genetic diversity of OrMV
<table>
<thead>
<tr>
<th>Analyses</th>
<th>Statistic</th>
<th>Observed Mean (95% HPD)</th>
<th>Null Mean (95% HPD)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI</td>
<td></td>
<td>1.920 (1.492, 2.314)</td>
<td>3.094 (2.659, 3.464)</td>
<td>&lt; 0.001***</td>
</tr>
<tr>
<td>PS</td>
<td></td>
<td>18.785 (18.000, 19.000)</td>
<td>23.989 (22.051, 25.597)</td>
<td>&lt; 0.001***</td>
</tr>
<tr>
<td>MC (Japan)</td>
<td></td>
<td>1.987 (2.000, 2.000)</td>
<td>1.281 (1.000, 2.000)</td>
<td>0.080 ns</td>
</tr>
<tr>
<td>MC (Israel)</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>MC (Australia)</td>
<td></td>
<td>2.178 (1.000, 3.000)</td>
<td>1.154 (1.000, 1.999)</td>
<td>0.050*</td>
</tr>
<tr>
<td>MC (China)</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>MC (India)</td>
<td></td>
<td>2.224 (2.000, 3.000)</td>
<td>1.485 (1.000, 2.129)</td>
<td>0.180 ns</td>
</tr>
<tr>
<td>MC (New Zealand)</td>
<td></td>
<td>1.012 (1.000, 1.000)</td>
<td>1.114 (1.000, 1.767)</td>
<td>1.000 ns</td>
</tr>
<tr>
<td>MC (South Africa)</td>
<td></td>
<td>1.000 (1.000, 1.000)</td>
<td>1.014 (1.000, 1.021)</td>
<td>1.000 ns</td>
</tr>
<tr>
<td>MC (USA)</td>
<td></td>
<td>1.999 (2.000, 2.000)</td>
<td>1.158 (1.000, 2.000)</td>
<td>0.090 ns</td>
</tr>
<tr>
<td>MC (New Zealand)</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.080 (1.000, 1.767)</td>
<td>1.000 ns</td>
<td></td>
</tr>
<tr>
<td>Host</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI</td>
<td></td>
<td>1.262 (0.896, 1.652)</td>
<td>2.981 (2.451, 3.449)</td>
<td>&lt; 0.001***</td>
</tr>
<tr>
<td>PS</td>
<td></td>
<td>13.007 (13.000, 13.000)</td>
<td>22.136 (20.187, 23.975)</td>
<td>&lt; 0.001***</td>
</tr>
<tr>
<td>MC (Ornithogalum)</td>
<td></td>
<td>4.000 (4.000, 4.000)</td>
<td>1.706 (1.001, 3.000)</td>
<td>0.030*</td>
</tr>
<tr>
<td>MC (Chenopodium)</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>MC (Iris)</td>
<td></td>
<td>1.493 (1.000, 2.000)</td>
<td>1.224 (1.000, 1.996)</td>
<td>1.000 ns</td>
</tr>
<tr>
<td>MC (Narcissus)</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>MC (Vanilla)</td>
<td></td>
<td>1.000 (1.000, 1.000)</td>
<td>1.033 (1.000, 1.004)</td>
<td>1.000 ns</td>
</tr>
<tr>
<td>MC (Lachenalia)</td>
<td></td>
<td>2.000 (2.000, 2.000)</td>
<td>1.046 (1.000, 1.028)</td>
<td>0.020*</td>
</tr>
<tr>
<td>MC (Gladiolus)</td>
<td></td>
<td>2.224 (2.000, 3.000)</td>
<td>1.451 (1.000, 2.244)</td>
<td>0.200 ns</td>
</tr>
<tr>
<td>MC (Diuris)</td>
<td></td>
<td>2.178 (1.000, 3.000)</td>
<td>1.029 (1.000, 1.046)</td>
<td>0.010 **</td>
</tr>
<tr>
<td>MC (Crocosmia)</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>MC (Crocus)</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

*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.