Evolution of Viola stagnina and its sisterspecies by hybridisation and polyploidisation

Hof, K. van den

Chapter

2

Chalcone Synthase Gene Lineage Diversification confirms Allopolyploid Evolutionary Relationships of European Rostrate Violets[1]

K. van den Hof, R.G. van den Berg and B. Gravendeel

Phylogenetic relationships among and within the subsections of the genus *Viola* are still far from resolved. We present the first organismal phylogeny of predominantly western European species of subsection *Rostratae* based on the plastid *trn*S-*trn*G intron and intergenic spacer and the nuclear low-copy gene Chalcone Synthase (CHS) sequences. CHS is a key enzyme in the synthesis of flavonoids, which are important for flower pigmentation. Genes encoding for CHS are members of a multigene family. In *Viola,* three different CHS copies are present. CHS gene lineages obtained confirmed earlier hypotheses about reticulate relationships between species of *Viola* subsection *Rostrata*e based on karyotype data. Comparison of the CHS gene lineage tree and the plastid species phylogeny of *Viola* reconstructed in this study indicates that the different CHS copies present in *Viola* are the products of both recent and more ancient duplications.

Key words: Chalcone synthase, gene lineage diversification, phylogeny, *Viola* subsection *Rostratae*, allopolyploidy, *trn*S-*trn*G.

[1]Published as: van den Hof et al., 2008. Mol. Biol. Evol. 25: 2099-2109.

Introduction

Speciation through hybridization is considered a common process in higher plants. Although hybrids between distantly related taxa are usually sterile, they can become fertile again by doubling their chromosome numbers. The resulting chimeric species can have two or more sets of chromosomes derived from different parental species; this is called allopolyploidy (Stebbins, 1971; Song et al., 1995; Bennett, 2004; Hegarty and Hiscock, 2005). In contrast with allopolyploidy, which may occur in connection with hybridization between taxa that are not very closely related, hybrid speciation without a change in chromosome number may occur in cases where the parental species are closely related and their primary hybrid is somewhat fertile; this process is called homoploid hybrid stabilization (Rieseberg, 1997; Rieseberg et al., 2003; Abbott et al., 2005).

Polyploid evolution has been an important factor in the evolutionary history of land plants, and continues to be so also in extant lineages such as the plant genus *Viola* (Violaceae). In fact, the first report of an infrageneric series of polyploid levels was from *Viola* (Miyaji, 1913). The base chromosome number of *Viola* is believed to be x=6 or x=7, but the vast majority of north-temperate taxa have been shown to be paleo-allotetraploid with secondary base numbers of x=10 or x=12 (Nordal and Jonsell, 1998; Marcussen and Nordal, 1998; Karlsson et al., 2009). These are hereafter referred to as secondary diploids. Further polyploidy based on these secondary diploid chromosome numbers has been demonstrated especially within the species-rich subsections of section *Viola* (Karlsson et al., 2009).

Within Section *Viola* subsection *Rostratae* Kupffer (sometimes treated as the separate section *Trigonocarpea* (Godr.) Vl. V. Nikitin), most species have retained the secondarily diploid chromosome number of $2n$=20. However, subsequent polyploidization events have led to the formation of higher-ploids with chromosome numbers of $2n$=40 (octoploid), 60 (dodecaploid) or even 58 (sub-dodecaploid); these are hereafter referred to as secondary tetraploids and (sub-)hexaploids, respectively. Nearly all of these secondary polyploids, a total of ten species, are native to western Eurasia, and their relatively recent polyploid parentages have been investigated in a series of cytological studies in the late 1950s and early 1960s (fig. 1) (Valentine 1950, 1958; Moore and Harvey 1961; Harvey, 1966).
The subsection consists of about fifty species with a northern temperate distribution in North America and Eurasia. Most species have white to dark lilac flowers and grow in woodlands. Subsection *Rostratae* is characterized primarily by primitive characters. Phylogenetic analyses based on nuclear ribosomal Internal Transcribed Spacer (nrITS) sequences have recovered that the subsection is paraphyletic with respect to a number of other north-temperate groups (Ballard et al., 1999; Yoo et al., 2005). In Europe, where subsection *Rostratae* is morphologically most diverse, the subsection has traditionally been subdivided in a variable number of morphologically defined groups, usually at the series level. Series *Rosulantes* is characterized by having a basal rosette and flowers produced only from the lateral aerial shoots; this growth form is found also in other sections and may be considered as primitive within the genus. Series *Mirabiles* differs from the *Rosulantes* in producing flowers also from the basal leaf rosette, series *Arosulatae* in lacking the basal rosette altogether, and series *Repentes* in being stoloniferous and producing flowers from the rosettes. However, the recognition of series is problematic for

two main reasons. First, the series typically define small groups of species by a very limited number of autapomorphies, thereby rendering the remaining groups paraphyletic and defined by synplesiomorphies only. Second, several of these series cannot be considered monophyletic because of the alloploid relationships between taxa of different series.

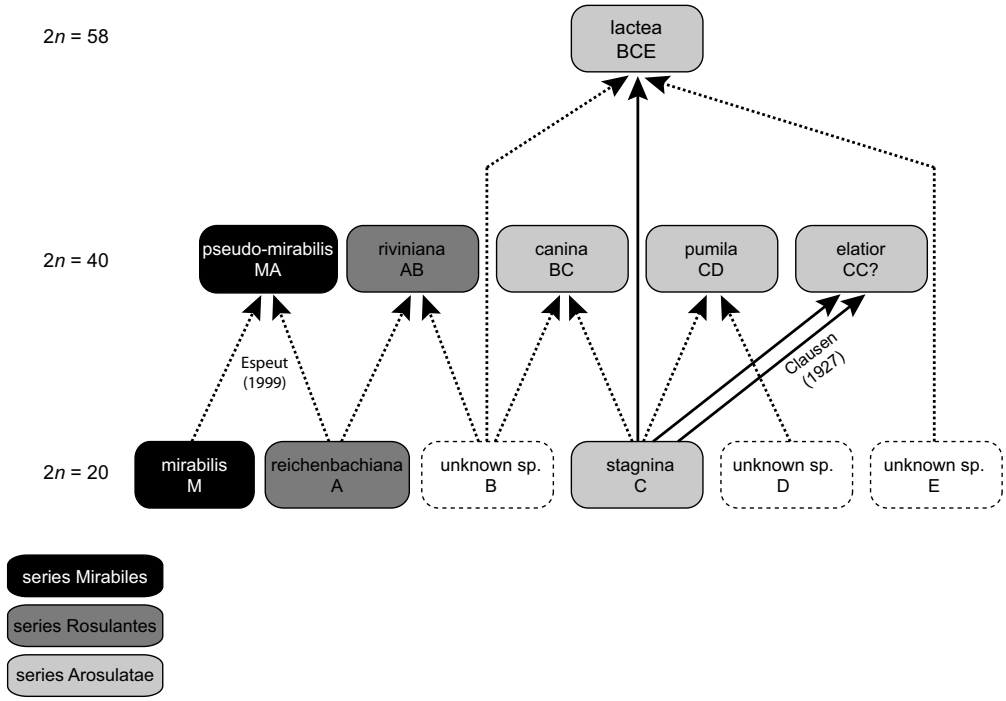

Fig. 1. Hypotheses of relationships between different genome types (A, B, C, D, E and M) in species of *Viola* subsection *Rostratae*. Series affinity is indicated with shades of gray: series *Mirabiles* black; series *Rosulantes* dark grey; series *Arosulatae* light gray. Presumably extinct taxa are indicated with dashed lines. Data from Moore and Harvey (1961) except where indicated (Clausen, 1927; Espeut, 1999).

Series *Arosulatae* defines a small group of five western Eurasian species. Species of the series are specialists of temporarily flooded habitats, rather than woodland, and are easily characterized by lacking leaf rosettes and by their leaf and stipule characters. Traditionally seven species have been recognized, but two (the East Asian *V. acuminata* and the submediterranean *V. jordanii*) must be omitted on the basis of having a leaf rosette and the quite different choices of habitat. The remaining five species are all Central European. *Viola stagnina* is the only secondary diploid with 2*n*=20, whereas three species are secondary tetraploids with 2*n*=40 (*V. canina*, *V. elatior*, *V. pumila*) next to a secondary sub-hexaploid with 2*n*=58 (*V. lactea*) (Moore and Harvey, 1961).

Especially the study by Ballard et al. (1999) indicated that the taxonomy of the genus *Viola* needs revision and that more molecular phylogenetic studies are called for. Although the nrITS region used by Ballard et al. (1999) was useful for recognizing infrageneric groups of the genus *Viola*, nrITS is generally not useful for examining evolutionary relationships among polyploid lineages. This is because recombination and concerted evolution between orthologous nrITS copies often lead to retention of only one copy type and erasion of the

other parental copy (Wendel et al., 1995; Álvarez and Wendel, 2003). This is usually also the case in *Viola*, as nearly all investigated species have retained only one nrITS copy type regardless of ploidal level (Ballard et al., 1999; Malécot et al., 2007). The nrITS as a phylogenetic marker is therefore not suitable for recovering the reticulate relationships within the genus. Plastid markers generally also have the problem of retention of a single parental copy as these are usually uniparentally inherited in plants; furthermore, sequence variation in plastid markers is usually low (Corriveau and Coleman, 1988; Taberlet et al., 2007). Again, reticulate relationships therefore remain obscure.

Álvarez and Wendel (2003) suggested using single or low copy nuclear markers to circumvent the problem of concerted evolution causing misleading phylogenetic reconstructions of polyploid species. Phylogenetic analysis of paralogous and orthologous copies of single or low copy genes in alloploid species is also a good method to reveal the parental contributors to alloploid genomes. This method has been successfully applied in numerous studies (e.g. Popp and Oxelman, 2001; Smedmark et al., 2005).

We utilized the low copy nuclear Chalcone synthase (CHS) gene as a phylogenetic marker in *Viola* subsection *Rostratae*. As an independent dataset we chose the *trn*S-*trn*G intergenic spacer and intron as plastid phylogenetic marker, as this region proved to be sufficiently informative in *Viola* to assess interspecific relationships.

CHS is the first enzyme in the flavonoid synthesis pathway and is encoded by a small gene family (Durbin et al., 1995). Flavonoids are important secondary metabolites responsible for a multitude of tasks in plants, ranging from flower and fruit coloration and protection against UV radiation to pathogen defense and pollen development (Harborne, 1994). In *Viola cornuta*, three different CHS gene copies were found to be expressed from early stages of flower coloration onwards (Farzad et al., 2003). In general, genes of the CHS family consist of one intron flanked by two exons. There is high variation in the number of CHS copies among angiosperms. In asterids the number of CHS copies ranges from a single copy in *Antirrhinum* (Sommer and Saedler, 1986) to six copies in *Ipomoea* (Clegg and Durbin, 2003) and eight in *Petunia* (Koes et al., 1987). Similarly for the rosids, both *Arabidopsis* (Wang et al., 2007) and *Populus* (Tuskan et al., 2006) have two CHS copies, whereas both *Vitis* (Sparvoli et al., 1994; Jaillon et al., 2007) and *Viola cornuta* cultivars (Farzad et al., 2003; 2005) have three CHS copies.

We collected different CHS paralogues in species of *Viola* subsection *Rostratae* and analyzed these phylogenetically to 1) test earlier hypotheses about reticulate relationships of several allopolyploid taxa based on karyotype data in subsection *Rostratae* (e.g. between *V. stagnina*, a possible Dutch endemic, and its closest relatives), 2) make a comparison with a species phylogeny of *Viola* subsection *Rostratae* based on sequences of the plastid *trn*S-*trn*G intron and intergenic spacer to infer how many duplications of CHS took place during the evolution of *Viola*.

Materials and Methods

Taxon sampling

In total, 30 *Viola* taxa with a predominantly western European origin were sampled, of which 21 taxa belong to *Viola* subsection *Rostratae*. The nine taxa outside subsection *Rostratae* represent sections *Andinium*, *Boreali-Americanae*, *Chamaemelanium*, *Erpetion*,

and *Melanium* and subsection *Viola* of section *Viola*. These species appeared to be either closely or more distantly related to the species of subsection *Rostratae* in a previous molecular phylogenetic study of *Viola* (Ballard et al., 1999). DNA was obtained from freshly collected material from the field and from herbarium collections.

For reconstruction of the CHS gene lineage tree, two different parts of the gene were sampled, the intron and exon 2. Exon 2 lineages available in Genbank from representatives of major Angiosperm clades were included in the analysis to find out whether the different CHS copies present in *Viola* are the products of recent or more ancient duplications. The following lineages were sampled: Gymnosperms: *GbCHS* (*Ginkgo biloba*, AY647263) and *PsCHS* (*Pinus sylvestris*, X60754); Monocots: *IhCHS* (*Iris x hollandica*, AB232914), *HvCHS* (*Hordeum vulgare*, X58339), and *ZmCHS* (*Zea mays,* AY728478, X60204); Core eudicots: *VvCHS* (*Vitis vinifera*, AB015872, AB066275, EF192464, AM 454341, X75969); Rosids: *GmCHS* (*Glycine max*, AY262686), *PsCHS* (*Pisum sativum*, D88263, D88262, D88261, D88260, X63333), *PtCHS* (*Populus* spp., DQ371804, EF147137, EF147091, DQ371802), *VcCHS* (*V. cornuta* cultivar, AY497407, AY497414); Asterids: *AmCHS* (*Antirrhinum majus*, X03710*), DcCHS* (*Daucus carota*, D16255), and *PhCHS* (*Petunia hybrida*, X14597).

The phylogenetic analyses performed were all rooted differently. There were several reasons for this. First of all, plastid sequences of non-Violaceae were not used for phylogenetic analyses. The Angiosperm Phylogeny Group topology was used to constrain the analyses instead (see below). For the plastid phylogeny, closely related genera of *Viola* were used as outgroups. Second, CHS intron sequences outside *Viola* could not be aligned with CHS intron *Viola* data because of too high sequence divergence. For the CHS intron analyses, we therefore tentatively used *Viola CHS3* as outgroup. Third, CHS gene duplication events could only be assessed with a broad taxonomic sampling. For the CHS exon 2 analyses, we therefore used gymnosperm lineages for rooting.

DNA Extraction, Polymerase Chain Reaction Amplification, Cloning, and Sequencing

Total genomic DNA was extracted using the Dneasy Plant Mini Kit (Qiagen, Hilden, Germany) and the cetyltrimethylammonium bromide (CTAB) method of Doyle JJ and Doyle JL (1987) with some modifications. Leaf material was ground using a Ratch Mill. In total, 750 µl CTAB buffer (Doyle and Doyle, 1987) was added to the ground material together with proteinase K and RNase. After incubating for 30 minutes at 60°C, 750 µl of chloroform-isoamyl alcohol (24:1) was added. The samples were briefly vortexed and then centrifuged for 10 minutes at 12,000 rpm. The upper aqueous layer was transported to a clean 2 ml tube. A total of 500 µl chloroform-isoamyl alcohol (24:1) was added and the samples were again centrifuged at 12,000 rpm. After 5 minutes spinning, the upper phase was transferred to a new 2 ml tube. The DNA was then precipitated by adding cold 500 µl isopropanol. The samples were shaken 5 to 10 minutes and subsequently centrifuged for 15 minutes at 12,000 rpm. The supernatant was removed and 70% ethanol was added. The samples were subsequently shaken vigorously for 2 minutes, after which the ethanol was poured off. The remaining ethanol was removed by evaporation. The resulting DNA pellet was dissolved in 200 µl 0.1x Tris-EDTA buffer.

In total, one plastid region (*trn*S-*trn*G spacer and intron) and one nuclear region (CHS intron and exon 2) were amplified and sequenced. Polymerase chain reaction (PCR) amplification of the plastid spacer and intron was performed with primers designed by

Shaw et al. (2005). For the *trn*S spacer, the primers trnS^{GCU} (5′-AGA TAG GGA TTC GAA CCC TCG GT-3′) and trnG2S (5′-TTT TAC CAC TAA ACT ATA CCC GC-3′) were used. The *trn*G intron was amplified with the primers trnG^{UUC} (5′-GTA GCG GGA ATC GAA CCC GCA TC-3′) and trnG2G (GCG GGT ATA GTT TAG TGG TAA AA).

The primers CHSX1F (5′-AGG AAA AAT TCA AGC GCA TG-3′) and CHSX2RN (5′-TTC AGT CAA GTG CAT GTA ACG -3′) designed by Strand et al. (1997) were used for amplifying the CHS intron. The primers CHS forward (5′-TAY CAR CAR GGN TGY TTY GC-3′) and CHS reverse (5′-GGR TGD GCD ATC CAR AAV A-3′) from Farzad et al. (2003) were used to amplify exon 2 of the CHS gene (fig. 2a). The generated intron and exon sequences did not overlap as the intermediate part turned out to be too large and too heterogeneous for this. PCR fragments for several *Viola* species of the CHS intron are shown in fig. 2b. Per individual, 12 clones were analyzed and consensus sequences were compiled from 3-7 individual clones.

Fig. 2. Amplified regions of chalcone synthase (CHS):

Fig 2a. Map of the two exons and interjected intron of the CHS gene. Primers and their binding sites are indicated for the CHS intron data set (black triangles) and the CHS exon 2 data set (white triangles).

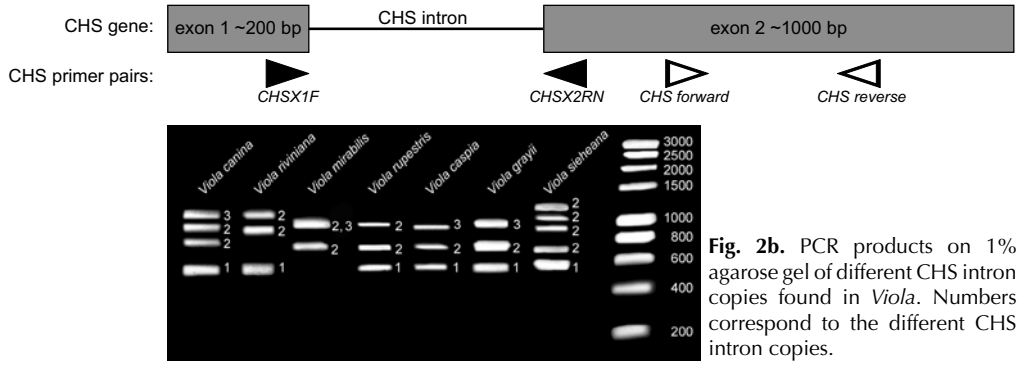

Fig. 2b. PCR products on 1% agarose gel of different CHS intron copies found in *Viola*. Numbers correspond to the different CHS intron copies.

PCR amplification conditions for the *trn*S spacer consisted of denaturing for 50 s at 95°C, annealing for 1 min at 53°C, and extension for 2 min at 72°C. This cycle was repeated 35 times. The *trn*G intron was amplified with the same conditions except for the annealing temperature which was 56°C. PCR amplification conditions for the CHS intron consisted of denaturing for 1 min at 95°C, annealing for 90 s at 53°C, and extension for 2 min at 72°C. This cycle was repeated 35 times. The conditions for amplifying the CHS exon 2 consisted of denaturing for 45 s at 95°C, annealing for 1 min at 55°C and extension for 1 min at 72°C. This cycle was repeated 40 times.

PCR products were purified using the Promega Wizard Purification System, cloned using the pGEM®-T Easy Vector System and sequenced using the M13 primers with 30 s denaturing at 95°C, annealing for 30 s at 50°C, and extension for 1 min at 72 °C. This cycle was repeated 35 times. PCR products were purified and analyzed on an ABI 377 (Applied

Biosystems Inc., Foster City, CA) or a MegaBACE Sequence Analyzer 4.0 (Amersham Biosciences, Uppsala, Sweden) automated sequencer using the manufacturers' protocols.

Phylogenetic analyses

DNA sequences were aligned using McClade 4.06 (Maddison DR and Maddison WP, 2003) with the pairwise alignment option and manual adjustment where necessary. Individual insertion and deletion events were manually added as additional binary characters.

MrModeltest version 2.2 (Nylander, 2004) was used to find the best model of sequence evolution (Posada and Crandall, 1998). The models used for Bayesian analyses were the symmetrical model with separate gamma distributions and a separate proportion of invariant sites for CHS exon 2 (SYMIG model), the General Time Reversible model with gamma distribution for CHS intron (GTRG model), and the General Time Reversible model with gamma distribution and a separate proportion of invariant sites for *trn*S-*trn*G (GTRIG model). Maximum Parsimony (MP) analyses were carried out with PAUP* 4.0b10 (Swofford, 2003). Phylogenies were obtained using the heuristic search option, with twenty random sequence additions and Tree Bisection-Reconnection branch swapping. After each sequence addition, a maximum of 10,000 trees was saved.

For MP, bootstrap support (Felsenstein, 1985) was calculated with 2,000 bootstrap replicates, using only ten random sequence additions each bootstrap replicate. After every random sequence addition replicate a maximum of 2,500 trees were saved. Bayesian inference analyses were performed using MrBayes 3.1 (Huelsenbeck and Ronquist, 2001). Markov Chain Monte Carlo analyses (MCMC) were run for eight million generations with five simultaneous MCMCs, saving one tree per 100 generations. The burn-in values were identified using the program Tracer 1.3 (Rambaut and Drummond, 2004).

To convert the CHS gene tree composed of multiple paralogous lineages from allopolyploid taxa into a species tree to assess gene duplications, GeneTree version 1.3 (Page and Charleston, 1997) was used. The analyses were run with default settings. GeneTree requires fully resolved organismal and gene trees as input. For the organismal tree, one of the 200,000 fully resolved most parsimonious trees (MPTs) of *trn*S-*trn*G data was chosen randomly. This analysis was constrained for all non violets to the latest angiosperm phylogeny topology as depicted on the Angiosperm Phylogeny Website (version 8, June 2007) (www.mobot.org/MOBOT/research/Apweb/). For the CHS exon 2 gene tree, the 95% most probable Bayesian tree was used.

Homology Assessment of CHS Copies

The CHS lineages found were assigned to different copies based on size, sequence divergence and phylogenetic position (Helariutta et al., 1996; Doyle and Davies, 1998; Smedmark et al., 2005). CHS fragments within one species with only minor divergence and gaps were interpreted as alleles. The different CHS copies of *V. cornuta* published by Farzad et al. (2003) always ended up in a single clade in all analyses performed here. We therefore used a single representative sequence only. When size difference and sequence divergence were more apparent, e.g. by the presence of large indel events, the CHS fragment was treated as a paralogous copy. Our classification of alleles and paralogous copies was further confirmed by topological positions in the phylogenies obtained.

Results

*trn*S-*trn*G

MP analyses of the *trn*S-*trn*G alignment produced a total of 200,000 MPTs with 537 steps (consistency index [CI] = 0.8239; retention index [RI] = 0.7312). The majority rule consensus tree (data not shown) has a similar topology as the Bayesian tree (fig. 3). We plotted both the Bootstrap Support values (BS) and Posterior Probability Index values (PPI) on the latter. All species sampled of *Viola* subsection *Rostratae* ended up in five different, poorly to well supported subclades (<50–98% BS; 0.56–1.00 PPI). The largest subclade consists of *V. stagnina*, *V. elatior*, *V. lactea*, *V. canina*, *V. sieheana*, *V. jordanii*, *V. oligyrtia*, *V. rupestris*, and *V. pumila*.

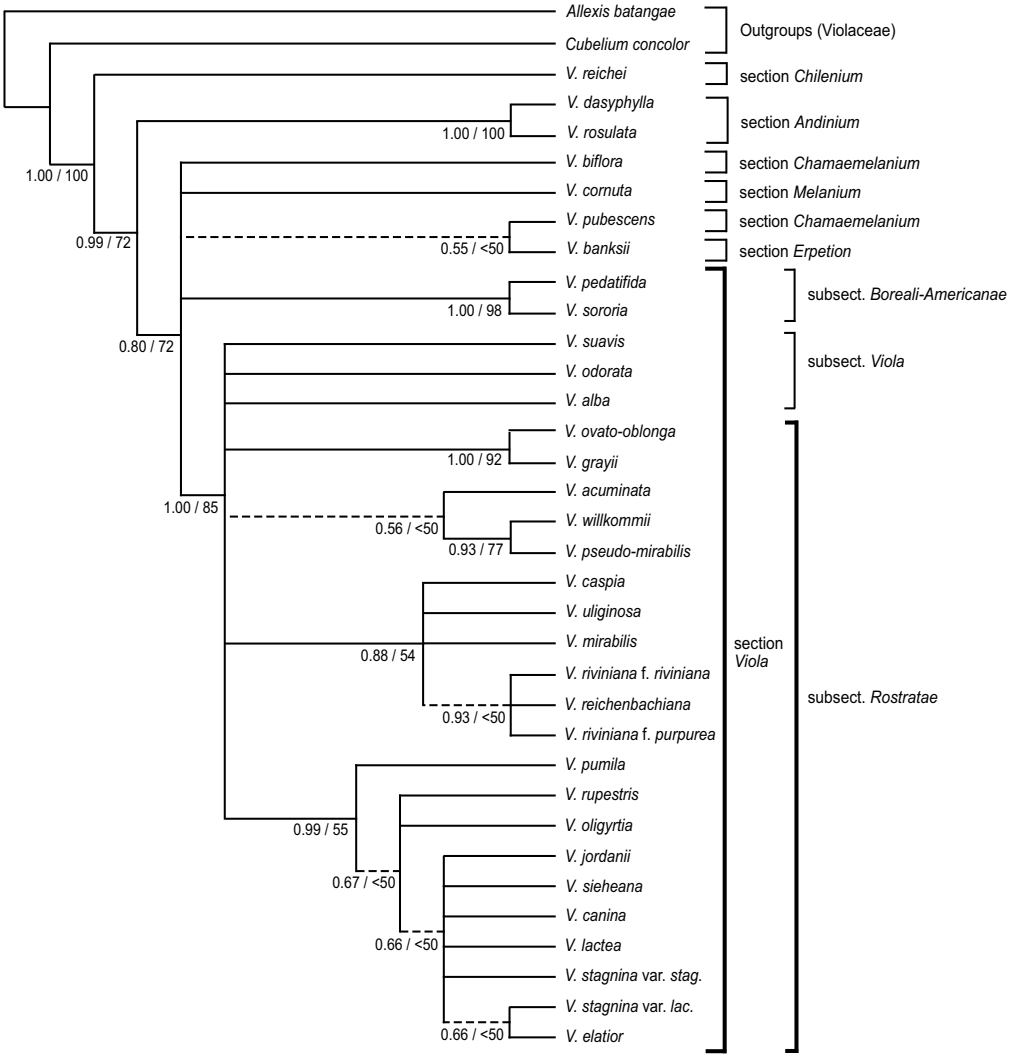

Fig. 3. Plastid *trn*S-*trn*G phylogeny of *Viola*.
(MP majority rule consensus of 200,000 trees; ci = 0.8239, ri = 0.7312, 537 steps. Numbers on branches refer to PPI and BS values).

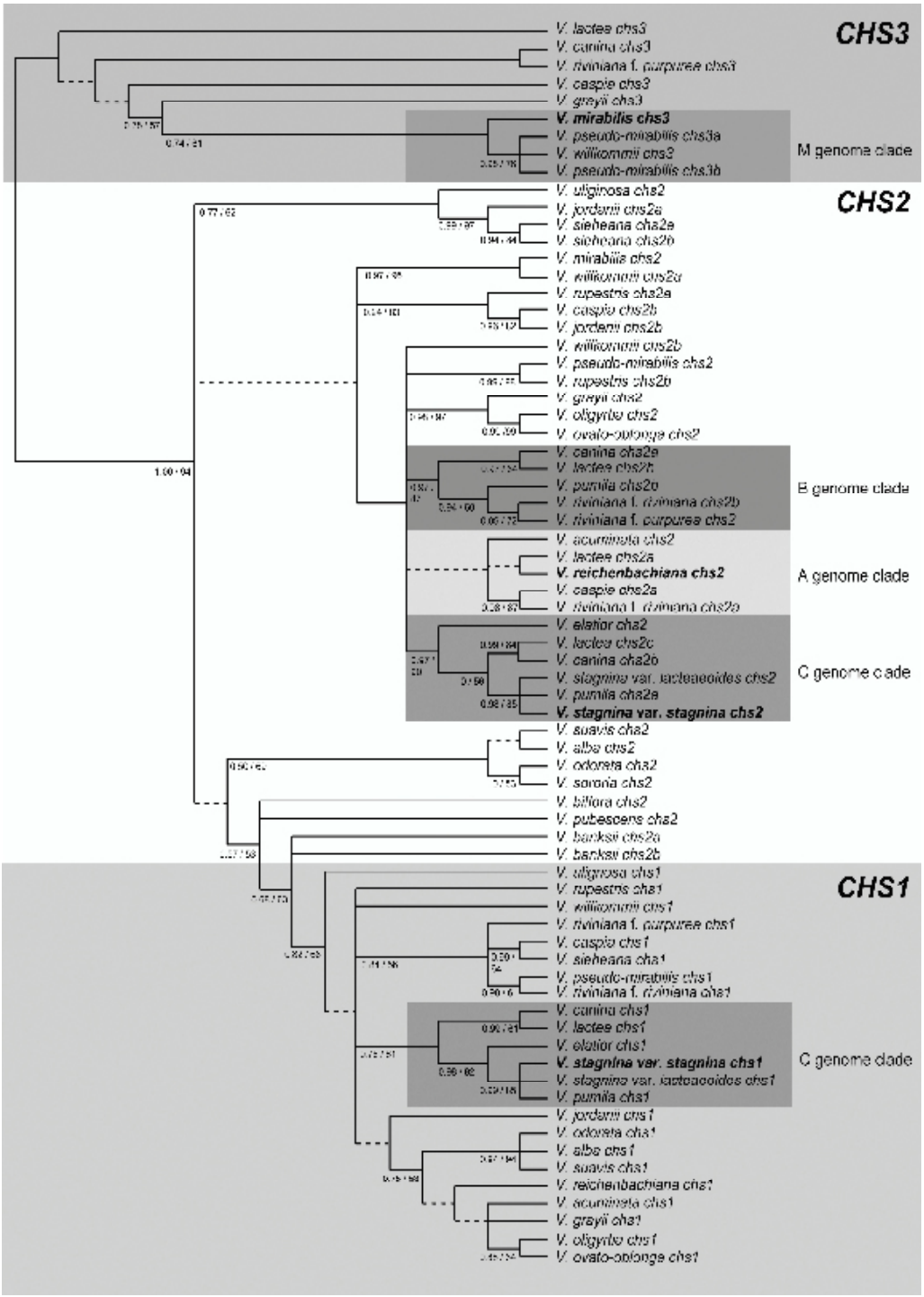

Fig. 4. CHS intron gene lineage tree.
(MP majority rule consensus of 200,000 trees; ci = 0.7828 ri = 0.8824, 921 steps. Numbers on branches refer to PPI and BS values) The genomes described in Figure 1 which could be recognized are indicated. Taxa in bold refer to extant secondary diploids.

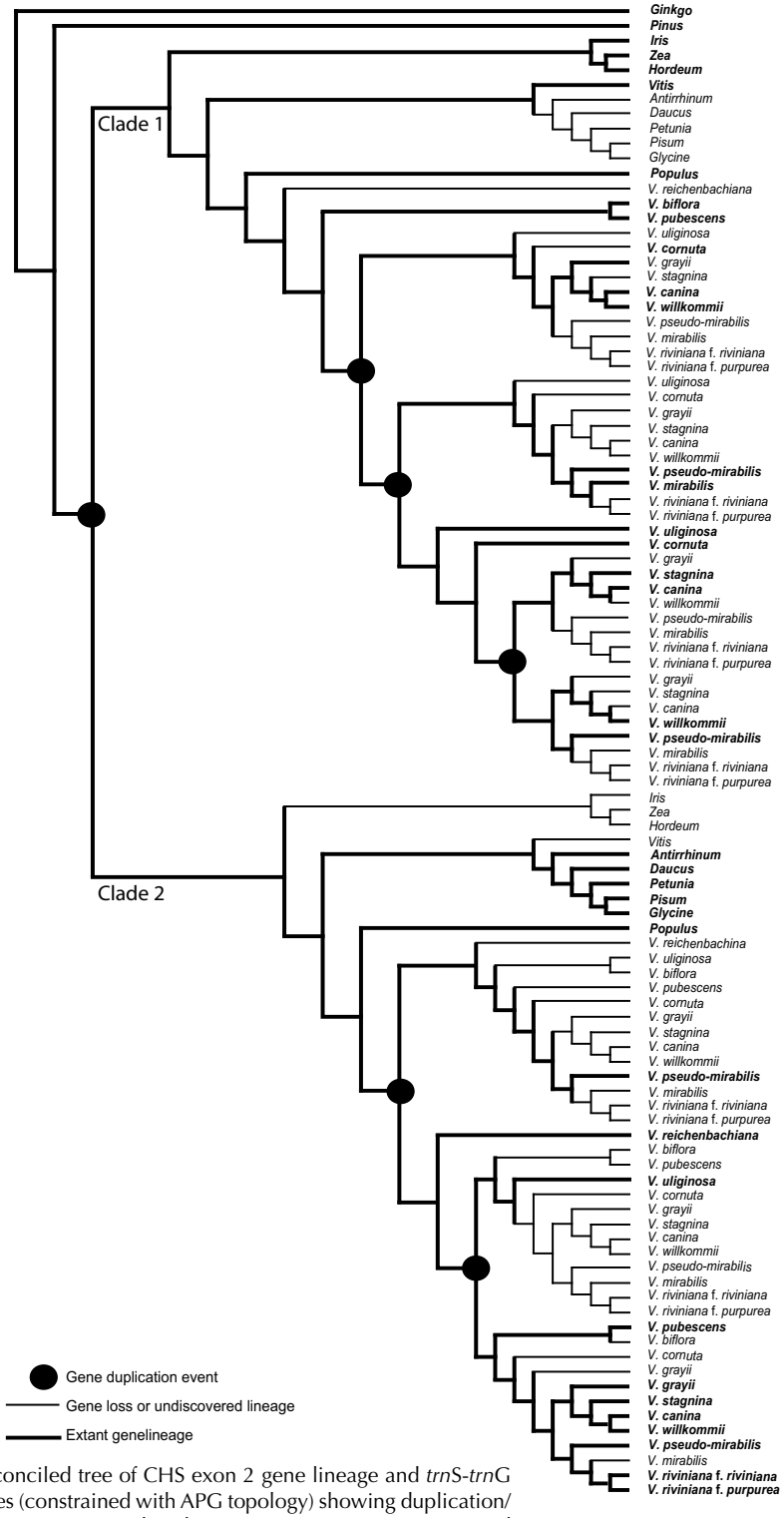

Fig. 5. Reconciled tree of CHS exon 2 gene lineage and *trn*S-*trn*G phylogenies (constrained with APG topology) showing duplication/loss events, reconstructed with GeneTree version 1.3 (Page and Charleston, 1997).

CHS intron and exon 2

Three copies of the CHS intron were found in the sampled species of *Viola* subsection *Rostratae* (figs. 2b and 4). Copies 1 (*CHS1*, 600 base-pairs [bp]) and 2 (*CHS2*, 735–1100 bp) have a relatively similar sequence identity and were found in all *Viola* species sampled with the exception of *V. biflora*, *V. banksii*, and *V. pubescens*. The third copy was not found in all species and was the largest sized (*CHS3,* 775–1,160 bp). The *CHS3* copy is probably present in more taxa, but amplification failures probably led to an under sampling of this particular copy. It differed quite substantially from the other two copies in size and sequence similarity. In contrast with *CHS1*, multiple paralogs/orthologs were found in *CHS2* and *CHS3*.

The complete CHS intron alignment consisted of 2,980 bp after exclusion of a 64 bp segment that was too variable for proper alignment. A total of 302 characters were phylogenetically informative, of which 12 were indel characters (indels varying in size between 5 and 422 bp). Two indels, found in *CHS1* and *CHS2*, seemed to be the result of slip strand mispairing as many repeats were found in these regions. The first (TGATTT) and second repeat (TGTT) were repeated up to four times. The other ten indels lacked a repetitive structure. Most of the indels occurred in *CHS2*. In *CHS3*, two large indels were found of 182 and 422 bp, respectively.

MP analyzed of CHS intron sequences produced 200,000 MPTs (921 steps, CI = 0.7828, RI = 0.8824). The majority rule consensus tree (fig. 4) had a similar topology as the Bayesian tree (data not shown).

At least two different copies of CHS exon 2 were found in *Viola*. Both copies had a similar size (984 bp) and were retrieved from almost all *Viola* species analyzed. They differed substantially in sequence similarity. Of the 984 bp retrieved, 498 were phylogenetically informative. One autapomorphic gap was found. Bayesian analyses of CHS exon 2 sequences produced a topology similar to MP (data not shown), in which two main clades were present comprising the different copies of CHS exon 2.

Reconciliation of Gene Tree and Species Tree

A reconciled tree (fig. 5) reconstructed with the program GeneTree (Page and Charleston 1997) was used to visualize CHS exon 2 duplications during the evolution of *Viola*. It seems that assuming six CHS gene duplication events is sufficient to make the gene and species tree congruent.

Discussion

Polyploidy in Viola Subsection *Rostratae*

The internal topology of the *CHS2* intron clades (fig. 4) is in general agreement with previously inferred relationships between parental species and their allopolyploid hybrids in *Viola* subsection *Rostratae* (Moore and Harvey, 1961). Moore and Harvey (1961) could recognize the parental karyotypes in the genomes of artificially constructed allopolyploid *Viola* hybrids by the unique size and shape of the chromosomes. Subsequently, they

used observations on chromosome pairing to formulate hypotheses regarding the origin of allopolyploids (fig. 1). In their study, five different types of genomes (A–E) could be recognized, each referring to the secondary diploid level (2*n*=20). Only the A and C genome occurred in extant secondary diploids, namely in *V. reichenbachiana* (A) and *V. stagnina* (C). The other four genome types were found only in combination with other genome types in the secondary tetraploid (2*n*=40) or sub-hexaploid taxa (2*n*=58). From this, they concluded that *V. stagnina* (C) contributed a C genome to *V. canina* (BC) and its close relative *V. lactea* (BCE) and possibly also to *V. pumila* (CD). Similarly, *V. reichenbachiana* (A) would have contributed an A genome to *V. riviniana* f. *riviniana* (AB). Thus, the species possessing the B genome was involved in the origin of both *V. riviniana* f. *riviniana* (AB) and *V. canina* (BC). The authors attributed the three "missing" genomes (B, D, and E) to secondary diploids species that might have become subsequently extinct, at least in Europe. *Viola elatior* was not included in this study.

In the *CHS2* intron tree, *V. canina* was found to have one orthologue in common with *V. stagnina* (corresponding to genome C) and a second in common with *V. riviniana* (corresponding to genome B). The second orthologue of *V. riviniana* was found to be closely related to *V. reichenbachiana* (corresponding to genome A). Like *V. canina*, *V. pumila* had one orthologue in common with *V. stagnina* (genome C) while its second copy was found to be closely related to *V. canina* and *V. riviniana* suggesting that genome D could have been derived from genome B.

The phylogenetic position of the gene lineages retrieved from *V. elatior* suggests that this particular species probably contains the C genome because its *CHS2* intron copy ended up close to *V. stagnina*. The chromosome number of *V. elatior* (2*n*=40) suggests that it is a secondary tetraploid, but we were not successful in detecting more than one orthologue in this species. Unpublished isozyme studies also reveal a lower number of allozymic bands than usual in species of this ploidal level. These findings, together with earlier observations of quadrivalents in the meiosis of the species (Clausen, 1927) indicate that *V. elatior* may be an autopolyploid derivative of some *stagnina*-like ancestral species possessing the C genome.

In contrast with the *Arosulatae* series, which was found to be monophyletic for *CHS1* and part of *CHS2*, CHS lineages retrieved from species assigned to the *Rosulantes* and *Mirabiles* series did hardly ever end up in the same clades. This is probably caused by the fact that the morphological characters used to delimit these series are phylogenetically uninformative. Unknown hybridization and polyploidisation events within and between these series probably also cause paraphyly.

The small series *Mirabiles* consists of one secondarily diploid (2*n*=20) species, *V. mirabilis*, and two secondary tetraploids (2*n*=40). The local endemic *V. pseudo-mirabilis* of Les Grands Causses in southern France has been variously interpreted on morphological grounds as an intermediate between *V. mirabilis* and *V. riviniana* (Valentine et al., 1968) or as a polyploid derivative of *V. mirabilis* and *V. reichenbachiana* (Espeut, 1999). The latter view has later been confirmed by own unpublished isozyme data and a chromosome count of 2*n*=40 (Verlaque and Espeut, 2007). In the present study, *V. pseudo-mirabilis* ends up as sister to *V. riviniana* (61% BS; 0.99 PPI), which is not in contradiction to the previous findings because *V. riviniana* is itself a polyploid derivative of *V. reichenbachiana*. The other secondary tetraploid *Mirabiles* species, *V. willkommii*, endemic to northern Spain, is morphologically rather similar to *V. mirabilis* but differs considerably in choice of habitat. This particular species is a secondary tetraploid and has been supposed to have

originated from *V. mirabilis* and *V. rupestris* (Marcussen, personal communication). Our CHS data confirm the parentage of *V. mirabilis* but the second parent of *V. willkommii* remains unclear. Two paralogues of the *CHS2* intron were retrieved from *V. willkommii*, of which one ended up in a strongly supported clade (96% BP; 0.97 PPI) with *V. mirabilis* and the other in an unresolved polytomy. The *CHS1* intron fragment retrieved from *V. willkommii* ended up in a basal dichotomy with the lineage of *V. rupestris*. Future genome type data should be collected of the *Mirabiles* series to confirm hypotheses about reticulate relationships suggested by the CHS gene lineages obtained here.

Closest Relatives of Viola stagnina

Viola subsection *Rostratae* as mentioned before has been taxonomically subdivided in series *Arosulatae*, *Mirabiles*, *Repentes* and *Rosulantes* (Valentine, 1958). *Viola stagnina* is considered to be a member of the *Arosulatae* series together with *V. canina*, *V. elatior*, *V. lactea*, and *V. pumila*, based on the lack of a basal leaf rosette. This taxonomic placement is supported by the fact that in our study, two *Arosulatae* clades are present in the CHS intron gene lineage tree (fig. 4), one consisting of *CHS1* intron lineages and one of *CHS2* intron lineages. These lineages were all retrieved from species assigned to the *Arosulatae* series.

Based on previous morphological and karyological studies and our results, we can conclude that the closest relatives of *V. stagnina* are *V. pumila*, *V. elatior*, *V. canina* and *V. lactea*. Fingerprinting techniques are currently being applied to assess gene flow between different European populations of *V. stagnina* and its close relatives to determine whether a Dutch variety of *V. stagnina* deserves a different taxonomic status.

CHS Lineage Diversification in Viola

In the reconciled tree (fig. 5), two main clades are present. Unfortunately, GeneTree does not provide statistical support for individual nodes. The congruence in topology with Huang et al. (2004) and Yamazaki et al. (2001) indicates that a general phylogenetic signal was recovered, though. The first clade consists of monocot, core eudicots and *Viola* representatives. This clade indicates that at least one duplication event in the CHS gene family took place before the split between the monocots and the eudicots. The second main clade consists of rosid, asterid and *Viola* representatives. This indicates that another duplication event in the CHS gene family took place during the split between the core eudicots and rosids/asterids. Similar results were also found by Huang et al. (2004).

Yang and Gu (2006) also describe multiple rounds of CHS gene duplications during the evolution of the angiosperms. According to these authors, the most ancestral gene lineages originated during the divergence of different plant families, such as Solanaceae, Convolvulaceae and Asteraceae in a first round of duplications. Derived CHS genes further duplicated and diverged, which led to the occurrence of various CHS plant family specific genes in subsequent rounds of duplication.

The CHS intron gene lineage tree suggests that three CHS copies are present in *Viola*, whereas the CHS exon 2 gene lineage tree (data not shown) only contains two copies. The oldest duplication event in clade I (fig. 5) suggests the possibility of the presence of a third copy of CHS in *Viola*. This might also explain why CHS intron copies 1 and 2 have a close resemblance. The fact that not all copies were retrieved does not mean they are not there

but could explain why taxa from different sections end up nested in section *Viola* in the reconciled tree. Our interpretation of the CHS paralogues in *Viola* is different from Farzad et al. (2003, 2005). We consider the CHS paralogues in *V. cornuta* to be different alleles whereas the latter study identified them as different copies. The plant analyzed by Farzad et al. (2003, 2005) was a garden cultivar of hybrid origin. The nominal species is in itself a high-polyploid (Marcussen et al., forthcoming). Furthermore, our analyses of a larger sample of different *Viola* lineages showed that the interpretation of Farzad et al. (2005) was incomplete. Farzad et al. (2005) showed that the three CHS paralogous in *V. cornuta* are all still expressed and fully functional. Expression patterns were found to be slightly different, which might indicate subfunctionalization. Subfunctionalization of duplicate CHS genes in angiosperms appears to have happened by differentiation of their regulatory elements (Yang and Gu, 2006). It would be interesting to further investigate the mechanisms of subfunctionalization in a wider array of *Viola* species.

Acknowledgements

We would like to acknowledge T. Marcussen for providing material and for his valuable feedback on earlier versions of this manuscript. Furthermore, we like to acknowledge P. van Beers, R.L. Eckstein, H. den Held, F. Hellberg, W.J. Holverda, the late R. van der Meijden, R. van Moorsel, I. Nordal, P.B. Pelser, N. Schidlo, E.J. Weeda, and B. Wijlens for providing material and assistance in the field; Kew Botanical Gardens for providing DNA samples; and M.C.M. Eurlings, B.J. van Heuven, L. McIvor, S. Bollendorff, and D.L.V. Co for technical support. We would like to acknowledge the following organizations and departments for providing access to plant material and collecting permits: Staatsbosbeheer, Natuurmonumenten, Overijssels Landschap, and Kalmar county environmental department.