*Syst. Biol.* 61(1):107–126, 2012

DOI:10.1093/sysbio/syr096
Advance Access publication on September 14, 2011

# Inferring Species Networks from Gene Trees in High-Polyploid North American and Hawaiian Violets (*Viola*, Violaceae)

THOMAS MARCUSSEN[1,2,*], KJETILL S. JAKOBSEN[1], JIŘÍ DANIHELKA[3,4], HARVEY E. BALLARD[5], KIM BLAXLAND[†], ANNE K. BRYSTING[1,2], AND BENGT OXELMAN[2]

[1]*Department of Biology, Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo, PO Box 1066 Blindern, NO-0316 Oslo, Norway;* [2]*Department of Plant and Environmental Sciences, University of Gothenburg, PO Box 461, 405 30 Gothenburg, Sweden;* [3]*Department of Botany & Zoology, Masaryk University, Kotlářská 2, CZ-611 37 Brno, Czech Republic;* [4]*Institute of Botany, Academy of Sciences of the Czech Republic, Lidická 25/27, CZ-657 20 Brno, Czech Republic; and* [5]*Department of Environmental and Plant Biology, Porter Hall, Ohio University, Athens, OH 45701, USA;*
**Correspondence to be sent to: Department of Plant and Environmental Sciences, University of Gothenburg, PO Box 461, 405 30 Gothenburg, Sweden; E-mail: thmsmrcssn@gmail.com.*
[†]*Deceased.*

*Received 1 October 2010; reviews returned 8 March 2011; accepted 27 June 2011*
*Associate Editor: Susanne Renner*

*Abstract*.—The phylogenies of allopolyploids take the shape of networks and cannot be adequately represented as bifurcating trees. Especially for high polyploids (i.e., organisms with more than six sets of nuclear chromosomes), the signatures of gene homoeolog loss, deep coalescence, and polyploidy may become confounded, with the result that gene trees may be congruent with more than one species network. Herein, we obtained the most parsimonious species network by objective comparison of competing scenarios involving polyploidization and homoeolog loss in a high-polyploid lineage of violets (*Viola*, Violaceae) mostly or entirely restricted to North America, Central America, or Hawaii. We amplified homoeologs of the low-copy nuclear gene, glucose-6-phosphate isomerase (*GPI*), by single-molecule polymerase chain reaction (PCR) and the chloroplast *trn*L-F region by conventional PCR for 51 species and subspecies. Topological incongruence among *GPI* homoeolog subclades, owing to deep coalescence and two instances of putative loss (or lack of detection) of homoeologs, were reconciled by applying the maximum tree topology for each subclade. The most parsimonious species network and the fossil-based calibration of the homoeolog tree favored monophyly of the high polyploids, which has resulted from allodecaploidization 9–14 Ma, involving sympatric ancestors from the extant *Viola* sections *Chamaemelanium* (diploid), *Plagiostigma* (paleotetraploid), and *Viola* (paleotetraploid). Although two of the high-polyploid lineages (*Boreali-Americanae*, *Pedatae*) remained decaploid, recurrent polyploidization with tetraploids of section *Plagiostigma* within the last 5 Ma has resulted in two 14-ploid lineages (*Mexicanae*, *Nosphinium*) and one 18-ploid lineage (*Langsdorffianae*). This implies a more complex phylogenetic and biogeographic origin of the Hawaiian violets (*Nosphinium*) than that previously inferred from rDNA data and illustrates the necessity of considering polyploidy in phylogenetic and biogeographic reconstruction. [Allopolyploidy; BEAST; homoeolog loss; low-copy nuclear gene; PADRE; single-molecule PCR; species network; *Viola*.]

Polyploidy, the condition where a genome consists of more than two chromosome sets, is a widespread and important evolutionary phenomenon in plants (e.g., Leitch and Leitch 2008). Interspecific hybridization in combination with genome duplications (allopolyploidy) can lead to instantaneous speciation by formation of a fertile novel species that is reproductively isolated from its parents. Polyploidization may account for up to 15% of all speciation events in angiosperms, and 35% in ferns (Wood et al. 2009), and there is growing evidence that practically all angiosperm lineages are ancient polyploids (Soltis et al. 2009). In vascular plants, infrageneric variation in ploidy level is common, and very high ploidy levels, such as 16-ploid or 18-ploid, have been reported from at least 20 angiosperm genera (e.g., Grant 1981; Elven 2007 onwards). Genome duplications also provide a copious source of gene duplication that opens up the possibility of sub- and neofunctionalization of the duplicated gene homoeologs (Adams and Wendel 2005).

Although the role of polyploidization in evolutionary processes has been generally acknowledged, molecular phylogenetic investigations in plants have been long dominated by the use of chloroplast DNA (cpDNA) and nuclear rDNA sequences. These markers are, however, unsuitable for reconstructing reticulate phylogenies, such as those generated by allopolyploidy, due to uniparental inheritance of cpDNA (Harris and Ingram 1991) and concerted evolution of rDNA (Wendel et al. 1995; Álvarez and Wendel 2003; Matyášek et al. 2007). This problem has been circumvented by inferring past events of polyploidy from incongruences in the rDNA and cpDNA phylogenies (e.g., McBreen and Lockhart 2006), but the usefulness of this approach is limited because other processes (e.g., allelic variation, gene duplication, horizontal gene transfer) can also lead to incongruence (Wendel and Doyle 1998). Furthermore, this approach will succeed in correctly identifying instances of hybridization only if rDNA and cpDNA represent substantially differentiated parental genomes, and furthermore, it cannot be used to trace successive polyploidization events where more than two genomes are involved (e.g., Popp et al. 2005; Brysting et al. 2011).

Low-copy nuclear genes are in general less prone to concerted evolution than rDNA and are therefore more likely to have conserved gene homoeologs for each of the ancestral genomes (Mort and Crawford 2004; Duarte et al. 2010). By individually sequencing these homoeologs in allopolyploids, using either specific primers, single-molecule (sm) polymerase chain reaction (PCR) (see Kraytsberg and Khrapko 2005), or *in vivo* cloning

(reviewed in Brysting et al. 2011), reticulate organism phylogenies can be untangled (e.g., Popp and Oxelman 2001; Sang 2002; Smedmark et al. 2003; Howarth and Baum 2005; Popp et al. 2005; Huber et al. 2006; Brysting et al. 2007; Popp and Oxelman 2007; Fortune et al. 2008; Kim et al. 2008; Mason-Gamer 2008; Mandáková et al. 2010; Marcussen et al. 2011; Brysting et al. 2011). The raw data are a set of multilabeled trees (or MUL trees), i.e., gene trees that contain more than one sequence for some of their included species as a result of gene duplication (paralogy) and/or polyploidy (homoeology), and these are then transformed into a species network. Huber et al. (2006) devised a method to derive a network that minimizes the number of hybridization nodes from a multilabeled tree, that is particularly suitable for allopolyploid networks. However, if some of the homoeologs have become extinct or remain undetected, or if the gene trees differ from the species or genome tree due to incomplete sorting events, the method will fail to recover the "true" network. Recent advances have introduced methods and software to overcome some of these problems (Lott et al. 2009), but a method utilizing optimality criteria is still wanting. Here, we explore extinction and hybridization on a single-gene phylogeny including species with genomes ranging from diploids to 18-ploids.

The cosmopolitan genus *Viola* (Violaceae), with 500–600 species of violets and pansies, comprises numerous hybrid and polyploid complexes in the northern hemisphere (Miyaji 1913; Moore and Harvey 1961; Clausen 1964; Fabijan et al. 1987; Ballard et al. 1998; Nordal and Jonsell 1998; van den Hof et al. 2008; Hepenstrick 2009; Marcussen et al., 2011). From a putative base number of $x = 6$ or $x = 7$, extant chromosome numbers range from dysploid $2n = 4$ in *V. modesta*, the lowest number known in angiosperms and also found in five other genera unrelated to *Viola* and Violaceae, to at least 20-ploid $2n =$ ca. 160 in *Viola arborescens* (Valentine et al. 1968; Erben 1996). The genus had its origin in South America (Clausen 1929; Ballard et al. 1998) and dispersed into the northern hemisphere in the Early Miocene (ca. 18 Ma), based on evidence from a number of Eurasian fossil seed morphotypes (Dorofeev 1963; Kovar-Eder et al. 2001; Mai 2001; Arbuzova 2005; Nikitin 2007).

The genus *Viola* is represented by about 100 species in continental North America (i.e., including Mexico and Central America). Based on morphology, ploidy, and chromosome number, these are usually classified into four diploid ($2x$) to tetraploid ($4x$) sections, in addition to five high-polyploid groups that have been given different taxonomic rank and placement (Becker 1925; Clausen 1929, 1964). The four sections present in North America, *Chamaemelanium* ($2x$), *Melanium* ($4x$), *Plagiostigma* ($4x$), and *Viola* ($4x$), each have broad distributions extending well beyond North America (Clausen 1964), with centers of diversity in western North America (section *Chamaemelanium*), East Asia (section *Plagiostigma*) and western Eurasia (section *Melanium*, section *Viola*). While the lowest chromosome number of section *Chamaemelanium* is diploid ($2n = 12$), the three sections *Melanium* (dysploid, $2n$ probably between 4 and 16; unpublished data), *Plagiostigma* ($2n = 24$), and *Viola* ($2n = 20$) have been shown to be allotetraploids between diploid species from the lineage of section *Chamaemelanium* (the CHAM lineage) and another, unidentified diploid lineage (the MELVIO lineage) (Marcussen et al. 2011). Secondary polyploidy has occurred internally in all four sections (e.g., Miyaji 1913, 1929; Clausen 1964; McPherson and Packer 1974; Fabijan et al. 1987; Erben 1996; van den Hof et al. 2008; Marcussen et al. 2010).

The five high-polyploid species groups (the main focus of this study), *Boreali-Americanae*, *Langsdorffianae*, *Mexicanae*, *Nosphinium*, and *Pedatae*, are mostly or entirely restricted to continental North America, Beringia, and the Hawaiian islands (Table 1). Each group is morphologically distinct and monophyletic as indicated by rDNA sequences (Ballard et al. 1998; Ballard and Sytsma 2000), and in most instances these groups are also allopatric and differ in chromosome number ($2n = 54$, 80, or 102). The respective phylogenetic placements and taxonomic ranks of these species groups have varied (Table 1; see survey in Ballard et al. 1998; Ballard and Sytsma 2000) and for this reason we herein treat the individual species groups as (taxonomically unranked) lineages. The *Boreali-Americanae* lineage ($2n = 54$) ranges throughout North America and is sympatric with the *Pedatae* lineage in eastern North America (both $2n = 54$). The former is a taxonomically difficult group with numerous closely related and interfertile species, whereas the latter is monotypic (*V. pedata*). The *Mexicanae* lineage ($2n = 80$) consists of 10 species, eight of which occur in Mexico and Central America and two disjunctly in northern South America. The *Langsdorffianae* lineage ($2n = 102$ [ca. 96, ca. 120]) comprises one (*V. langsdorffii*) or a few species in Beringia southward to northern Japan and coastal California, and the *Nosphinium* lineage ($2n = 80$) comprises nine species in the Hawaiian Islands. Owing to differences in key morphological characteristics, particularly in style shape, the high-polyploid lineages have been assigned to different sections of the genus. *Boreali-Americanae*, *Langsdorffianae*, and *Pedatae* were all included in section *Plagiostigma* in spite of their divergent chromosome numbers (Clausen 1964). The Hawaiian *Nosphinium* lineage, differing in style shape and in pronounced woodiness in several of the species, was given a section of its own (Becker 1925), though the woody species have sometimes been transferred to the predominately South American section *Leptidium* (St. John 1989). However, rDNA phylogenies (Ballard et al. 1998; Ballard and Sytsma 2000) suggest a close relationship of all the high polyploids with section *Viola*. Particularly the unexpected rDNA affinity of the Hawaiian *Nosphinium* lineage with the Beringian *Langsdorffianae* lineage, with Japanese and American exemplars paraphyletic with respect to monophyletic *Nosphinium*, led to the inference that the Hawaiian violets had been derived from within *Langsdorffianae* and had colonized the archipelago from its range in Beringia (Ballard and Sytsma 2000) no longer than 1.2−2.0 Ma (Havran et al.

TABLE 1. Infrageneric groups of *Viola*, sensu Clausen (1964) and Gershoy (1928), occurring in North America and Hawaii

| Taxonomic group[a] | 2n[a] and ploidy[b] | Distribution[a,c] | Spp.[a,d] | Morphology | | Flower color | Leaves | Taxonomic placement | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Stems | Stolons | | | Becker (1925) | Clausen (1964) |
| Sect. *Chamaemelanium* subsect. *Nudicaules* | $2n = 2x = 12$ | E. Asia, North America | 10 | + | − |  |  | Sect. *Chamaemelanium* grex *Erectae* b. *Nudicaules* | Sect. *Chamaemelanium* subsect. *Nudicaules* |
| Sect. *Viola* subsect. *Rostratae* | $2n = 4x = 20$ | North temperate | 50 | +(−) | −(+) |  |  | Sect. *Nomimium* greges *Rostratae* & *Repentes* | Sect. *Rostellatae* subsect. *Rostratae* |
| Sect. *Plagiostigma* grex *Primulifoliae* | $2n = 4x = 24$ | North America (Caribbean, N. South America) | 6 | − | + |  |  | Sect. *Nomimium Stolonosae*, p.p. | Sect. *Plagiostigma* subsect. *Stolonosae*, p.p. |
| Grex *Boreali-Americanae* | $2n = 10x = 54$ | E. North America (W. North America) | 11−17 | − | − |  |  | Sect. *Nomimium* grex *Boreali-Americanae* | Sect. *Plagiostigma* subsect. *Boreali-Americanae* |
| Grex *Pedatae* | $2n = 10x = 54$ | E. North America | 1 | − | − |  |  | Sect. *Nomimium* grex *Pedatae* | Sect. *Plagiostigma* subsect. *Pedatae* |
| Grex *Mexicanae* | $2n = 14x = 80$ | Mexico and Central America (South America) | 10 | − | + or − |  |  | Sect. *Nomimium* grex *Mexicanae* | — |
| Grex *Nosphinium* | $2n = 14x = 80$ | Hawaii | 9 | + | − |  |  | Sect. *Nosphinium* | — |
| Grex *Langsdorffianae* | $2n = 18x = 102$ (ca. 96, ca. 120) | Beringia (NE. Asia, W. North America) | 1 − 3 | + | − |  |  | Sect. *Nomimium* grex *Langsdorffianae* | Sect. *Plagiostigma* subsect. *Vaginatae* |

This table is available in black and white in print and in color at *Systematic Biology* online. [a]Data from Becker (1925), Gershoy (1928), Clausen (1929, 1964), Miyaji (1929), Skottsberg (1940), Sokolovskaya (1960, 1963), Taylor and Mulligan (1968), Carr (1978, 1985), Canne (1987), Nishikawa (1998), Ballard et al. (1998), Volkova et al. (2003), and Probatova et al. (2007).

[b]Ploidy levels determined herein (see Results and Discussion).

[c]Regions of secondary dispersal in brackets.

[d]Number of currently recognized species worldwide.

2009). An alternative hypothesis is that the phylogenetic relationship between the *Langsdorffianae* and *Nosphinium* lineages may be the result of parallel ancient allopolyploidization events in the two lineages involving some of the same parental species.

An introductory survey of *Viola* using isoenzymes indicated that ancient gene duplications in the high polyploids had been preserved for cytosolic glucose-6-phosphate isomerase (Gpi; EC 5.3.1.9). Therefore, its corresponding highly conserved (Grauvogel et al. 2007) low-copy gene (*GPI*) appeared to be promising for resolving the phylogeny of the high polyploids.

The aim of the present study was to resolve the origin of the high-polyploid violets in a phylogenetic perspective. We use DNA sequence data from homoeologs of the low-copy nuclear gene *GPI*, and from the chloroplast *trn*L-F region (*trn*L intron, and the *trn*L-F spacer). We determine the most parsimonious species network by evaluating different competing scenarios of events of deep coalescence, gene loss, and allopolyploidization. Timing of polyploidization events is estimated by calibrating the multilabeled phylogeny at 10 internal nodes with four dated fossils. Using this suite of tools, we obtain insights into polyploid origins as well as the evolution and pace of radiation in the older polyploid lineages, and we use the results to reevaluate the complex origin and relationships of the Hawaiian violets.

## Materials and Methods

### *Plant Material,* In vitro *Cloning of Homoeologs, PCR, and Sequencing*

A total of 58 accessions representing 51 species and subspecies was sequenced for the low-copy nuclear gene *GPI* (glucose-6-phosphate isomerase; Table 2). *Allexis batangae* served as the outgroup (Tokuoka 2008). Each of the five high-polyploid violet lineages was represented by one accession, except for *Boreali-Americanae*, which was represented by two species (*V. sagittata* and the reportedly chromosomally divergent *V. clauseniana*; Clausen 1964). All major morphological groups occurring in North America were sampled, with dense sampling especially within the lineages putatively sister to the high polyploids (see Results section). DNA was extracted using a CTAB extraction protocol (Doyle and Doyle 1987). In most cases, DNA working solutions were made by diluting extractions 1:20, of which 1 μL was used per PCR reaction. For "difficult" DNA preparations, the obtained stock DNA solution was further cleaned using the DNeasy Blood & Tissue Kit (Qiagen, Düsseldorf, Germany), following the manufacturer's guidelines except omitting the first two steps.

*Viola*-specific primers (Table 3) were designed for a locus corresponding to exon 12 to exon 18 of the *GPI* gene in *Arabidopsis* (AB007647, NM123638) and with a length of ca. 2000 base pairs. In order to increase the chance of discovering all homoeologs, the *GPI* locus was amplified in two separate PCR reactions covering exon 12 to exon 16 (PCR1) and exon 13 to exon 18 (PCR2), respectively. Different approaches were used to amplify homoeologs in diploids, tetraploids, and high polyploids (Table 3). For diploids, PCR1 and PCR2 were performed using a single set of general primers each, and for tetraploids (sections *Plagiostigma* and *Viola*) using specific primers for their two homoeologs, CHAM and MELVIO. For the high-polyploid species, homoeologs were isolated *in vitro* by single molecule (sm) PCR at limiting dilution, where DNA concentration was so low that most of the reactions (approximately 60%) by pure chance did not receive any template molecules at all and thus produced no PCR product (see Kraytsberg and Khrapko 2005). Under such conditions, the positive reactions were most likely to have been initiated by a single template molecule. The smPCR protocol was performed in three steps. First, heteroduplex DNA was removed from the DNA template working solutions by 1 min denaturation at 85 °C followed by renaturation by slow cooling to room temperature over approximately 30 min. Second, optimal dilution of the DNA template was estimated in a PCR trial on a dilution series consisting of a 192 μL master mix divided into 16 reactions, in which the DNA template in reaction$_{n+1}$ was diluted 2:3 compared with reaction$_n$; hence, reaction$_{16}$ was 657 times more diluted than reaction$_1$. Finally, the highest DNA template dilution yielding a product, often a 1:25 to 1:50 dilution compared with reaction$_1$, was considered optimal and used for smPCR. smPCR was typically conducted with 32 replicates, that is, 384 μL master mix divided into 32 tubes. Forty PCR cycles generally amplified strong smPCR products. PCR products were visualized by electrophoresis on 1% agarose gels. Successful PCR products were diluted 1:10 and cycle sequenced directly using PCR primers with the BigDye 3.1 sequencing Kit (Applied Biosystems, Foster City, CA), then processed on a 3730 ABI DNA analyser (Applied Biosystems).

### *Phylogenetic Reconstruction*

All sequence chromatograms were edited manually and subsequently aligned with MUSCLE, as implemented in Geneious version 5.3.5 under standard settings. Alignments were further optimized manually in order to ensure that all putatively homologous indels (i.e., having identical length and position, and >80% sequence similarity for putative insertions) were consistently aligned. Indel characters were coded by using Simple indel-coding (Simmons and Ochoterena 2000) as implemented in the SeqState software (Müller 2005), except for length variation in polynucleotide motifs (one poly-A and one poly-AT in *trn*L-F); these were not coded. The final *GPI* alignment was 3602 bp long and contained 290 coded indels, of which 684 and 116 were parsimony informative, respectively (postedited from a MUSCLE-generated alignment that was 3569 bp long and contained 307 indels, of which 707 and 116 were parsimony informative). The final *trn*L-F alignment was

TABLE 2. Plant material

| Taxon | Section—infrasectional group[a] | GenBank accession IDs | Site—collector[b], date—herbarium voucher ID (herbarium) |
|---|---|---|---|
| *Viola clauseniana* | HPP—*Boreali-Americanae* | *GPI*: JF767038, JF767039, JF767040, JF767041, JF767042, JF767043. *trn*L-F: JF767205 | USA, SW Utah — 2008 — TM707 (O) |
| *V. sagittata* var. *ovata* | HPP—*Boreali-Americanae* | *GPI*: JF767124, JF767125, JF767126, JF767127. *trn*L-F: JF767206 | USA, Pennsylvania, Lancaster Co. — KB & TM 11.05.2007 — TM724 (O) |
| *V. langsdorffii* | HPP—*Langsdorffianae* | *trn*L-F: JF767200 | Cultivated — Russian Federation, Moscow Botanical Garden to TM — TM617 (O) |
| *V. langsdorffii* | HPP—*Langsdorffianae* | *GPI*: JF767071, JF767072, JF767073, JF767074, JF767075, JF767076, JF767077. *trn*L-F: JF767201 | USA, Alaska, Juneau Co. — Dr. Miki to KB 1994 — TM618 (O) |
| *V. grahamii* | HPP—*Mexicanae* | *GPI*: JF767059, JF767060, JF767061, JF767062, JF767063, JF767064. *trn*L-F: JF767203 | México, Edo. Michoacán, Municipio de Quiroga — A.C. Cortés-Palomec 07.2000 — A-T1-KK (BHO) |
| *V. tracheliifolia* | HPP—*Nosphinium* | *GPI*: JF767137, JF767138, JF767139, JF767140, JF767141. *trn*L-F: JF767202 | USA, Hawaii, Oahu, Kuaokala — W. Takeuchi et al. 14.07.1984 — UC-1598182 (UC) |
| *V. pedata* | HPP—*Pedatae* | *GPI*: JF767105, JF767106, JF767107, JF767108. *trn*L-F: JF767204 | USA, Pennsylvania, Lancaster Co. — KB & TM 12.05.2007 — TM729 (O) |
| *V. congesta* | *Andinium* | *GPI*: JF767046. *trn*L-F: JF767154 | Chile, VII Región, Maule Valley, e. of Talca — KB 15.12.2007 — TM641 (O) |
| *V. pusilla* | *Andinium* | *GPI*: JF767119. *trn*L-F: JF767153 | Chile, II Región, Taltal — KB 15.09.2006 — TM749 (O) |
| *V. biflora* | *Chamaemelanium—Biflorae* | *GPI*: JF767023. *trn*L-F: JF767165 | Norway, Oppland, Sel — TM 27.07.2007 — TM775 (O) |
| *V. canadensis* | *Chamaemelanium—Canadenses* | *GPI*: JF767033, JF767034. *trn*L-F: JF767163 | Canada, Québec — Montréal Botanical Garden to TM — TM638 (O) |
| *V. sheltonii* | *Chamaemelanium—Chrysanthae* | *GPI*: JF767130. *trn*L-F: JF767159 | USA, California, Humboldt Co. — TM 27.05.2007 — TM757 (O) |
| *V. brevistipulata* | *Chamaemelanium—Nudicaules* | *GPI*: JF767032. *trn*L-F: JF767167 | Japan — KB to TM 05.2007 — TM745 (O) |
| *V. glabella* | *Chamaemelanium—Nudicaules* | *GPI*: JF767057, JF767058. *trn*L-F: JF767164 | USA, California, Humboldt Co. — TM 27.05.2007 — TM755 (O) |
| *V. lobata* | *Chamaemelanium—Nudicaules* | *GPI*: JF767080. *trn*L-F: JF767161 | USA, California, Nevada Co. — TM 29.05.2007 — TM762 (O) |
| *V. pubescens* | *Chamaemelanium—Nudicaules* | *GPI*: JF767117. *trn*L-F: JF767162 | Canada, Québec — Montréal Botanical Garden to TM — TM637 (O) |
| *V. uniflora* | *Chamaemelanium—Nudicaules* | *GPI*: JF767146, JF767147. *trn*L-F: JF767166 | Russian Federation, Altai Republic, Shebalino Distr. — JD et al. 2005/159 11.08.2005 — (BRNU 580353) |
| *V. purpurea* | *Chamaemelanium—Nuttallianae* | *GPI*: JF767118. *trn*L-F: JF767160 | USA, California, Shasta Co. — TM 28.05.2007 — TM758 (O) |
| *V. rotundifolia* | *Chamaemelanium—Orbiculares* | *GPI*: JF767122. *trn*L-F: JF767168 | USA, Pennsylvania, Lancaster Co. — KB & TM 12.05.2007 — TM734 (O) |
| *V. maculata* | *Chilenium* | *GPI*: JF767083, JF767084. *trn*L-F: JF767158 | Argentina, Santa Cruz, Rio Gallegos — KB 25.11.1999 — TM947 (O) |
| *V. arguta* | *Leptidium* | *GPI*: JF767017, JF767018. *trn*L-F: JF767155 | Ecuador, Zamora-Chinchipe Prov., Estación Cientifica San Francisco — HEB 05.12.2002 — HEB02-309 (BHO) |
| *V. bicolor* | *Melanium* | *GPI*: JF767019, JF767020, JF767021, JF767022. *trn*L-F: JF767176 | USA, New Jersey, Bridgeport — TM 14.05.2007 — TM743 (O) |
| *V. tuberifera* | *Plagiostigma* | *GPI*: JF767142, JF767143. *trn*L-F: JF767180 | China, Qinghai, s. of Aba — KB 20.07.2001 — TM948 (O) |
| *V. principis* | *Plagiostigma—Australasiaticae* | *GPI*: JF767115, JF767116. *trn*L-F: JF767183 | China, Sichuan, Wolong valley — KB 05.06.2007 — TM795 (O) |
| *V. verecunda* | *Plagiostigma—Bilobatae* | *GPI*: JF767150, JF767151. *trn*L-F: JF767181 | China, Yunnan, Dali, Kangshan — KB 30.05.1996 — TM697 (O) |
| *V. diffusa* | *Plagiostigma—Diffusae* | *GPI*: JF767047, JF767048. *trn*L-F: JF767177 | Japan — KB 1993 — TM711 (O) |
| *V. papuana* | *Plagiostigma—Diffusae* | *GPI*: JF767103, JF767104 | Unknown — Plant World Seeds to TM, as *Viola "papuanum"*, 2009 — TM942 (O) |

(*Continued*)

TABLE 2. (Continued)

| Taxon | Section—infrasectional group[a] | GenBank accession IDs | Site—collector[b], date—herbarium voucher ID (herbarium) |
|---|---|---|---|
| *V. selkirkii* | *Plagiostigma—Estolonosae* | *GPI*: JF767128, JF767129.*trn*L-F: JF767179 | Norway, Oppland, Sel — TM 21.07.2001 — TM698 (O) |
| *V. somchetica* | *Plagiostigma—Estolonosae* | *GPI*: JF767131, JF767132. *trn*L-F: JF767178 | Georgia, Mt. Kazbek — ex B&T World Seeds — TM636 (O) |
| *V. blanda* | *Plagiostigma—Stolonosae* | *GPI*: JF767024, JF767025, JF767026, JF767027. *trn*L-F: JF767190 | USA, Pennsylvania, Lancaster/Lebanon Co. — KB & TM 11.05.2007 — TM727 (O) |
| *V. blanda* | *Plagiostigma—Stolonosae* | *GPI*: JF767028, JF767029, JF767030, JF767031. *trn*L-F: JF767191 | USA, Pennsylvania, Lancaster Co. — KB & TM 12.05.2007 — TM735(O) |
| *V. epipsila* subsp. *epipsila* | *Plagiostigma—Stolonosae* | *GPI*: JF767049, JF767050. *trn*L-F: JF767184 | Norway, Akershus, Asker — TM 06.2005 — TM661 (O) |
| *V. epipsila* subsp. *repens* | *Plagiostigma—Stolonosae* | *GPI*: JF767051, JF767052 | Canada, Yukon — KB 02.07.2009 — TM926 (O) |
| *V. epipsila* subsp. *repens* | *Plagiostigma—Stolonosae* | *GPI*: JF767053, JF767054. *trn*L-F: JF767185 | Canada, Yukon — KB 01.07.2009 — TM927 (O) |
| *V. epipsila* subsp. *repens* | *Plagiostigma—Stolonosae* | *GPI*: JF767055, JF767056. *trn*L-F: JF767186 | USA, Alaska, Fairbanks Co. — KB 27.06.2009 — TM928 (O) |
| *V. palustris* | *Plagiostigma—Stolonosae* | *GPI*: JF767095, JF767096, JF767097, JF767098 | USA, Alaska, Juneau Co. — KB 10.05.2003 — TM721 (O) |
| *V. palustris* | *Plagiostigma—Stolonosae* | *GPI*: JF767099, JF767100, JF767101, JF767102. *trn*L-F: JF767187 | Norway, Oslo — TM 08.2009 — TM933 (O) |
| *V. renifolia* | *Plagiostigma—Stolonosae* | *GPI*: JF767120, JF767121. *trn*L-F: JF767188 | Canada, Ontario, Dorion — KB 2008 — TM932 (O) |
| *V. jalapaënsis* | *Plagiostigma—Stolonosae — Primulifoliae* | *GPI*: JF767067, JF767068. *trn*L-F: JF767197 | Mexico, Edo. Veracruz, near Jalapa — HB 21.06.1993 — HEB045 (BHO) |
| *V. lanceolata* | *Plagiostigma—Stolonosae — Primulifoliae* | *GPI*: JF767069, JF767070. *trn*L-F: JF767189 | USA, Pennsylvania, Carbon Co. — KB & TM 13.05.2007 — TM739 (O) |
| *V. macloskeyi* | *Plagiostigma—Stolonosae — Primulifoliae* | *GPI*: JF767081, JF767082. *trn*L-F: JF767198 | USA, California, Nevada Co. — TM 29.05.2007 — TM761 (O) |
| *V. occidentalis* | *Plagiostigma—Stolonosae — Primulifoliae* | *GPI*: JF767087, JF767088. *trn*L-F: JF767199 | USA, California, Del Norte Co. — TM 26.05.2007 — TM753 (O) |
| *V. pallens* | *Plagiostigma—Stolonosae — Primulifoliae* | *GPI*: JF767089, JF767090. *trn*L-F: JF767192 | USA, Pennsylvania, Carbon Co. — KB & TM 13.05.2007 — TM736 (O) |
| *V. pallens* | *Plagiostigma—Stolonosae — Primulifoliae* | *GPI*: JF767091, JF767092 | USA, Pennsylvania, Carbon Co. — KB & TM 13.05.2007 — TM737 (O) |
| *V. pallens* | *Plagiostigma—Stolonosae — Primulifoliae* | *GPI*: JF767093, JF767094. *trn*L-F: JF767193 | USA, Maine, Freeport — Arthur Haines to HEB 03.06.2002 — TM920 |
| *V. primulifolia* | *Plagiostigma—Stolonosae — Primulifoliae* | *GPI*: JF767109, JF767110. *trn*L-F: JF767194 | USA, Pennsylvania, Lancaster Co. — KB & TM 12.05.2007 — TM730 (O) |
| *V. primulifolia* | *Plagiostigma—Stolonosae — Primulifoliae* | *GPI*: JF767111, JF767112. *trn*L-F: JF767195 | USA, West Virginia — HEB 2000 — TM921 / HEB00-020 (BHO) |
| *V. primulifolia* | *Plagiostigma—Stolonosae — Primulifoliae* | *GPI*: JF767113, JF767114. *trn*L-F: JF767196 | USA, Texas — Ross McCauley to HEB 10.2000 — TM922 / HEB-v3 (BHO) |
| *V. vaginata* | *Plagiostigma—Vaginatae* | *GPI*: JF767148, JF767149. *trn*L-F: JF767182 | Japan, Honshu, Hiroshima Prefecture — KB 30.04.1997 — TM946 (O) |
| *V. capillaris* | *Rubellium* | *GPI*: JF767035. *trn*L-F: JF767156 | Chile, VII Región, Talca to Termas de Chillan. — KB 17.12.2007 — TM748 (O) |
| *V. rubella* | *Rubellium* | *GPI*: JF767123. *trn*L-F: JF767157 | Chile, XIV Región, Valdivia, near Mehuin — G. Knoche 13.01.2002 — TM671 (O) |
| *V. laricicola* | *Viola—Rostratae* | *GPI*: JF767078, JF767079. *trn*L-F: JF767175 | France, Hautes-Alpes — TM 05.2002 — TM509 (O) |
| *V. mirabilis* | *Viola—Rostratae* | *GPI*: JF767085, JF767086. *trn*L-F: JF767172 | France, Hautes-Alpes — TM 05.2006 — TM675 (O) |

(*Continued*)

TABLE 2. (Continued)

| Taxon | Section — infrasectional group[a] | GenBank accession IDs | Site—collector[b], date—herbarium voucher ID (herbarium) |
|---|---|---|---|
| *V. stagnina* var. *lacteoides* | *Viola—Rostratae* | *GPI*: JF767133, JF767134 van den Hof 270 (L) | The Netherlands: Kienveen, IJssel valley — van den Hof et al. 05.2008 — |
| *V. striata* | *Viola—Rostratae* | *GPI*: JF767135, JF767136. *trn*L-F: JF767173 | USA, Pennsylvania, Lancaster Co. — KB & TM 12.05.2007 — TM731 (O) |
| *V. uliginosa* | *Viola—Rostratae* | *GPI*: JF767144, JF767145. *trn*L-F: JF767174 | Sweden, Gästrikland — TM 05.2005 — TM662 (O) |
| *V. chelmea* | *Viola—Viola* | *GPI*: JF767036, JF767037. *trn*L-F: JF767169 | Greece, Evvia Island — TM 05.2001 — TM352 (O) |
| *V. collina* | *Viola—Viola* | *GPI*: JF767044, JF767045. *trn*L-F: JF767171 | Russian Federation, Bashkortostan Republic — M. Kočí et al. 2007/201 — (BRNU 590933) |
| *V. hirta* | *Viola—Viola* | *GPI*: JF767065, JF767066. *trn*L-F: JF767170 | France, Hautes-Alpes — TM 06.2006 — TM682 (O) |
| *Allexis batangae* | (outgroup) | *GPI*: JF767016 | Cameroon, 2°54′ N 9 °54′ E — Bos et al. 1969 — Bos4241 (UPS) |
| *Allexis cauliflora* | (outgroup) | *trn*L-F: AY739760 | |
| *Noisettia orchidiflora* | (outgroup) | *trn*L-F: JF767152 | Brasil, Linhares/ES. — Paula-Souza et al. 5695 (ESA) |

[a]Infrasectional taxon names, including those of the high polyploids (HPP), are not given taxonomic rank as at least some of the taxonomic combinations in current use, derived primarily from Becker (1925) and Clausen (1964), are not validly published.

[b]For material collected by the authors only author initials are given.

TABLE 3. Standard PCR and sequencing primers and annealing temperatures used

| Accessions | | PCR1 (exon12 – exon16) | PCR2 (exon13 – exon 18) |
|---|---|---|---|
| Diploids / high polyploids | PCR primers | Gpi_vex12F (TGCTTCCTYTGTCTCTCCA), Gpi_vex16R (TGRCCATTTGTTCCAGGTTC) | Gpi_vex13F (AGCACAKTCSATAGATCAGCAT), Gpi18R1 (GCRAAAAAGTTGGACATGAG) |
| | Annealing temperature | 59 °C | 58 °C |
| | Sequencing primers | Gpi_vex13R (CTGATCTATSGAMTGTGCTC), Gpi_vex15R (GGCRAATTTCTCCAGAGC) | Gpi_vex13F, Gpi16F (AGCATGGAAAGTAAYGGCAA) |
| Sect. *Viola*: CHAM homoeolog | PCR primers | Gpi_vex12F, Gpi_cham16R (CTACATTGAAATAGAATTATACAGC) | Gpi_vex13F, Gpi_cham17R (CAACTTCWTGAATCTAAATCTTG) |
| | Annealing temperature | 59 °C | 58 °C |
| | Sequencing primers | Gpi_cham13R (TTACTTCAGTCKATGATTATAACAGA), Gpi_cham15R (TAAGATGGCCTGTGAGCAC) | Gpi_cham13F (CAAGTATCGTGTGGAATTT), Gpi_cham16F (GAAAGTAATGGCAAGGGAGTT) |
| Sect. *Viola*: MELVIO homoeolog | PCR primers | Gpi_M12F (CTCTCCAATATGGTTTCTCCATT), Gpi_melvio16R (GAAGTGGTAGACCATCAATAGAT) | Gpi_vex13F, Gpi_melvio17R (AACTTMTKGAATCTAAAAYCCTC) |
| | Annealing temperature | 58 °C | 56 °C |
| | Sequencing primers | Gpi_melvio13R (TTAAAAAACCATAAAGTGTGCATTCC), Gpi_melvio15R (TAAGATGGCCTGTGAGCAT) | Gpi_melvio13F (GTCGTGTGGAATTTGCAGG), Gpi_melvio16F (GAAAGTAATGGCAAGGGAGTA) |
| Sect. *Plagiostigma*: CHAM homoeolog | PCR primers | Gpi_C12Fpcr (TCCAATATGGTTTCTCCATG), Gpi_C16Rpcr (AAGTGGTAGACCATCAATAGAA) | Gpi_C13Fpcr (CGACTTTAGGTAGATTAAAGTG), Gpi_cham17R |
| | Annealing temperature | 58 °C | 56 °C |
| | Sequencing primers | Gpi_C13Rseq (GCATACACATGCACTTATACC), Gpi_cham15R | Gpi_C13Fseq (TGTTTTCGTTTACTGTTAACATTCA), Gpi_cham16F |
| Sect. *Plagiostigma*: MELVIO homoeolog | PCR primers | Gpi_M12F, Gpi_melvio16R | Gpi_vex13F, Gpi_melvio17R |
| | Annealing temperature | 58 °C | 56 °C |
| | Sequencing primers | Gpi_melvio13R, Gpi_melvio15R | Gpi_melvio13F, Gpi_melvio16F |

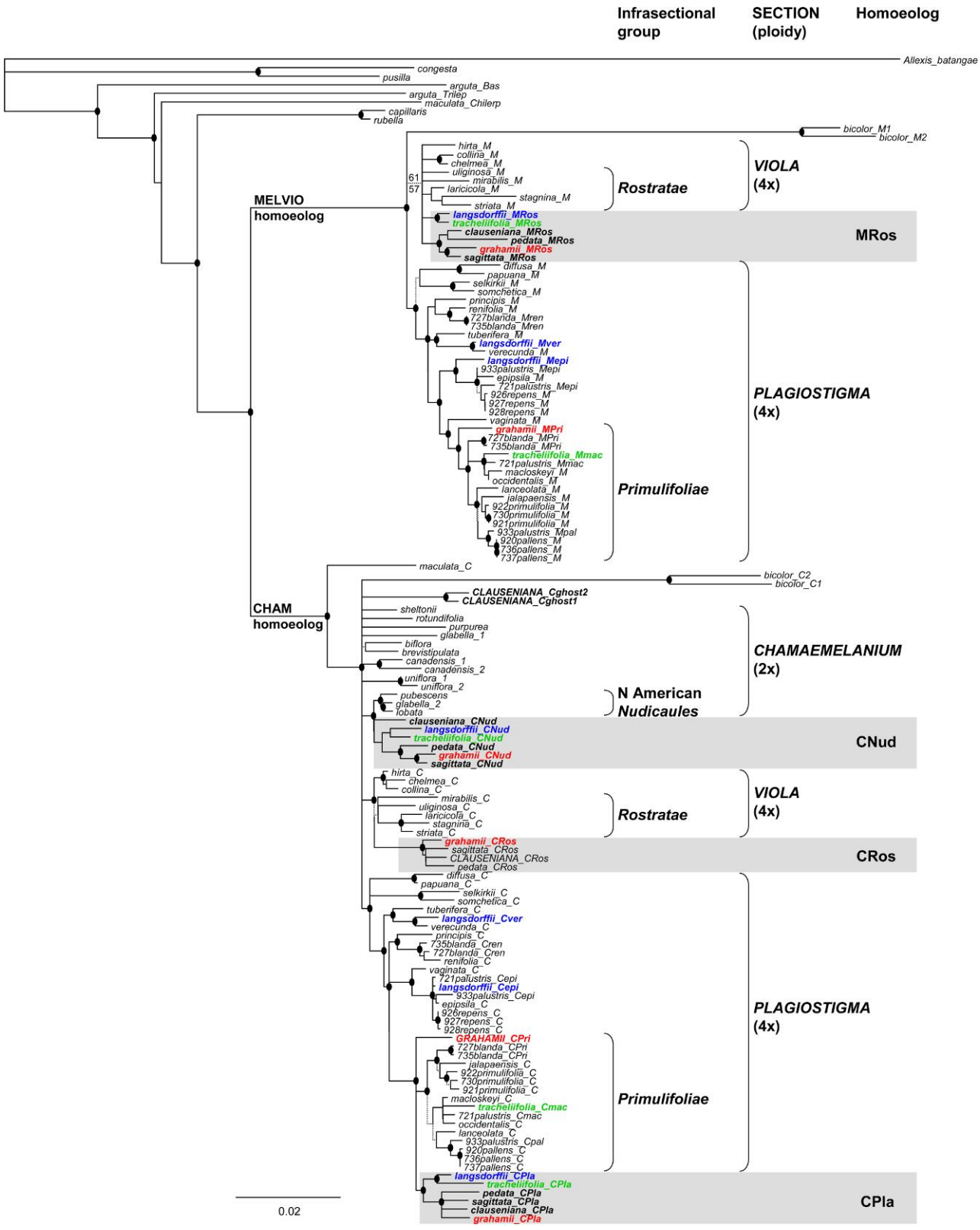
Notes: All primer sequences read from 5′ end to 3′ end. PCR mix: 25 μL reactions; 0.2 mM dNTPs, 0.25 μM of each of the primers, 1× Phusion HF buffer, 0.008 U/μL Phusion polymerase. The PCR conditions were as follows: initial denaturation at 95 °C for 30 s followed by 35 cycles of 95 °C for 120 s, annealing at a temperature specified for 30 s, and 72 °C for 30 s; the PCR ended with 7:30 min at 72 °C and subsequent soak at 10 °C.

1309 bp long and contained 68 coded indels, of which 74 and 15 were parsimony-informative, respectively (post-edited from a MUSCLE-generated alignment that was 1145 bp long and contained 75 indels, of which 85 and 28 were parsimony informative). The *GPI* and *trn*L-F phylogenies were constructed using maximum parsimony (MP) and maximum likelihood (ML). MP analysis was performed with Tree analysis using New Technology (TNT) version 1.1 (Goloboff et al. 2008) with "Traditional Search," Tree Bisection-Reconnection branch swapping, 10 replicates (sequence addition), and 10 trees saved per replication in effect. The consistency index (CI) and retention index (RI) were calculated. ML analysis was performed with Treefinder version of March 2008 (Jobb et al. 2004). Nucleotide substitution models for the exon and intron partitions were proposed by Treefinder based on the AICc model selection criterion. For *GPI*, three data partitions were defined and analyzed with different nucleotide substitution models, each with four rate categories: exons with HKY+Γ (A .284, C .198, G .207, T .311; TC = AG .354, TA = TG = CA = CG .073; alpha .366), introns with GTR+Γ (A .255, C .167, G .188, T .390; TC .256, TA .085, TG .099, CA .105, CG .130, AG .325; alpha 2.250), and coded indels with JC. The *trn*L-F data were analysed in the same way as *GPI* except that the entire nucleotide sequence was analysed with GTR+Γ (A .370, C .159, G .166, T .305; TC .331, TA .032, TG .146, CA .129, CG .120, AG .241; alpha .429) and, as for *GPI*, coded indels with JC. ML and MP bootstrapping were performed with the same settings as above using 1000 replicates. The *GPI* and *trn*L-F alignments and tree files are deposited at Dryad (doi:10.5061/dryad.68722) and TreeBASE (study number S11380).

### *Estimation of Divergence Times*

We calibrated the phylogeny with four *Viola* seed fossils using a Bayesian relaxed clock as implemented in BEAST 1.5.4 (Drummond et al. 2006; Drummond and Rambaut 2007), from a reduced data set containing only species of the CHAM and MELVIO clades; CHAM and MELVIO denote the homoeologs present in the tetraploid lineages (sections) *Melanium*, *Plagiostigma*, and *Viola* of the northern hemisphere (cf. Marcussen et al. 2010). The data matrix was partitioned with respect to exon, intron, and coded indels, using the substitution models HKY+Γ, GTR+Γ, and the simple binary substitution model, respectively. The analyses used a speciation model that followed a Yule tree prior, with rate variation across branches uncorrelated and lognormally distributed. One Markov chain Monte Carlo (MCMC) chain was run for 37.5 million generations, with parameters sampled every 1000 step. A visually determined burn-in of 4 million generations was discarded. Effective sample sizes for all estimated parameters and node ages were well above 200, as recommended. Additionally, two shorter MCMC chains were run for 5 million generations to control for similar convergence to the long chain; these were not used further.

Seeds of *Viola* are easy to identify to genus, owing to a characteristic transverse cellular pattern of the inner surface of the testa (Van der Burgh 1987). Comprehensive morphological study of seeds across the entire genus has not been published to date, although Gil-ad (1997) used seed micromorphology in part to recognize distinct species and infer hybridization in *Boreali-Americanae*. At least 15 *Viola* morphospecies are known from the Eurasian Neogene (Dorofeev 1963; Mai 2001; Arbuzova 2005; Nikitin 2007). At least some of these can easily be assigned to currently recognized infrageneric groups based on unique traits and, in certain instances, show affinities with particular extant species. Four seed fossils were used for calibration of the phylogeny. Calibration 1 (appearance of the genus in the northern hemisphere) is linked to the almost synchronous appearance of several *Viola* seed morphotypes in Lower Miocene sediments (Dorofeev 1963) and the basal polytomy in the CHAM and MELVIO subclades, indicating rapid radiation (see Results section). The oldest fossil flora containing violet seeds, from Austria, has been dated to 17–18 Ma (Kovar-Eder et al. 2001). We used a lognormal prior probability that the basal node in either of the CHAM and MELVIO subclades was at least 18 myr old, with a 95% confidence interval of 5 myr (offset = 18; log(mean) = log(SD) = 0.6). Calibration 2 corresponded to the age of subsection *Rostratae*, which from the Upper Miocene (13.7–5.3 Ma) appears with several seed morphotypes or species (Van der Burgh 1987; Geissert et al. 1990). The oldest and most accurately dated seed fossil, attributed to the extant "*V. canina*" on morphological grounds (the irregularly folded testa), has been described from western Germany (Van der Burgh 1987) and dated to 9–10 myr old (Schäfer et al. 2004). We applied a lognormal prior probability that subsection *Rostratae* (node basal to *V. laricicola*, *V. mirabilis*, *V. stagnina*, *V. striata*, *V. uliginosa*) is at least 10 myr old, with a 95% confidence interval of 2 myr (offset = 10; log(mean) = 0; log(SD) = 0.4). Calibration 3 (age of the Eurasian subsection *Viola*) is based on the occurrence of three fossil seed species (*Viola* "sp. 1," "sp. 2," "sp. 3") in Russian Pliocene (5.2–2.6 Ma) sediments (Arbuzova 2005). We used a lognormal prior probability that subsection *Viola* (node basal to *V. chelmea*, *V. collina*, *V. hirta*) is at least 5.2 myr old, with a 95% confidence interval of 2 myr (offset = 10; log(mean) = 0; log(SD) = 0.4). Calibration 4 corresponded to the appearance of the allo-octoploid *V. palustris* in Europe. European *V. palustris* is an allopolyploid of *V. epipsila* subsp. *epipsila* and *V. pallens* (data herein), and fossil seeds of this bog species, identifiable by their luster and splitting seed coats, are common in European Tertiary and Quartenary sediments (Van der Burgh 1983). Its oldest fossil, from Lower Pliocene (5.3–3.6 Ma) (Van der Burgh 1983, Arbuzova 2005), was used to constrain the *epipsila-palustris* node and the *pallens-palustris* node in each of CHAM and MELVIO subclades to being at least 3.6 myr old with identical lognormal prior probabilities (offset = 3.6; log(mean) = 0; log(SD) = 0.4). Finally, we put a normally distributed age constraint on the root height of 26 Ma (SD = 1.2), inferred from a


Infrasectional group
SECTION (ploidy)
Homoeolog
Allexis_batangae
congesta
pusilla
arguta_Bas
arguta_Trilep
maculata_Chilerp
capillaris
rubella
bicolor_M1
bicolor_M2
MELVIO homoeolog
61
57
hirta_M
collina_M
chelmea_M
uliginosa_M
mirabilis_M
laricicola_M
stagnina_M
striata_M
Rostratae
VIOLA (4x)
langsdorffii_MRos
trachelilifolia_MRos
clauseniana_MRos
pedata_MRos
grahamii_MRos
sagittata_MRos
MRos
diffusa_M
papuana_M
selkirkii_M
somchetica_M
principis_M
renifolia_M
727blanda_Mren
735blanda_Mren
tuberifera_M
langsdorffii_Mver
verecunda_M
langsdorffii_Mepi
933palustris_Mepi
epipsila_M
721palustris_Mepi
926repens_M
927repens_M
928repens_M
vaginata_M
PLAGIOSTIGMA (4x)
grahamii_MPri
727blanda_MPri
735blanda_MPri
trachelilifolia_Mmac
721palustris_Mmac
macloskeyi_M
occidentalis_M
lanceolata_M
jalapaensis_M
922primulifolia_M
730primulifolia_M
921primulifolia_M
933palustris_Mpal
920pallens_M
736pallens_M
737pallens_M
Primulifoliae
maculata_C
bicolor_C2
bicolor_C1
CLAUSENIANA_Cghost2
CLAUSENIANA_Cghost1
CHAM homoeolog
sheltonii
rotundifolia
purpurea
glabella_1
biflora
brevistipulata
canadensis_1
canadensis_2
uniflora_1
uniflora_2
pubescens
glabella_2
lobata
N American Nudicaules
CHAMAEMELANIUM (2x)
clauseniana_CNud
langsdorffii_CNud
trachelilifolia_CNud
pedata_CNud
grahamii_CNud
sagittata_CNud
CNud
hirta_C
chelmea_C
collina_C
mirabilis_C
uliginosa_C
laricicola_C
stagnina_C
striata_C
Rostratae
VIOLA (4x)
grahamii_CRos
sagittata_CRos
CLAUSENIANA_CRos
pedata_CRos
CRos
diffusa_C
papuana_C
selkirkii_C
somchetica_C
tuberifera_C
langsdorffii_Cver
verecunda_C
principis_C
735blanda_Cren
727blanda_Cren
renifolia_C
vaginata_C
721palustris_Cepi
langsdorffii_Cepi
933palustris_Cepi
epipsila_C
926repens_C
927repens_C
928repens_C
GRAHAMII_CPri
727blanda_CPri
735blanda_CPri
jalapaensis_C
922primulifolia_C
730primulifolia_C
921primulifolia_C
macloskeyi_C
trachelilifolia_Cmac
721palustris_Cmac
occidentalis_C
lanceolata_C
933palustris_Cpal
920pallens_C
736pallens_C
737pallens_C
Primulifoliae
PLAGIOSTIGMA (4x)
langsdorffii_CPla
trachelilifolia_CPla
pedata_CPla
sagittata_CPla
clauseniana_CPla
grahamii_CPla
CPla
0.02

comprehensive study of the genus (Marcussen et al. in preparation). The BEAST data files are available as online Appendix 1 (input .xml file) and online Appendix 2 (output .tree file) at Dryad (doi:10.5061/dryad.68722).

### *Reconstruction of the Most Parsimonious Allopolyploid Network*

We used the computer software PADRE (Lott et al. 2009) to construct allopolyploid species networks from the *GPI* multilabeled tree. As an input "species tree" topology for PADRE, we used the maximum tree (in the terminology of Liu et al. 2010) topology for *GPI*, reconciled using the youngest coalescent age estimates for each clade previously obtained from BEAST (Fig. 3). The maximum tree minimizes the number of deep coalescences and has been demonstrated to be a consistent estimator of the species tree (Liu et al. 2010). Hence, in the tetraploids, the maximum tree topology was reconciled based on coalescent ages in the CHAM and MELVIO subclades, and in the high polyploids this was done based on ages in the four homoeolog subclades CNud, CPla, CRos, and MRos.

A challenge in the reconstruction of allopolyploid networks is to correctly identify polyploidizations even if the associated homoeologs have become "lost", either due to gene deletion or our failure to detect them (e.g., due to primer mismatch). Two cases of putative homoeolog loss were observed in the *GPI* data: CRos in *V. langsdorffii* and *V. tracheliifolia*, and a hypothesized "MPla" in all the high polyploids (see Results section). To assess whether this absence was primary, that is, a result of the allopolyploid origin itself, or secondary, we generated four input tree files reflecting each of the four combinations of presence and absence of the CRos and "MPla" homoeologs. The four tree files were then analyzed separately in PADRE and the results were compared for the most parsimonious solution, that is, the one requiring the fewest polyploidizations and gene losses to explain the observed data. The four input tree files and resulting networks used for PADRE are available in online Appendix 3 at Dryad (doi:10.5061/dryad.68722).

## RESULTS

### *Phylogenetic Reconstruction of Homoeologs*

There is a general correspondence among chromosome number, ploidy level, and number of *GPI* homoeologs (Table 2). Diploids are mostly homozygous for *GPI*. In cases where more *GPI* alleles or homoeologs are found within the genome, each copy is either numbered (e.g., canadensis_1, canadensis_2), or given an appended letter code referring to the clade to which it belongs (C = CHAM clade, M = MELVIO clade). For the high polyploids, we have named the homoeologs resulting from the decaploidization (see below) CNud, CRos, CPla, and MRos, with reference to their closest sister clades, that is, CHAM-Nudicaules, CHAM-Rostratae, CHAM-Plagiostigma, and MELVIO-Rostratae, respectively (Fig. 1).

Results from the MP and ML analyses give congruent tree topologies for both *GPI* (Fig. 1) and *trn*L-F (Fig. 2). The *GPI* phylogenies (MP: 60 most parsimonious trees, CI = 0.74, RI = 0.92) confirm that the tetraploid sections *Plagiostigma* and *Viola* possess *GPI* homoeologs from two main clades, CHAM and MELVIO (Fig. 1), and the chloroplast *trn*L-F phylogenies (Fig. 2) (MP: 22 most parsimonious trees, CI = 0.84, RI = 0.86) correspond well to the CHAM clade in the *GPI* phylogeny. The CHAM and MELVIO subclade topologies are congruent (Fig. 1), with three exceptions: (i) sequences from species of section *Chamaemelanium* ($2x$) are only present in the CHAM clade, where they form a polytomy for *GPI* (Fig.1) and a more or less resolved clade for *trn*L-F (Fig. 2); (ii) sequences of *V. vaginata*, *V. jalapaënsis*, and *V. primulifolia* take well-supported but somewhat different positions in the CHAM and MELVIO subclades (Fig. 1); (iii) there is incomplete additivity of the CHAM and MELVIO subtree topologies for the high polyploids (see below; Fig. 1), suggesting either duplication or loss of homoeologs (or, alternatively, failure of detection). *GPI* homoeologs of the high polyploids formed four subclades, nested within clades containing sequences of the three sections *Chamaemelanium* (CNud), *Plagiostigma* (CPla), and *Viola* (CRos, MRos) (Fig. 1). Here the chloroplast is inherited from section *Plagiostigma*. Additional CHAM and MELVIO homoeologs are found in three of the high-polyploid species. The *Chamaemelanium*-derived homoeolog (CNud) is sister to a clade consisting of North American members of subsection *Nudicaules*. Of the section *Viola*-derived homoeologs, CRos is sister to a clade corresponding to the northern hemisphere subsection *Rostratae* s.lat., whereas the position of MRos was not resolved; CRos was not recovered in *V. langsdorffii* or in *V. tracheliifolia*. The homoeolog CPla, which is derived from the tetraploid section *Plagiostigma*, is sister to a large clade of North American species, the *Primulifoliae* lineage, but only the CHAM homoeolog is found in the high

←

FIGURE 1. Nuclear *GPI* ML bootstrap consensus tree for North American high-polyploid *Viola* based on 1000 bootstrap replicates. Branches receiving strong (≥ 80%) MP or ML bootstrap support are indicated with a terminal dot; weakly supported branches (50–65%) are indicated with broken lines. Bootstrap values are shown (MP above and ML below branch) for the MELVIO homoeolog of section *Viola*. Where applicable, number prefixes to taxon names distinguish accessions within species, and appended numbers or letter codes distinguish gene copies within an individual. Localization of the ancestral CHAM and MELVIO lineages is shown. Ploidy levels and names for sections and infrasectional taxa are shown only for clades that contain high-polyploid gene copies. *GPI* homologs for the high polyploids are indicated in bold. The four clades containing only high-polyploid homoeologs (CNud, CPla, CRos, MRos), each recovered as monophyletic in the MP consensus tree, are shaded with gray. Homoeologs for the three high-polyploid species that have additional homoeologs outside of these clades are shown in different colors (*V. grahamii*, *V. langsdorffii*, *V. tracheliifolia*). Pseudogenized homologs (in *V. clauseniana* and *V. grahamii*) are indicated in capital letters.

polyploids (Fig. 1). Three of the high polyploids possess further *Plagiostigma*-derived CHAM and MELVIO homoeologs, suggesting secondary polyploidy: *V. grahamii* harbors an additional *Primulifoliae* genome (CPri, MPri); *V. langsdorffii* contains additional genomes from *V. epipsila*-like and *V. verecunda*-like ancestors (Cepi, Mepi, Cver, Mver), and *V. tracheliifolia* possesses an additional genome from the *V. macloskeyi/occidentalis* complex (Cmac, Mmac). In all these cases, the chloroplast is inherited from the higher-ploid parent. *Viola clauseniana* has a deviant version of the *Nudicaules* homoeolog. Pseudogenized *GPI* homoeologs, identified by frameshifts and premature stop codons, are detected in the two high polyploids *V. clauseniana* (CRos) and *V. grahamii* (C_Pri). In addition, *V. clauseniana* possesses two pseudogenized *GPI* copies that are placed in the basal polytomy of the CHAM clade (Cghost1 and Cghost2); the higher frequency at which they are amplified using smPCR suggests that these occur in multiple copies in the genome and are paralogs rather than homoeologs.

In addition, the allo-octoploid origins of *V. blanda* and *V. palustris*, section *Plagiostigma*, are resolved. The two accessions of *V. blanda* (8*x*) both possess two *V. renifolia*-like (4*x*) *GPI* homoeologs (Cren, Mren) and the chloroplast sequence and two *GPI* homoeologs phylogenetically nested within the *Primulifoliae* lineage (4*x*) (Figs. 1 and 2). The two accessions of *V. palustris* (8*x*) are not monophyletic: the Alaskan sample (721 palustris) has homoeologs in common with the Pacific *V. epipsila* subsp. *repens* (4*x*; Cepi, Mepi) and *V. macloskeyi/occidentalis* (4*x*; Cmac, Mmac), whereas the Norwegian sample (933 palustris) shares homoeologs with the sympatric *V. epipsila* subsp. *epipsila* (4*x*; Cepi, Mepi) and northeastern North American *V. pallens* (4*x*; Cpall, Mpall), with which it also shares the chloroplast (Figs. 1 and 2).

### *Estimation of Divergence Times*

A chronogram with average divergence time estimates and 10 calibration points denoted is presented in Figure 3. The four clades consisting of sequences from high-polyploid species (CNud, CPla, CRos, MRos) coalesce with their lower-ploid sister clades at 11.5 Ma (CPla) to 13.0 Ma (CRos). This narrow time interval with overlapping credibility intervals (8.6–16.1 Ma) suggests a single allopolyploidization event. Dating polyploidization events in general is not trivial, but as alleles/orthologs can be older but not younger than their species (Doyle and Egan 2010), a maximum age for the polyploidization is given by the youngest stem age estimate, that is, 11.5 (9.4–13.6) Ma (CPla). This value may be close to the actual date of allopolyploidization since the individual stem lineages are relatively short (0.6–2.7 Ma). The deviant coalescent age of 18.1 Ma for two "extra" *GPI* copies in *V. clauseniana* (Cghost1 and Cghost2) lends further support (see above) to the hypothesis that these are paralogs, not homoeologs, and therefore irrelevant in the reconstruction of allopolyploid relationships. The youngest coalescent age of the "western" and "eastern" high-polyploid lineages (Fig. 4) is 8.4 (6.2–10.7) Ma (CPla). For the "western" high polyploids, the youngest coalescence time for *V. langsdorffii* and *V. tracheliifolia* is 3.9 (1.0–7.1) Ma (MRos). This value is close to the inferred minimum age of the secondary polyploidization of *V. tracheliifolia* with the *V. macloskeyi-occidentalis* lineage at 3.7 (1.9–5.5) Ma (Mmac), whereas those associated with *V. langsdorffii* and the clades of *V. epipsila* and *V. verecunda* are inconsistently younger, 1.31 Ma (poor node support) and 2.2 (0.5–4.3) Ma, respectively. Coalescence times suggest allopolyploid origins in the last 3–4 Ma also for the high-polyploid *V. grahamii* (3.2 [1.3–5.4] Ma) and for the *Plagiostigma* octoploids *V. blanda* (3.4 [1.2–6.1] Ma) and the polyphyletic *V. palustris* (fossil used as calibration).

### *Reconstruction of the Most Parsimonious Allopolyploid Network*

Our comparison of four scenarios reflecting all possible combinations of gene loss yielded a most parsimonious network that required 11 changes (2 gene losses and 9 polyploidizations), as compared with 12–15 changes in the other three scenarios (Table 4). The most parsimonious PADRE network implies loss of (or failure to detect) both CRos (in *V. langsdorffii* and *V. tracheliifolia*) and the hypothesized "MPla" (in all the high polyploids) and supports nine genome mergers in the ingroup (Table 4 and Fig. 4): (1) a tetraploidization basal to sections *Viola* + *Plagiostigma*, (2) a decaploidization basal to the high polyploids, individual allopolyploidizations marking the origins of (3) *V. glabella* (4*x*), (4 and 5) *V. palustris* (8*x*; two origins) and (6) *V. blanda* (8*x*), and three secondary allopolyploidizations within the high-polyploid clade marking the origins of (7) *V. grahamii* (14*x*), (8) *V. langsdorffii* (18*x*), and (9) *V. tracheliifolia* (14*x*). Two of these genome mergers (2 and 8), representing the origin of the decaploids and the origin of *V. langsdorffii* (18*x*), involve three lineages, and each can be inferred as two subsequent polyploidizations.

## DISCUSSION

### *Allopolyploid Origin and Ancestry of the High-Polyploid Violets*

We have shown that the high-polyploid violets mostly native to continental North America, Beringia, and Hawaii all possess genomes derived from three lower-ploid, sympatric lineages, corresponding to the North American lineage of section *Chamaemelanium* subsection *Nudicaules* (diploid), the *Primulifoliae* lineage within section *Plagiostigma* (tetraploid), and section *Viola* (tetraploid) (Figs. 1 and 4). The most parsimonious species network suggests that the high polyploids are monophyletic at the decaploid level (Fig. 4 and

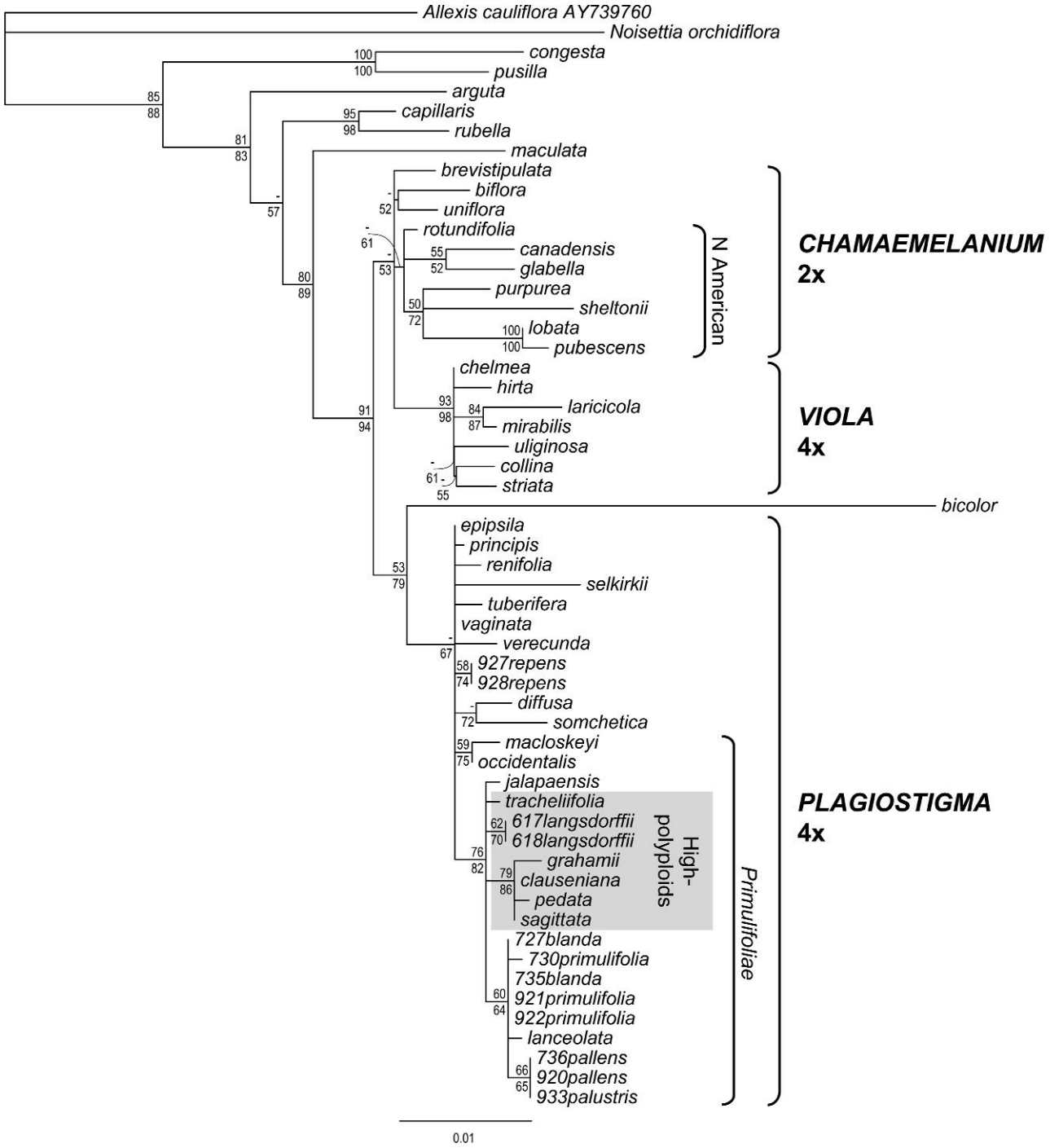
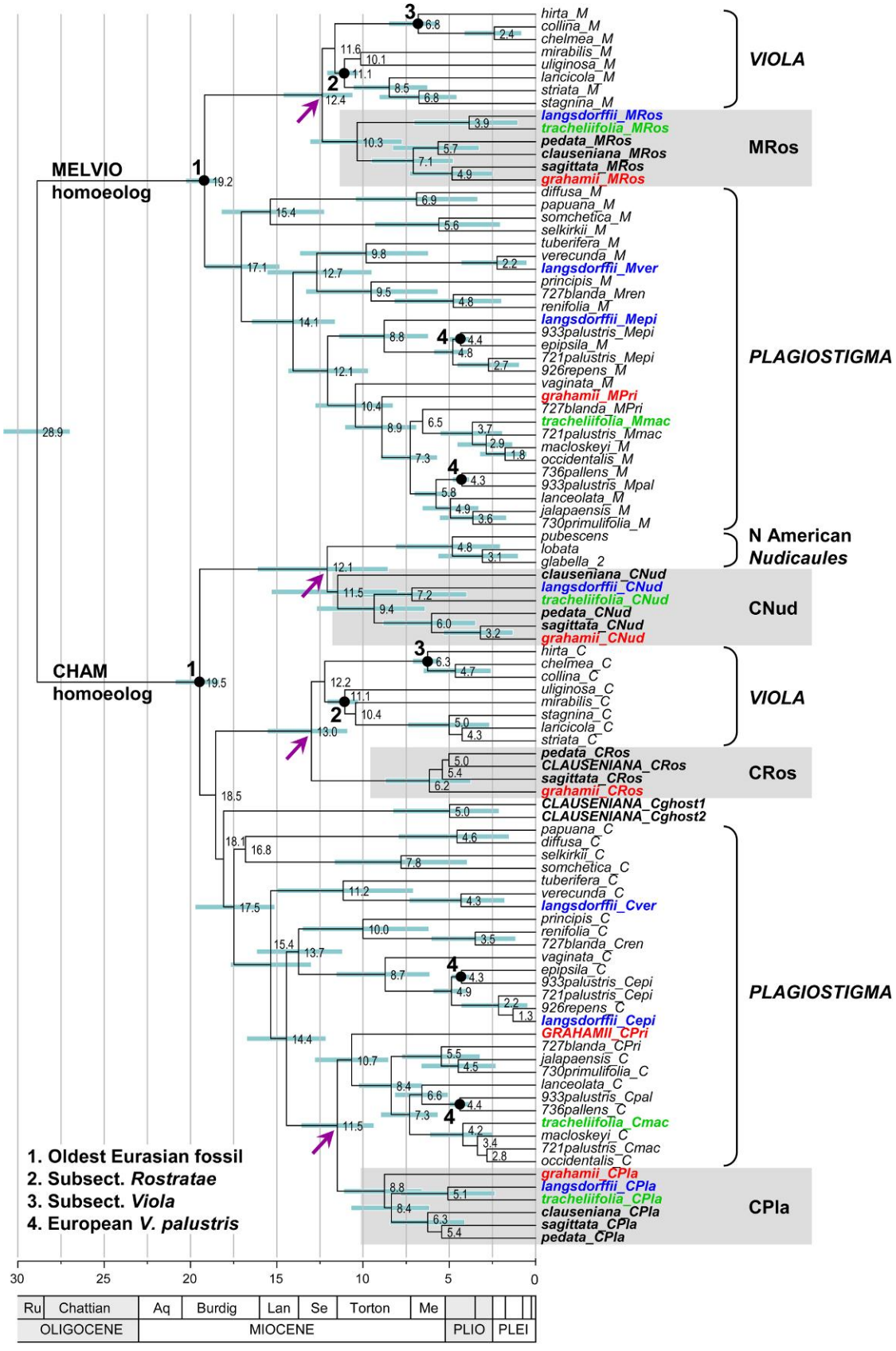



FIGURE 2. ML bootstrap consensus tree of the chloroplast *trn*L-F region for North American high-polyploid *Viola* based on 1000 bootstrap replicates. Bootstrap frequencies based on 1000 replicates are indicated above (MP) and below (ML) branches; branches indicated with a terminal dot receive bootstrap support $\geq$ 95% for MP and ML. Where applicable, number prefixes to taxon names distinguish accessions within species. Section names and the phylogenetic position of the high polyploids (shaded with gray) within section *Plagiostigma* are indicated.

Table 4). The decaploid level could only have been attained in two successive events of hybridization and genome duplication, but owing to the extinction of species at intermediate ploidy levels it is not possible to assess the sequence of individual hybrid combinations formed over the entire allopolyploidization process. In the following, we therefore refer to the decaploidization as a single event. The decaploids all have the *Plagiostigma* chloroplast (Fig. 2), showing that the *Plagiostigma* parent must have been the maternal parent, at least in the second hybridization. The occurrence of a distinct and deep-coalescent *Nudicaules*


MELVIO homoeolog
CHAM homoeolog
hirta_M
collina_M
chelmea_M
mirabilis_M
uliginosa_M
lancicola_M
striata_M
stagnina_M
VIOLA
langsdorffii_MRos
tracheliifolia_MRos
pedata_MRos
clauseniana_MRos
sagittata_MRos
grahamii_MRos
MRos
diffusa_M
papuana_M
somchetica_M
selkirkii_M
tuberifera_M
verecunda_M
langsdorffii_Mver
principis_M
727blanda_Mren
renifolia_M
langsdorffii_Mepi
933palustris_Mepi
epipsila_M
721palustris_Mepi
926repens_M
vaginata_M
grahamii_MPri
727blanda_MPri
tracheliifolia_Mmac
721palustris_Mmac
macloskeyi_M
occidentalis_M
736pallens_M
933palustris_Mpal
lanceolata_M
jalapaensis_M
730primulifolia_M
PLAGIOSTIGMA
pubescens
lobata
glabella_2
N American Nudicaules
clauseniana_CNud
langsdorffii_CNud
tracheliifolia_CNud
pedata_CNud
sagittata_CNud
grahamii_CNud
CNud
hirta_C
chelmea_C
collina_C
uliginosa_C
mirabilis_C
stagnina_C
lancicola_C
striata_C
VIOLA
pedata_CRos
CLAUSENIANA_CRos
sagittata_CRos
grahamii_CRos
CRos
CLAUSENIANA_Cghost1
CLAUSENIANA_Cghost2
papuana_C
diffusa_C
selkirkii_C
somchetica_C
tuberifera_C
verecunda_C
langsdorffii_Cver
principis_C
renifolia_C
727blanda_Cren
vaginata_C
epipsila_C
933palustris_Cepi
721palustris_Cepi
926repens_C
langsdorffii_Cepi
GRAHAMII_CPri
727blanda_CPri
jalapaensis_C
730primulifolia_C
lanceolata_C
933palustris_Cpal
736pallens_C
tracheliifolia_Cmac
macloskeyi_C
721palustris_Cmac
occidentalis_C
PLAGIOSTIGMA
grahamii_CPla
langsdorffii_CPla
tracheliifolia_CPla
clauseniana_CPla
sagittata_CPla
pedata_CPla
CPla
1. Oldest Eurasian fossil
2. Subsect. Rostratae
3. Subsect. Viola
4. European V. palustris
30 25 20 15 10 5 0
Ru Chattian Aq Burdig Lan Se Torton Me
OLIGOCENE MIOCENE PLIO PLEI

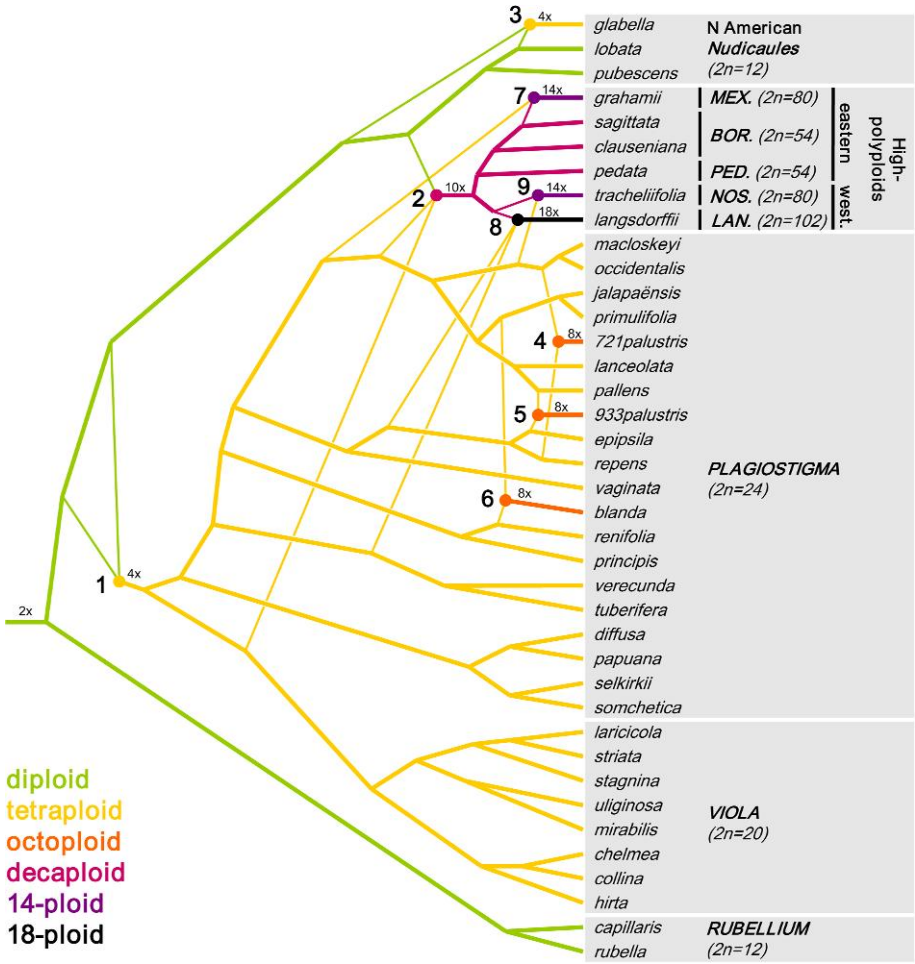



FIGURE 4. Most parsimonious PADRE reconstruction of allopolyploid relationships within North American *Viola*, which requires two homoeolog losses and 11 allopolyploidizations (Table 4). Genome mergers are numbered and shown as filled circles at line junctions, along with ploidy levels (2*x* to 18*x*): (1) tetraploidization basal to the sections *Plagiostigma* and *Viola* involving the diploid CHAM and MELVIO lineages; (2) decaploidization basal to the high polyploids, involving one diploid *Chamaemelanium* (*Nudicaules*) genome and one tetraploid genome from each of the sections *Plagiostigma* and *Viola*; (3) tetraploidization of *V. glabella*; (4–5) two independent octoploidizations of the diphyletic *V. palustris*; (6) octoploidization of *V. blanda*; (7) 14-ploidization of *V. grahamii* involving one decaploid and one tetraploid *Plagiostigma* genome; (8) 18-ploidization of *V. langsdorffii* involving one decaploid and two tetraploid *Plagiostigma* genomes; and (9) 14-ploidization of *V. tracheliifolia* involving one decaploid and one tetraploid *Plagiostigma* genome. Two mergers, (2) and (8), combine three lineages and thus each represent two subsequent polyploidization events whose order and lineage combinations remain unresolved. Chromosome numbers are shown for all higher infrageneric taxa (here unranked), and geographic affinity to western and eastern North America is indicated for the high polyploids.

allele in *V. clauseniana* (Fig. 1) suggests that this particular polyploidization happened more than once, which seems to be the rule in polyploids (e.g., Soltis and Soltis 1999). Although the *Boreali-Americanae* and *Pedatae* lineages have remained at the decaploid level, the *Mexicanae*, *Nosphinium*, and *Langsdorffianae* lineages are all products of additional independent polyploid events incorporating genomes from various lineages within section *Plagiostigma*—once in *Mexicanae* (14*x*), once in *Nosphinium* (14*x*), and twice in *Langsdorffianae* (18*x*) (Fig. 4). Our fossil-calibrated phylogeny (Fig. 3) shows that the decaploidization may have happened 9–14 Ma and the secondary polyploidizations, probably less than 3–4 Ma. This suggests that the initial decaploid differentiated substantially over a period of a few million years before the individual secondary polyploidizations that

←

FIGURE 3. Calibrated multilabeled chronogram based on Bayesian relaxed clock analysis of *GPI* sequence data for *Viola*, constrained with four fossils (1–4) at—owing to polyploidy—10 nodes. The high polyploids are indicated in bold, and the four homoeolog clades nested within lower-ploid ancestral lineages are shaded. The nodes basal to the four homoeolog clades, defining the maximum age of the decaploidization, are indicated with arrows. Genomes of *V. grahamii*, *V. langsdorffii*, and *V. tracheliifolia* are shown in color to indicate the secondary acquisition of tetraploid *Plagiostigma* genomes by polyploidization in these lineages. Pseudogenized homoeologs are indicated in capital letters. For explanation of homoeolog names, see Figure 1.

TABLE 4. Summary of the four PADRE analyses to assess whether the absence of two *GPI* homoeologs, CRos in *Viola langsdorffii* and *V. tracheliifolia* and a hypothesized "MPla" in all the high polyploids, is primary or due to (secondary) loss

| PADRE analyses[a] | Inferred mergers | Total inferred changes[b] |
|---|---|---|
| 1. No loss (i.e., primary absence) | 15 | 15 |
| 2. Loss of CRos | 14 | 15 |
| 3. Loss of "MPla" | 11 | 12 |
| 4. Loss of CRos and "MPla" | 9 | 11[c] |

[a]Online Appendix 3.
[b]Sum of the number of lineage fusions (i.e., polyploidizations) and the number of homoeolog losses.
[c]The most parsimonious network (i.e., the one implying the fewest changes), assuming independent loss of both *GPI* homoeologs (Fig. 3).

ultimately generated the ancestors of the five modern-day high-polyploid lineages.

Beyond compelling phylogenetic evidence from the *GPI* locus, the triple-hybrid decaploid hypothesis outlined above is corroborated by chromosome numbers. Given the lowest chromosome number for the sections *Chamaemelanium* ($2n = 2x = 12$), *Plagiostigma* ($2n = 4x = 24$), and *Viola* ($2n = 4x = 20$), their raw decaploid would be expected to have $2n = 56 (= 12 + 24 + 20)$, the 14-ploids $2n = 80 (= 56 + 24)$, and the 18-ploids $2n = 104 (= 56 + 24 + 24)$. This is indeed very close to the actual counts for these high-polyploid lineages, $2n = 54$, 80, and 102, respectively (Table 1). In *Langsdorffianae*, slightly deviating counts of $2n =$ ca. 96 (Miyaji 1929; Sokolovskaya 1960, 1963; Volkova et al. 2003; Probatova et al. 2007) and $2n =$ ca. 120 (Taylor and Mulligan 1968) presumably reflect partly the great difficulty in counting many small chromosomes, and partly the wish to align counts with multiples of $x = 12$, the base number attributed to *Langsdorffianae* by early authors (Miyaji 1929; Clausen 1964). Counts of $2n =$ ca. 60 and $2n =$ ca. 72 (Sokolovskaya and Probatova 1986) are difficult to interpret in light of our results and could be errors or counts made on hybrids of *V. langsdorffii* with, for example, tetraploids ($2n = 11x = 63$) and octoploids ($2n = 13x = 75$).

In spite of the long time since the decaploidization took place, at a time when the parental sections themselves had differentiated only for about 8 Ma, hybrids between members of the three parental sections can still be made artificially and, in certain cases, are vigorous, especially intersectional crosses involving *Rostratae* and *Plagiostigma* (Gershoy 1928, 1934). There are also reports of natural hybrids between *Boreali-Americanae* species and species of the other two lineages, but such hybrids are apparently rare (Russell 1955).

### *Diversification of the "Eastern" Decaploid Clade*

The four decaploid species of the "eastern" decaploid clade (*V. clauseniana*, *V. grahamii*, *V. pedata*, *V. sagittata*; Fig. 4) are currently distributed in most of North America but do not reach the Pacific coast. The individual *GPI* homoeolog subclades have incongruent topologies for these species (Fig. 1), but the reconciled maximum tree (Fig. 4), constructed from the minimum coalescent ages, places the *Pedatae* lineage as sister to the other two lineages and *Mexicanae* nested within *Boreali-Americanae*. The isolated position of *Pedatae* is not surprising: its single species, *V. pedata*, differs sharply in having deeply divided leaves, often differently colored petals, lack of cleistogamy, and is the only violet known to be (partially) self-incompatible (Becker and Ewart 1990). The second lineage, *Boreali-Americanae*, largely sympatric with the former, consists of numerous morphotypes and ecotypes that are variably distinct, variably sympatric, and variably interfertile (Clausen 1962; Gil-ad 1997). In our opinion, this reflects primary diversification rather than secondary breakdown of species barriers as a result of disturbance as previously suggested (Clausen 1962. This would fit with the relatively young age of the complex (5.9 Ma) and the apparently simple genetics of the few characters that separate taxa (Brainerd 1913, 1924). The *Mexicanae* lineage appears to have its origin in secondary allopolyploidization of a *Boreali-Americanae* species and a *Primulifoliae* species (section *Plagiostigma*) some 3.2 Ma. As previously mentioned, corresponding hybrids are not uncommon even among modern species (Russell 1955). The *Mexicanae* species are allopatrically distributed and particularly diverse in the high mountains of Mexico and Central America (but secondarily dispersed to northern South America).

### *Origin of the Hawaiian Violets Revisited*

Previous studies based on rDNA have hypothesized that the Hawaiian violets, the *Nosphinium* lineage, were derived from within the amphi-Beringian *Langsdorffianae* complex (Ballard and Sytsma 2000). Our low-copy nuclear gene data contradict such a scenario. For the lineages in question, herein represented by the "western" high polyploids *V. langsdorffii* and *V. tracheliifolia*, rDNA has apparently been homogenized toward the MELVIO-section *Viola* genome in the polyploid phylogeny (Fig. 5), which means that phylogenetic signals from the seven or eight other diploid genomes present in the allopolyploid were not captured. Indeed, the homoeolog lineages of *V. langsdorffii* and *V. tracheliifolia* continue to form well-supported clades also with *GPI*. The two lineages may have split only about 4 Ma at the time of their respective allopolyploidizations, with the *V. macloskeyi-occidentalis* clade (*V. tracheliifolia*) and with the clades of *V. epipsila* and *V. verecunda* (*V. langsdorffii*). Hence, the Hawaiian violets are not

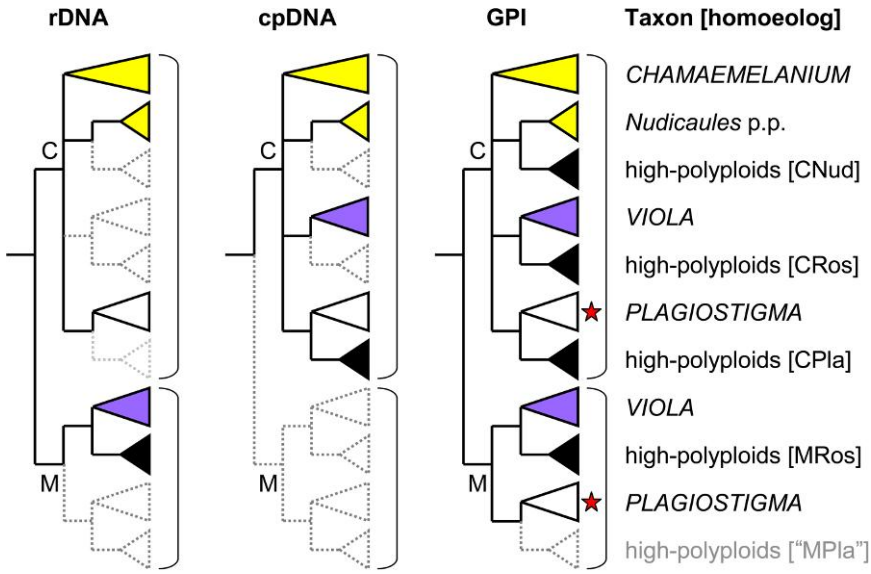



FIGURE 5. Simplified comparison of the inheritance of rDNA (ITS) (Ballard et al. 1998; Yoo et al. 2005), the chloroplast (cpDNA), and a low-copy nuclear gene (*GPI*) in *Viola* polyploids. Taxa refer to sections (capitalized) and one unranked infrasectional group (*Nudicaules*). The branch labels C and M refer to the CHAM clade and MELVIO clade, respectively. Some of the high-polyploids (Fig. 4) possess additional *GPI* homoeologs derived from section *Plagiostigma* by allopolyploidization (indicated with asterisks). This figure is available in black and white in print and in color at *Systematic Biology* online.

derived from within the *Langsdorffianae*; however, the two lineages are in one sense phylogenetic "sisters" in that they share a most recent common ancestor—the western decaploid—besides three additional tetraploid ancestors that they do not share.

Our data provide regional biogeographic evidence that the now exclusively Hawaiian *Nosphinium* lineage originated, by allopolyploidy, somewhere near the Pacific coast of North America, as this is where its extant "sister" species all occur (*V. macloskeyi*, *V. occidentalis*, *V. langsdorffii*). The fact that these species all have pronounced boreal affinities continues to lend support to Ballard and Sytsma's (2000) idea that the Hawaiian violets have an "Arctic origin." In a recent study, Havran et al. (2009) presented evidence that the ancestral *Nosphinium* violet first colonized the Maui Nui Complex of the Hawaiian archipelago. The emergence of this island, which later broke into four smaller islands (Maui, Moloka'i, Lāna'i, and Kaho'olawe), has been dated to 1.2–2.0 Ma (Price and Elliott-Fisk 2004) and thereby provides also a maximum age for the colonization event. However, the estimated age for the *Nosphinium* lineage based on our fossil-calibrated phylogeny is older, 3.7 (1.9–5.5) Ma, and the two time ranges are barely overlapping. Assuming these dates are accurate, this age discrepancy may be taken as evidence that the *Nosphinium* lineage persisted for some time on the American mainland, maybe up to 3–4 Ma, before dispersing to Hawaii. In Hawaii, it underwent rapid ecological and morphological radiation (Havran et al. 2009) while eventually becoming extinct on the mainland.

There is some evidence that polyploids are better adapted than diploids for establishment on oceanic islands (Harbaugh 2008; Soltis et al. 2009; Baldwin and Wagner 2010; see also Mummenhoff and Franzke 2007). Indeed, the highest incidence of polyploidy known is in the Hawaiian flora (Carr 1998). Hybridization and/or polyploidization shortly before dispersal to the Hawaiian Islands has been proposed for a number of angiosperm colonists outside of *Viola* (reviewed by Baldwin and Wagner 2010): hybrid origins may have aided their establishment or evolutionary success, by elevating genetic or genomic variation and potentially allowing for extensive recombination and expression of diverse phenotypes on which natural selection could act.

## CONCLUSIONS

Although polyploidy is of paramount phylogenetic importance in plant evolution, the way it has been approached traditionally in phylogenetic and biosystematic studies was often simplistic. Markers such as cpDNA and rDNA markers are de facto unsuitable for detecting reticulate evolutionary histories (e.g., Álvarez and Wendel 2003) and can lead to downstream misinterpretations of, for instance, historical biogeographic patterns, character evolution, and not least, biosystematics.

Herein, we demonstrated a multipronged approach to handle problems of polyploidy in phylogenetics: First, we amplified a low-copy nuclear gene (*GPI*) for which the duplications resulting from polyploidization were known to be (mostly) conserved. Second, we used *in vitro* cloning of gene homoeologs by smPCR rather than, for example, *in vivo* cloning that is less efficient when numerous gene copies are present. Third, we used fossil data to infer the ages of the different polyploidizations and for reconciling the maximum tree

topology (as a proxy for the species tree) from gene homoeolog trees with conflicting topology. Fourth, we applied an analytical tool (PADRE) for generation of species networks from multilabeled gene trees. Finally, we applied objective criteria to find the most parsimonious species network among competing scenarios of events of gene loss and polyploidization.

Using such a combined approach, we resolve both the ancient (9–14 myr old) common decaploid origin of a geographically confined but morphologically diverse high-polyploid lineage from three ancestral lower-ploid sections of *Viola* (*Chamaemelanium*, *Plagiostigma*, and *Viola*) and the more recent (less than 3–4 myr old) allopolyploidizations that contributed to the further diversification of the high polyploids and the formation of allopatric lineages at the 14- and 18-ploid levels.

The problems dealt with herein are of a general nature and may be encountered in numerous plant groups, as similarly high or even higher ploidy levels have been reported from more than 20 angiosperm genera (including *Cerastium*, *Curcuma*, *Draba*, *Fragaria*, *Fumaria*, *Papaver*, *Poa*, *Potentilla*, *Rumex*, *Saxifraga*, *Senecio*; e.g., Grant 1981; Elven 2007 onwards). Our approach highlights the need to consider phylogenetic trees at the appropriate level of organization (i.e., genome trees) in phylogenetic inference and, in particular, calls for the development of effective algorithms that can handle multilabeled trees and take coalescent and gene duplication/loss models into consideration for allopolyploids.

## Supplementary Material

Supplementary material, including data files and/or online-only appendices, can be found in the Dryad data repository (doi:10.5061/dryad.68722).

## Author Contributions

T.M. planned, designed, and led the project, collected plant material, did the laboratory work, ran all analyses, and wrote most of the text. All authors have contributed in the preparation of the study, and commented on and approved of the final manuscript. Particular contributions include suggestions on analysis designs (B.O., K.S.J.), information about fossils and chromosome counts from, mainly, Russian sources (J.D.), and collection of material and expertise with North American violet groups (K.B., H.E.B.).

## Funding

This work was supported by the Norwegian Research Council (170832: "Allopolyploid evolution in plants: patterns and processes within the genus *Viola*" to T.M. and K.S.J.); the Swedish Research Council (2009-5202 to B.O.); the Ministry of Education of the Czech Republic (MSM0021622416 and LC06073 to J.D.); and the Academy of Sciences of the Czech Republic (AV0Z60050516 to J.D.).

## Acknowledgements

In memory of Kim Blaxland. We wish to thank R. DeBry, S. Renner, M. Bendiksby, and two anonymous referees for suggesting valuable improvements of the paper. The curator of EPS is thanked for providing a rare sample of *Viola tracheliifolia* for DNA extraction, A. C. Cortés-Palomec for material of *V. grahamii*, G. Knoche for material of *V. rubella*, A. Skog and F.-I. Arias Sánchez for doing part of the molecular work, and Z. Kvaček and A. Hval' for providing us with advice and literature on palebotanical matters. Analyses in TreeFinder were carried out on the freely available Bioportal (www.bioportal.uio.no).

## References

Adams K.L., Wendel J.F. 2005. Polyploidy and genome evolution in plants. Curr. Opin. Plant Biol. 8:135–141.

Álvarez I., Wendel J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29:417–434.

Arbuzova O. 2005. *Viola* L. In: Budantsev L., editor. Iskopaemye tsvetkovye rastenija Rossii i sopredel'nyh gosudarstv [Fossil flowering plants of Russia and adjacent states]. Vol. 4. Nyctaginaceae – Salicaceae. Saint Petersburg (Russia): Isskustvo-SPB. p. 48–52 [in Russian].

Baldwin B.G., Wagner W.L. 2010. Hawaiian angiosperm radiations of North American origin. Ann. Bot. 105:849–879.

Ballard H.E., Sytsma K.J., Kowal R.R. 1998. Shrinking the violets: Phylogenetic relationships of infrageneric groups in *Viola* (Violaceae) based on internal transcribed spacer DNA sequences. Syst. Bot. 23:439–458.

Ballard H.E., Sytsma K.J. 2000. Evolution and biogeography of the woody Hawaiian violets (*Viola*, Violaceae): Arctic origins, herbaceous ancestry and bird dispersal. Evolution. 54:1521–1532.

Becker W. 1925. *Viola* L. In: Engler A., editor. Die natürlichen Pflanzenfamilien. Vol. 21. Parietales und Opuntiales. Leipzig (Germany): Wilhelm Engelmann. p. 363–376.

Becker W.A., Ewart L.C. 1990. Pollination, seed set and pollen tube growth investigations in *Viola pedata* L. Acta Hort. 272:33–36.

Brainerd E. 1913. Four hybrids of *Viola pedatifida*. Bull. Torrey Bot. Club. 40:249–260.

Brainerd E. 1924. Some natural violet hybrids of North America. Bull. Vermont Agric. Exp. Stn. 239:1–255.

Brysting A.K., Oxelman B., Huber K.T., Moulton V., Brochmann C. 2007. Untangling complex histories of genome mergings in high polyploids. Syst. Biol. 56:467–476.

Brysting A.K., Mathiesen C., Marcussen T. 2011. Challenges in polyploid phylogenetic reconstruction: a case story from the arctic-alpine *Cerastium alpinum* complex. Taxon. 60:333–347.

Canne J.M. 1987. Determinations of chromosome numbers in *Viola* (Violaceae). Can. J. Bot. 65:653–655.

Carr G.D. 1978. Chromosome numbers of Hawaiian flowering plants and the significance of cytology in selected taxa. Am. J. Bot. 65: 236–242.

Carr G.D. 1985. Additional chromosome numbers of Hawaiian flowering plants. Pac. Sci. 39:302–306.

Carr G.D. 1998. Chromosome evolution and speciation in Hawaiian flowering plants. In: Stuessy T.F., Ono M., editors. Evolution and speciation of island plants. Cambridge (UK): Cambridge University Press. p. 5–47.

Clausen J. 1929. Chromosome number and relationship of some North American species of *Viola*. Ann. Bot. 63:741–764.

Clausen J. 1962. Stages in the evolution of plant species. New York: Hafner Publishing Company.

Clausen J. 1964. Cytotaxonomy and distributional ecology of western North American violets. *Madroño*. 17:173–197.

Dorofeev P.I. 1963. Tretichnye flory zapadnoi Sibiri (The Tertiary floras of western Siberia). Moscow: Izdatel'stvo Akademii Nauk SSSR.

Doyle J.J., Doyle J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19:11–15.
Doyle J.J., Egan A.N. 2010. Dating the origins of polyploidy events. New Phytol. 186:73–85.
Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88.
Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.
Duarte J.L., Wall P.K., Edger P.P., Landherr L.L., Ma H., Pires J.C., Leebens-Mack J., dePamphilis C.W. 2010. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. BMC Evol. Biol. 10:61.
Elven R., editor. 2007 onwards. Checklist of the Panarctic Flora (PAF) Vascular Plants. Version: May 2007. http://www.binran.ru/infsys/paflist/index.htm.
Erben M. 1996. The significance of hybridization on the forming of species in the genus *Viola*. Bocconea. 5:113–118.
Fabijan D.M., Packer J.G., Denford K.E. 1987. The taxonomy of the *Viola nuttallii* complex. Can. J. Bot. 65:2562–2580.
Fortune P.M., Pourtau N., Viron N., Ainouche M.L. 2008. Molecular phylogeny and reticulate origins of the polyploid *Bromus* species from section *Genea* (Poaceae). Am. J. Bot. 95:454–464.
Geissert F., Gregor H.J., Mai D.H., Boenigk W., Guenther T. 1990. Die "Saugbaggerflora", eine Frucht- und Samenflora aus dem Grenzbereich Miozän-Pliozän von Sessenheim im Elsass (Frankreich). Doc. Nat. 57:1–207.
Gershoy A. 1928. Studies in North American violets I. General considerations. Bull. Vermont Agric. Exp. Stn. 279:1–18.
Gershoy A. 1934. Studies in North American violets III. Chromosome numbers and species characters. Bull. Vermont Agric. Exp. Stn. 367:3–91.
Gil-ad N.L. 1997. Systematics of *Viola* subsection *Boreali-Americanae*. Boissiera. 53:1–130.
Goloboff P.A., Farris J.S., Nixon K.C. 2008. TNT, a free program for phylogenetic analysis. Cladistics. 24:774–786.
Grant V. 1981. Plant speciation. New York: Columbia University Press.
Grauvogel C., Brinkmann H., Petersen J. 2007. Evolution of the glucose-6-phosphate isomerase: the plasticity of primary metabolism in photosynthetic eukaryotes. Mol. Biol. Evol. 24:1611–1621.
Harbaugh D.T. 2008. Polyploid and hybrid origins of pacific island sandalwoods (*Santalum*, Santalaceae) inferred from low-copy nuclear and flow cytometry data. Int. J. Plant Sci. 169: 677–685.
Harris S.A., Ingram R. 1991. Chloroplast DNA and biosystematics: the effects of interspecific diversity and plastid transmission. Taxon. 40:393–412.
Havran J.C., Sytsma K.J., Ballard H.E. 2009. Evolutionary relationships, interisland biogeography, and molecular evolution in the Hawaiian violets (*Viola*: Violaceae). Am. J. Bot. 96: 2087–2099.
Hepenstrick D. 2009. Tracing allopolyploid evolution with microsatellites in *Viola* [dissertation]. Zürich: Ecological Genetics and Evolution, WSL Swiss Federal Research Institute, Biology.
Howarth D.G., Baum D.A. 2005. Genealogical evidence of homoploid speciation in an adaptive radiation of *Scaevola* (Goodeniaceae) in the Hawaiian islands. Evolution. 59:948–961.
Huber K.T., Oxelman B., Lott M., Moulton V. 2006. Reconstructing the evolutionary history of polyploids from multilabeled trees. Mol. Biol. Evol. 23:1784–1791.
Jobb G., von Haeseler A., Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol. Biol. 4:9.
Kim S.T., Sultan S.E., Donoghue M.J. 2008. Allopolyploid speciation in *Persicaria* (Polygonaceae): insights from a low-copy nuclear region. Proc. Natl. Acad. Sci. U.S.A. 105:12370–12375.
Kovar-Eder J., Kvacek Z., Meller B. 2001. Comparing Early to Middle Miocene floras and probable vegetation types of Oberdorf N Voitsberg (Austria), Bohemia (Czech Republic), and Wackersdorf (Germany). Rev. Palaebot. Palynol. 114:83–125.
Kraytsberg Y., Khrapko K. 2005. Single-molecule PCR: an artifact-free PCR approach for the analysis of somatic mutations. Expert Rev. Mol. Diagn. 5:809–815.
Leitch A.R., Leitch I.J. 2008. Genomic plasticity and the diversity of polyploid plants. Science. 320:481–483.
Liu L., Yu L., Pearl D.K. 2010. Maximum tree: a consistent estimator of the species tree. J. Math. Biol. 60:95–106.
Lott M., Spillner A., Huber K.T., Moulton V. 2009. PADRE: a package for analyzing and displaying reticulate evolution. Bioinformatics. 25:1199–1200.
Lott M., Spillner A., Huber K.T., Petri A., Oxelman B., Moulton V. 2009. Inferring polyploid phylogenies from multiply-labeled gene trees. BMC Evol. Biol. 9:216. doi:210.1186/1471-2148-1189-1216.
Mai D.H. 2001. Die mittelmiozänen und obermiozänen Floren aus der Meuroer und Raunoer Folge in der Lausitz. III. Fundstellen und Palaeobiologie. Palaeontogr. Abt. B 258:1–85.
Mandáková T., Joly S., Krzywinski M., Mummenhoff K., Lysak M.A. 2010. Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. Plant Cell. 22:2277–2290.
Marcussen T., Blaxland K., Windham M.D., Haskins K.E., Armstrong F. 2011. Establishing the phylogenetic origin, history and age of the narrow endemic Viola guadalupensis (Violaceae). Am. J. Bot. 98:1-11.
Marcussen T., Oxelman B., Skog A., Jakobsen K.S. 2010. Evolution of plant RNA polymerase IV/V genes: evidence of subneofunctionalization of duplicated *NRPD2/NRPE2*-like paralogs in *Viola* (Violaceae). BMC Evol. Biol. 10:45.
Mason-Gamer R.J. 2008. Allohexaploidy, introgression, and the complex phylogenetic history of *Elymus repens* (Poaceae). Mol. Phylogenet. Evol. 47:598–611.
Matyášek R., Tate J.A., Lim Y.K., Šrubařová H., Koh J., Leitch A.R., Soltis D.E., Soltis P.S., Kovařík A. 2007. Concerted evolution of rDNA in recently formed *Tragopogon* allotetraploids is typically associated with an inverse correlation between gene copy number and expression. Genetics. 176:2509–2519.
McBreen K., Lockhart P.J. 2006. Reconstructing reticulate evolutionary histories of plants. Trends Plant Sci. 11:398–404.
McPherson G.D., Packer J.G. 1974. A contribution to the taxonomy of *Viola adunca*. Can. J. Bot. 52:895–902.
Miyaji Y. 1913. [Untersuchungen über die Chromosomezahlen bei einigen *Viola*-Arten]. Bot. Mag. Tokyo. 27:443–460 [in Japanese].
Miyaji Y. 1929. Studien über die Zahlenverhältnisse der Chromosomen bei der Gattung *Viola*. Cytologia. 1:28–58.
Moore D.M., Harvey M.J. 1961. Cytogenetic relationships of *Viola lactea* Sm. and other West European arosulate species. New Phytol. 60: 85–95.
Mort M.E., Crawford D.J. 2004. The continuing search: low-copy nuclear sequences for lower-level plant molecular phylogenetic studies. Taxon. 53:257–261.
Müller K. 2005. SeqState - primer design and sequence statistics for phylogenetic DNA data sets. Appl. Bioinformatics. 4: 65–69.
Mummenhoff K., Franzke A. 2007. Gone with the bird: late Tertiary and Quaternary intercontinental long-distance dispersal and allopolyploidization in plants. Syst. Biodivers. 5:255–260.
Nikitin V.P. 2007. Paleogene and Neogene strata in Northeastern Asia: paleocarpological background. Russ. Geol. Geophys. 48: 675–682.
Nishikawa T. 1988. Chromosome counts of flowering plants of Hokkaido (11). J. Hokkaido Univ. Educ. Sect. IIB. 38:33–40.
Nordal I., Jonsell B. 1998. A phylogeographic analysis of *Viola rupestris*: Three post-glacial immigration routes into the Nordic area? Bot. J. Linn. Soc. 128:105–122.
Popp M., Oxelman B. 2001. Inferring the history of the polyploid *Silene aegaea* (Caryophyllaceae) using plastid and homoeologous nuclear DNA sequences. Mol. Phylogenet. Evol. 20:474–481.
Popp M., Erixon P., Eggens F., Oxelman B. 2005. Origin and evolution of a circumpolar polyploid species complex in *Silene* (Caryophyllaceae) inferred from low copy nuclear RNA polymerase introns, rDNA, and chloroplast DNA. Syst. Bot. 30:302–313.
Popp M., Oxelman B. 2007. Origin and evolution of North American polyploid *Silene* (Caryophyllaceae). Am. J. Bot. 94:330–349.
Price J.P., Elliott-Fisk D. 2004. Tophographic history of the Maui Nui Complex, Hawai'i, and its implications for biogeography. Pac. Sci. 58:27–45.

Probatova N.S., Barkalov V.Y., Rudyka E.G. 2007. [Caryology of the flora of Sakhalin and the Kurile Islands: Chromosome numbers, taxononomic and phytogeographical comments]. Vladivostok (Russia): Dal'nauka [in Russian].

Russell N.H. 1955. The taxonomy of the North American acaulescent white violets. Am. Midl. Nat. 54:481–494.

Sang T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit. Rev. Biochem. Mol. Biol. 37:121–147.

Schäfer A., Utescher T., Mörs T. 2004. Stratigraphy of the Cenozoic Lower Rhine Basin, northwestern Germany. Newsl. Stratigr. 40: 73–110.

Simmons M.P., Ochoterena H. 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49:369–381.

Skottsberg C. 1940. Observations on Hawaiian violets. Acta Horti. Gothob. 13:451–528.

Smedmark J.E.E., Eriksson T., Evans R.C., Campbell C.S. 2003. Ancient allopolyploid speciation in *Geinae* (Rosaceae): Evidence from nuclear granule-bound starch synthase (GBSSI) gene sequences. Syst. Biol. 52:374–385.

Sokolovskaya A.P. 1960. [Geographical distribution of polyploidy in plants: investigation of the flora of Sakhalin]. Vestn. Leningr. Univ. Ser. Biol. 4:42–58 [in Russian].

Sokolovskaya A.P. 1963. [Geographical distribution of polyploidy in plants: investigation of the flora of the Kamchatka Peninsula]. Vestn. Leningr. Univ. Ser. Biol. 15:38–52 [in Russian].

Sokolovskaya A.P., Probatova N.S. 1986. [Chromosome numbers in some representatives of the Asteraceae, Iridaceae, Poaceae, Primulaceae, Violaceae from the Far East of the USSR]. Bot. Zh. (Leningr.) 71:1423–1425 [in Russian].

Soltis D.E., Soltis P.S. 1999. Polyploidy: recurrent formation and genome evolution. Trends Ecol. Evol. 14:348–352.

Soltis D.E., Albert V.A., Leebens-Mack J., Bell C.D., Paterson A.H., Zheng C., Sankoff D., de Pamphilis C.W., Wall P.K., Soltis P.S. 2009. Polyploidy and angiosperm diversification. Am. J. Bot. 96:336–348.

St. John H. 1989. Revision of the Hawaiian species of *Viola* (Violaceae). Hawaiian plant studies no. 135. Bot. Jahrb. Syst. 111:165–204.

Taylor R., Mulligan G. 1968. Cytological aspects of the vascular plants. Ottawa (Ontario): Plant Research Institute, Central Experimental Farm.

Tokuoka T. 2008. Molecular phylogenetic analysis of Violaceae (Malpighiales) based on plastid and nuclear DNA sequences. J. Plant Res. 121:253–260.

Valentine D.H., Merxmüller H., Schmidt A. 1968. *Viola* L. In: Tutin T.G., Heywood V.H., Burges N.A., Moore D.M., Valentine D.H., Walters S.M., Webb D.A., editors. *Flora europaea*. Vol. 2. Cambridge (UK): Cambridge University Press. p. 270–282.

van den Hof K., van den Berg R.G., Gravendeel B. 2008. Chalcone synthase gene lineage diversification confirms allopolyploid evolutionary relationships of European rostrate violets. Mol. Biol. Evol. 25:2099–2108.

Van der Burgh J. 1983. Allochthonous seed and fruit floras from the Pliocene of the lower Rhine Basin. Rev. Palaebot. Palynol. 40:33–90.

Van der Burgh J. 1987. Miocene floras in the lower Rhenish Basin and their ecological interpretation. Rev. Palaebot. Palynol. 52: 299–366.

Volkova S.A., Boyko E.V., Antusheva I.V. 2003. [Chromosome numbers of some species of Kamchatka and Sakhalin flora]. Bot. Zh. (St. Petersbg.) 88:154–155 [in Russian].

Wendel J.F., Schnabel A., Seelanan T. 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc. Natl. Acad. Sci. U.S.A. 92:280–284.

Wendel J.F., Doyle J.J. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis D.E., Soltis P.S., Doyle J.J., editors. Molecular systematics of plants II: DNA sequencing. Dordrecht (the Netherlands): Kluwer Academic Press. p. 265–296.

Wood T.E., Takebayashic N., Barker M.S., Mayrosee I., Greenspoond P.B., Rieseberg L.H. 2009. The frequency of polyploid speciation in vascular plants. Proc. Natl. Acad. Sci. U.S.A. 106: 13875–13879.

Yoo K.O., Jang S.K., Lee W.T. 2005. Phylogeny of Korean Viola based on ITS sequences. Korean J. Plant Tax. 35:7–23 [in Korean].