Available online at www.sciencedirect.com
SCIENCE DIRECT®
Organisms, Diversity & Evolution 5 (2005) 275–283
ORGANISMS DIVERSITY & EVOLUTION
www.elsevier.de/ode

Allopolyploid evolution in Geinae (Colurieae: Rosaceae) – building reticulate species trees from bifurcating gene trees

Jenny E.E. Smedmark[a,b,*], Torsten Eriksson[a,b], Birgitta Bremer[a,b]

[a]*Bergius Foundation, Royal Swedish Academy of Sciences, Box 50017, SE-104 05 Stockholm, Sweden*
[b]*Department of Botany, Stockholm University, SE-106 91 Stockholm, Sweden*

Received 17 September 2004; accepted 27 December 2004

Abstract

A previous phylogenetic study of paralogous nuclear low-copy granule-bound starch synthase (GBSSI) gene sequences from polyploid and diploid species in Geinae indicated that the clade has experienced two major allopolyploid events in its history. These were estimated to have occurred several million years ago. In this extended study we test if the reticulate phylogenetic hypothesis for Geinae can be maintained when additional sequences are added. The results are compatible with the hypothesis and strengthen it in minor aspects. We also attempt to identify extant members of one of the inferred ancestral lineages of the allopolyploids. On the basis of previous molecular phylogenies, one specific group has been proposed to be the descendants of this taxon. However, none of the additional paralogues belong to this ancestral lineage. A general method is proposed for converting a bifurcating gene tree, with multiple paralogous low-copy gene sequences from allopolyploid taxa, into a reticulate species tree.

Keywords: Geinae; Rosaceae; Allopolyploidy; Reticulate phylogeny; GBSSI

Introduction

The group Geinae Schulze-Menz has been suggested to have been shaped by allopolyploidy (Gajewski 1957, 1958), an evolutionarily important process among plants (Stebbins 1971; Grant 1981) involving interspecific hybridisation followed by chromosome doubling. To address this question, a phylogenetic study of low-copy granule-bound starch synthase (GBSSI) gene sequences from species in this clade was performed (Smedmark et al. 2003). When reconstructing allopolyploid speciation with single- or low-copy nuclear gene sequences, paralogous copies from a putative allopolyploid species are analysed phylogenetically together with sequences from closely related species. This method provides a direct reconstruction of phylogenetic relationships of parental lineages of allopolyploid taxa (Sang and Zhang 1999). In an allopolyploid, there are homoeologous copies of the gene, contributed by different ancestral taxa. These gene copies will be sister to orthologues in the ancestral taxa in a phylogenetic tree, rather than to each other. Autopolyploids, on the other hand, have paralogous loci that will be more closely related to each other than to orthologous sequences in closely related species. Several studies have used this method successfully to infer complex reticulate relationships among plant species, e.g., in *Paeonia* (Sang

*Corresponding author. Current address: Swedish Museum of Natural History, Department of Phanerogamic Botany, Box 50007, SE-104 05 Stockholm, Sweden.
E-mail address: jenny.smedmark@nrm.se (J.E.E. Smedmark).

1439-6092/$ - see front matter
doi:10.1016/j.ode.2004.12.003

and Zhang 1999), *Gossypium* (Cronn and Wendel 2004), *Glycine* (Doyle et al. 2004), and *Elymus* (Mason-Gamer 2004). The study of Geinae also provided strong evidence for reticulate historical relationships between lineages (Smedmark et al. 2003). A hypothesis suggesting two allopolyploidisation events with major impact on extant species diversity in the group was formulated, based on the GBSSI-1 gene tree. The analysis included paralogous copies from seven polyploid species in Geinae, along with copies from two diploid species in this group. These diploids had been shown to belong to the same clade (Smedmark and Eriksson 2002), which is the only one in Geinae known to contain extant diploids. This previous study only identified one of the two original ancestral lineages of the polyploids. Although the phylogenetic position of the other lineage can be inferred from the gene tree, it turned out that no extant representatives were included in the study. With reference to a previous phylogenetic study with wider taxon sampling (Smedmark and Eriksson 2002), a specific clade, the sister group of the remainder of Geinae, was suggested to be potential descendants of this second parental lineage. In this extended study we test whether the hypothesis proposed by Smedmark et al. (2003) about reticulate relationships within Geinae holds true with increased sampling of GBSSI paralogues. We also attempt to identify the unknown ancestral lineage of the allopolyploids.

It is worth noting that, despite the fact that low-copy gene sequences have been used successfully in several studies to infer the ancestry of polyploids, no method for converting bifurcating gene trees into reticulate species tree has been published. We here describe the method that we have used, a method that seems to be generally applicable to this type of problems.

Materials and methods

Samples, DNA extraction, amplification, cloning, and sequencing

Four species in Colurieae and two Rosoideae species outside this clade were selected (Table 1) to be incorporated in an existing data set (Smedmark et al. 2003). One species from the suggested paternal lineage was included, the decaploid *Oncostylus leiospermus*, as well as the closest relatives of Geinae, *Sieversia* Willd. and *Fallugia* Endl., and the tetraploid *Novosieversia glacialis*. Based on data from the *trnL–trnF* region (Smedmark and Eriksson 2002), the last of the above species was placed in a group that would correspond to the hypothesised allopolyploid clade.

DNA extraction, PCR amplification, and cloning followed the procedures described by Smedmark et al. (2003). Sampling was limited to available living specimens due to difficulties in amplifying GBSSI using extractions from dried material. The amplification of GBSSI-1 for *N. glacialis*, *O. leiospermus*, and *Sieversia pentapetala* was carried out with the primers 1F1C and 9R1C (Smedmark et al. 2003), whereas *Fallugia paradoxa*, *Filipendula vulgaris*, and *Fragaria vesca* were amplified with the primers 1F and 9R (Alice 1997).

Sequencing reactions were performed with a DYEnamic ET termination cycle sequencing premix *kit*

Table 1. List of taxa

Species	Voucher	Origin	Clone	No. clones screened	EMBL accession
Fallugia paradoxa (D. Don) Endl.	T. Eriksson No. 796 (SBT)	USA (Colorado)	*F. paradoxa* 2–2	7	AJ871485
Sieversia pentapetala (L.) Greene	T. Eriksson No. 749 (SBT); cult. Göteborg Botanic Garden	Unknown	*S. pentapetala* 2–2	5	AJ871484
Novosieversia glacialis (Adams) F.Bolle	A. Batten	USA (Alaska)	*N. glacialis* 1–2	22	AJ871488
Oncostylus leiospermus (Petrie) F. Bolle	M. Chase; cult. Royal Botanic Gardens Kew	New Zealand	*O. leiospermus* 1–3, *O. leiospermus* 1–6, *O. leiospermus* 2–2	14	AJ871490, AJ871489, AJ871491
Fragaria vesca L.	T. Eriksson & J.E.E. Smedmark 43 (SBT)	Sweden	*F. vesca* 3–2	2	AJ871486
Filipendula vulgaris Moench.	T. Eriksson 821 (SBT)	Sweden	*F. vulgaris* 3–3	1	AJ871487

The list follows the classification of Bolle (1933) and includes information on vouchers, origins, clones included in the analyses, numbers of GBSSI-1 clones screened, and EMBL accession numbers of the sequences (for information on *Rosa multiflora* and *Rubus odoratus* see Evans et al. 2000, for the remaining sequences see Smedmark et al. 2003).

Table 2. Sequencing primers used for *Fragaria vesca* and *Filipendula vulgaris*

Primer	Sequence (5′–3′)
Ros1_4F	GGACAACCAACTTAGATTCAG
Ros1_4R	GCTGAATCTAAGTTGGTTGTCC
Ros1_7F	GCTTACCAAGGCAGATTTGC
Ros1_7R	AATGCAAATCTGCCTTGG

(Amersham Pharmacia Biotech), on a MegaBACE 1000 capillary machine (Amersham Pharmacia Biotech). Protocols followed those provided by the manufacturer. The four Colurieae species were sequenced with the six primers used by Smedmark et al. (2003). For the two other species new sequencing primers were constructed based on GBSSI-1 sequences from *Waldsteinia*, *Rubus*, *Rosa*, *Physocarpus*, and *Prunus* (Table 2).

Phylogenetic analyses

The Staden package (Staden 1996) was used for sequence editing and assembly. Multiple sequence alignment was performed using ClustalX (v1.81; Thompson et al. 1997) with the default settings, followed by corrections made by eye in the sequence alignment editor Se–Al (v2.0a11; Rambaut 1996). All aligned positions were included in the analyses. Maximum likelihood analyses were performed with PHYML (v2.0.3; Guindon and Gascuel 2003). In addition, data were analysed using a Bayesian approach with MrBayes (v3.0; Huelsenbeck and Ronquist 2001). The model of sequence evolution was chosen based on results from Modeltest (v3.06; Posada and Crandall 1998). The Hierarchical Likelihood Ratio Tests and Akaike Information Criterion both selected the Hasegawa–Kishino–Yano (HKY; Hasegawa et al. 1985) nucleotide substitution model. Among-site rate variation was allowed to follow a gamma distribution (Yang 1996); in all analyses the default setting of the different programs with four rate categories was used. All trees were rooted on the branch leading to *Filipendula vulgaris*, based on the results by Morgan et al. (1994).

A search for optimal trees was performed with PHYML. The algorithm implemented in this program starts from an initial neighbour-joining tree and adjusts tree topology and branch lengths simultaneously during hill-climbing, which makes it fast. The model parameters, transition/transversion ratio; and gamma shape parameter were estimated during the hill-climbing process.

Support for each node was assessed using bootstrap and Bayesian analyses. Nonparametric bootstrap proportions (BPs; Felsenstein 1985) were estimated with PHYML from 10,000 pseudo-replicate data sets, which were assembled by SEQBOOT in PHYLIP (v3.5; Felsenstein 1993). The bootstrap majority-rule consensus trees were constructed by CONSENSE (also PHYLIP). This process was automated using the BootPHYML (Nylander 2003), which ties the bootstrap analysis together. The same model and parameter settings as above were used.

Bayesian posterior probabilities for clades (PPs; Larget and Simon 1999) were estimated with MrBayes. The Monte Carlo–Markov chain was run for 1,000,000 generations, and sampled every 10 generations. Log-likelihood values were considered to be stable after 7560 generations, and therefore the last 99,244 sampled trees were used to construct a majority-rule consensus tree and estimate Bayesian PPs. To make sure that the Markov chain did not get stuck in one region of parameter space in this analysis, failing to visit other regions where the posterior is as high or higher, we performed two additional analyses with the same settings, each one starting from a random tree. To retrieve the single sampled tree with the highest posterior probability, the output parameter and tree files were scanned using a script that is available from the second author.

Results

DNA sequences

Six new GBSSI-1 sequences from four Colurieae species and two from other Rosoideae species (Table 1) were analysed together with 24 sequences from a previous study (Smedmark et al. 2003). The sequences ranged in length from 1754 to 1913 base pairs. The exons were free of stop codons, except for two of the *O. leiospermus* sequences (clones 1–3 and 2–2), in which deletions in the exon sequence have damaged the reading frame. The aligned sequence length was 2116 base pairs, of which 0.2% was scored as missing data, and 1261 were variable sites.

Phylogenetic analyses

The best tree from the PHYML analysis (ln-likelihood –14243.289) is shown in Fig. 1. The one tree with the highest posterior probability that was sampled in the three Bayesian analyses (ln-likelihood –14252.489) had a topology identical to that of the optimal PHYML-tree. Nonparametric BPs based on 10,000 pseudo-replicates in PHYML are shown in Fig. 2. The three Bayesian analyses found the same nodes with high PPs, and the topologies of the majority-rule consensus trees were identical to that from the bootstrap analysis in

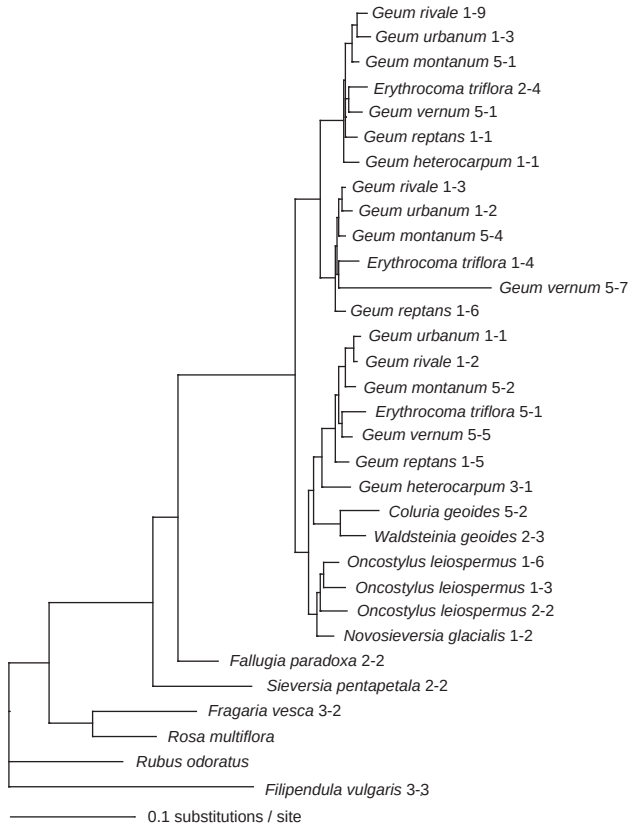

Fig. 1. The best tree from the PHYML analysis (ln-likelihood –14243.288), with branch lengths drawn in proportion to the amount of change. Numbers after species names are clone identifiers.

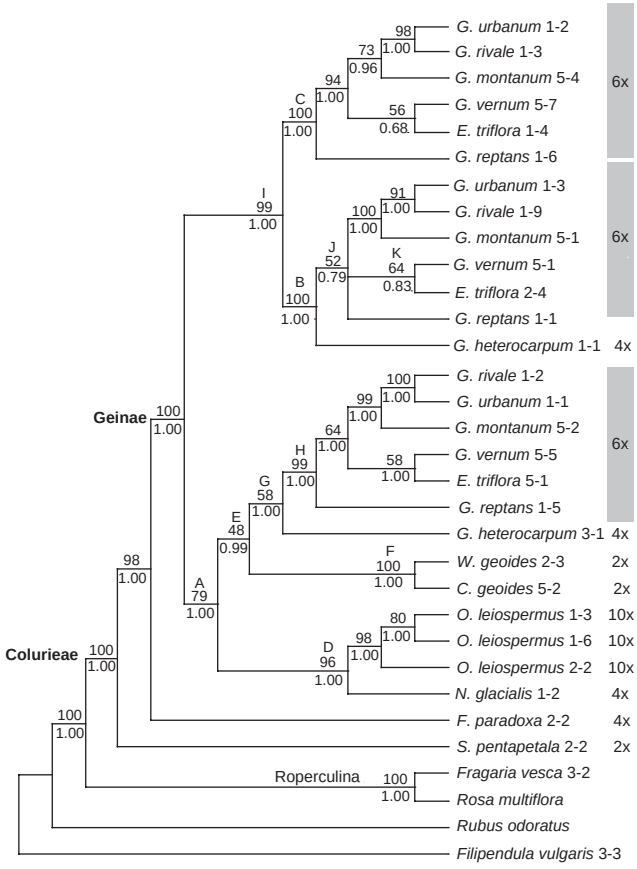

Fig. 2. The optimal tree, with support values indicated for each node. Bootstrap proportions obtained with PHYML, based on 10,000 pseudo-replicates, are shown above branches, and Bayesian posterior clade probabilities below. Nodes discussed in the text are marked with capital letters. Numbers after species names are clone identifiers, followed by ploidy level.

PHYML. All nodes but five in the majority-rule consensus trees had a posterior probability of 1.00 in the three analyses. The support for these five nodes differed by 0.01–0.05 between the separate runs. The values from one of the analyses are shown in Fig. 2.

In the optimal tree from the PHYML analysis (Figs. 1 and 2), *Rubus odoratus* is the sister of a clade comprising the two lineages Colurieae and Roperculina. Within Colurieae (Fig. 2), *S. pentapetala* is sister to the rest of the clade, in contrast to previous analyses (Smedmark and Eriksson 2002) that suggested *Fallugia* as the sister to the remainder of the group. In Geinae (Fig. 2) there are three major clades of GBSSI-1 clones: A, B and C. Clade A contains paralogues from all included Geinae species, whereas the other two, which are sister groups, only consist of additional paralogues from polyploids. Paralogues from each of the six hexaploid species can be found in all of the three major clades. Clades C, H, and J contain only copies from hexaploid species. Two of these have identical topologies (C and H), whereas the third is less resolved but congruent with the others. The two paralogues from the tetraploid species, *G. heterocarpum*, also appear in different parts of the tree (clades B and G). Each one is sister to a clade of copies from hexaploid species. Within clade A, the three paralogues from *O. leiospermus*, together with the one from *N. glacialis*, form the sister-group (clade D) of the remainder of the clade (clade E).

Discussion

Allopolyploidy in Geinae

The topology of the GBSSI-1 phylogeny with extended species sampling is compatible with the hypothesis about reticulate species history in Geinae (Smedmark et al. 2003). Two of the loci present in the hexaploids are more closely related to paralogues from other species than they are to each other (Fig. 2, nodes H and J). Homoeologues of the tetraploid *Geum heterocarpum* are sister to each of these. Therefore, two of the subgenomes of the hexaploids were hypothesised to have been inherited from the ancestral lineage of

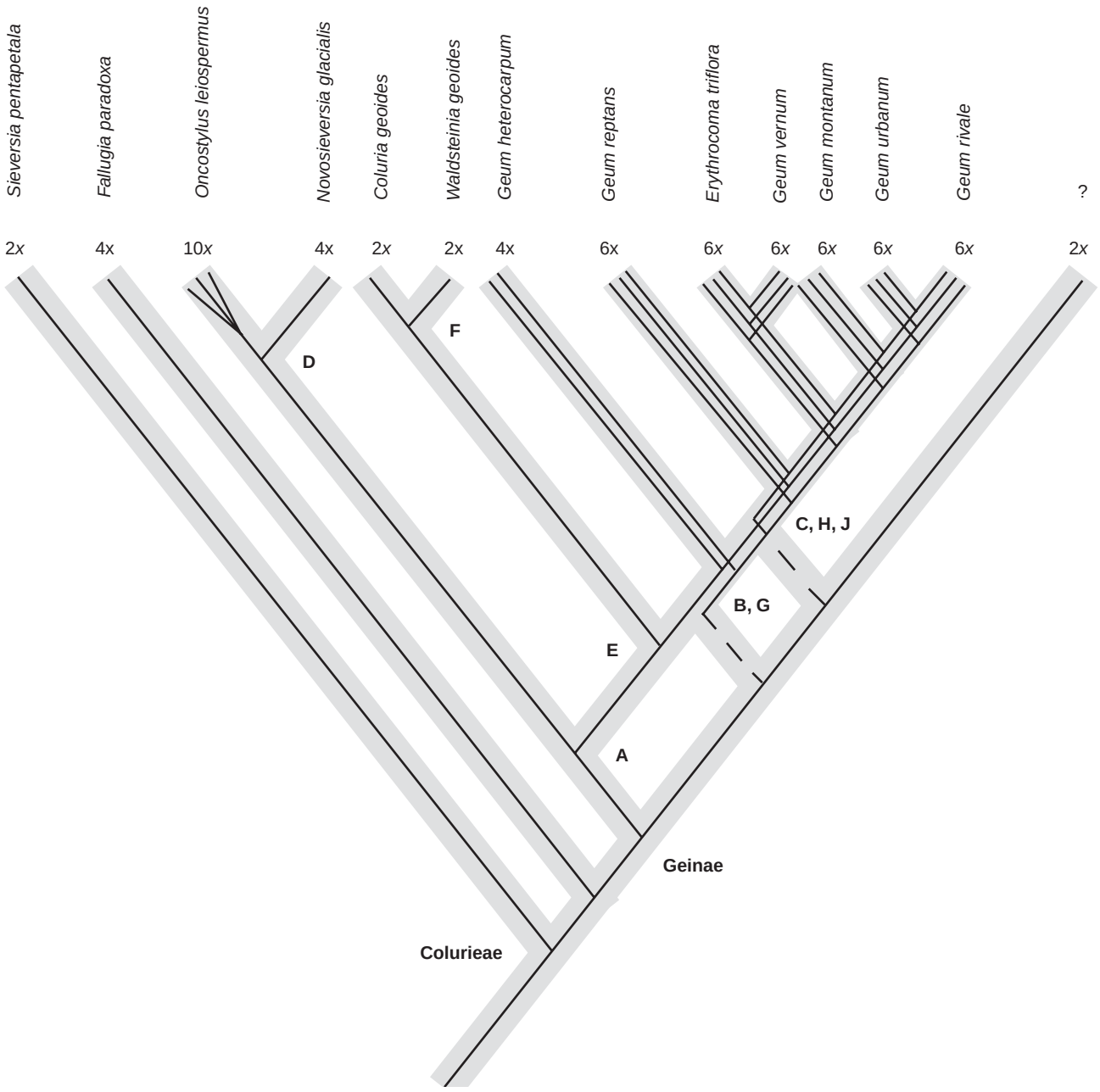

Fig. 3. Hypothesis of reticulate historical relationships within Colurieae. Thick grey lines represent organismal lineages, thin black lines the nuclear low-copy gene GBSSI-1. The phylogeny includes two instances where new lineages have arisen as a result of allopolyploidy. Capital letters refer to clades in Fig. 2.

G. heterocarpum (Fig. 3). The only group within Geinae known to include diploids, here represented by *Waldsteinia geoides* and *Coluria geoides*, (Fig. 2, node F) is the sister group of one of these polyploid clades (node G). Hence its ancestral lineage was hypothesised to have been one of the parents in an initial hybridisation that gave rise to the tetraploid lineage, represented by *G. heterocarpum* (Fig. 3). The other two loci of the hexaploids, clades C and J, were hypothesised to have originated in an unidentified diploid ancestral lineage (Fig. 3). Three nodes are present in the maximum likelihood tree (Fig. 2, nodes G, J, and K) which were not supported in the previous analysis (Smedmark et al. 2003), and one node present in that analysis is missing here. All of these changes strengthen the hypothesis about allopolyploid speciation.

Based on the evidence presented here, it is not possible to identify any extant members of the other of the two diploid lineages involved in the origin of the allopolyploids. The hypothesis suggesting that this ancestor belonged to the *Oncostylus* lineage (Smedmark et al. 2003), which is the sister group of the remainder of

Geinae (Smedmark and Eriksson 2002), can, however, not be fully dismissed. In the previous GBSSI study (Smedmark et al. 2003), one clade (corresponding to Fig. 2, node E) was congruent with the tree based on the *trnL*/*trnF* region of the chloroplast (Smedmark and Eriksson 2002), and thus was assumed to reflect the phylogeny of homoeologues of maternal descent. Another clade (Fig. 2, node I) comprised two other GBSSI-1 loci. These were suggested to be derived from the pollen parent lineage involved in the allopolyploidisations. If included in the analysis, a paralogue from the paternal lineage would be the sister of either clade I or C (Figs. 2 and 3), depending on whether the species diverged before or after the first allopolyploidisation event. Instead, the *Oncostylus* paralogues in the present study are found in clade D, which is the sister-group of clade E (Figs. 2 and 3). Only three distinct paralogues were found of this species, although the species is decaploid and thus might be expected to have five GBSSI-1 loci. The fact that they form a group indicates that the species, at least in part, is of autopolyploid origin. In the same clade the only identified paralogue of the tetraploid *N. glacialis* is also found. As a consequence of the fact that all possible paralogues of *O. leiospermus* or *N. glacialis* were not found, the hypothesis can neither be falsified nor verified. There is a possibility that there are homoeologous copies of the gene from these species belonging to the other ancestral lineage of the allopolyploids. If instead all five putative paralogues of the decaploid *O. leiospermus*, or the two of the tetraploid *N. glacialis*, had been found to belong to clade D, the ancestral lineage of these species would have been shown not to have been involved in the origin of the allopolyploids. Considerable effort was put into finding additional paralogues. For *N. glacialis*, 22 clones from six separate PCRs were screened. If both paralogues amplify and clone equally well, sampling 22 colonies results in a 4.8×10^{-7} probability of not sequencing both paralogues by chance. For *O. leiospermus*, 14 clones from five PCRs were screened, which gives a probability of 1.0×10^{-7} of not finding all five paralogues, provided that they amplify and clone equally well. There are many possible explanations to why not all putative paralogues were found. For example, they may have diverged too much for the primers to match, in which case they would not be amplified; they also could have been deleted from the genome, undergone homogenisation, or we may have failed to sample them by chance.

Although the second diploid ancestral lineage may be extinct, there remains a possibility that it just has not been sampled for this study. If the latter is the case, we suggest that possible candidates may be found among the species that have been classified in *Coluria* or *Acomastylis*. Distinction between these two genera has not always been straightforward (Bolle 1933). While the polyploid *Acomastylis* has been shown to be polyphyletic (Smedmark and Eriksson 2002), the diploid *Coluria* is likely to be monophyletic based on morphological evidence. It remains possible that some species classified in one of these groups actually belong to the inferred diploid ancestral lineage.

All the analyses in this paper give strong support (Fig. 2; BP 98, PP 1.00) to a topology that differs from previously reported results in the respect that *Sieversia*, rather than *Fallugia*, is the sister of the remainder of Colurieae. The *trnL*/*trnF* and ITS regions both render moderate or low support for the monotypic *Fallugia* being the sister of Colurieae, whereas *Sieversia*, consisting of the two species *S. pentapetala* and *S. pusilla*, is resolved as the second branch of extant species in Colurieae (Smedmark and Eriksson 2002). The second major paralogue of GBSSI found in Rosaceae (Evans et al. 2000), GBSSI-2, also supports this latter topology (Smedmark, unpublished data). The different topology presented in this paper is also not retrieved when introns are pruned from the data set (not shown). The two *Sieversia* species are diploid, whereas *F. paradoxa* is tetraploid, but only one paralogue each of GBSSI-1 and GBSSI-2 has been found and sequenced for the latter. Seven GBSSI-1 clones from *F. paradoxa* were screened, in an attempt to find the inferred second paralogue. This corresponds to a 1.6% probability of not sequencing both paralogues by chance. However, all seven clones were identical. The divergent result of GBSSI-1 presented here is difficult to explain based on available evidence. Visual comparison of the GBSSI-1 paralogue with the GBSSI-2 paralogue from *S. pentapetala* shows that, throughout the sequence, each one is more similar to the corresponding paralogue of *F. paradoxa* than they are to each other. This indicates that recombination between the two *S. pentapetala* paralogues has not taken place. Possible scenarios, which cannot be addressed in detail here, may involve, e.g., hybridisation or lineage sorting.

A method for converting a bifurcating gene tree into a reticulate tree

Commonly used algorithms for phylogenetic inference reconstruct strictly diverging relationships among ancestral lineages. Therefore the process of converting a gene tree, with multiple paralogous loci from polyploids, into a reticulate taxon tree may appear to be somewhat arbitrary. Here we describe, step by step, the method that we have used to accomplish this.

The first step is to determine whether a reticulate tree is the best description of species relationships. To do this, the different gene copies from each polyploid taxon are located. If copies from the same species form a group, they may either represent allelic variation or

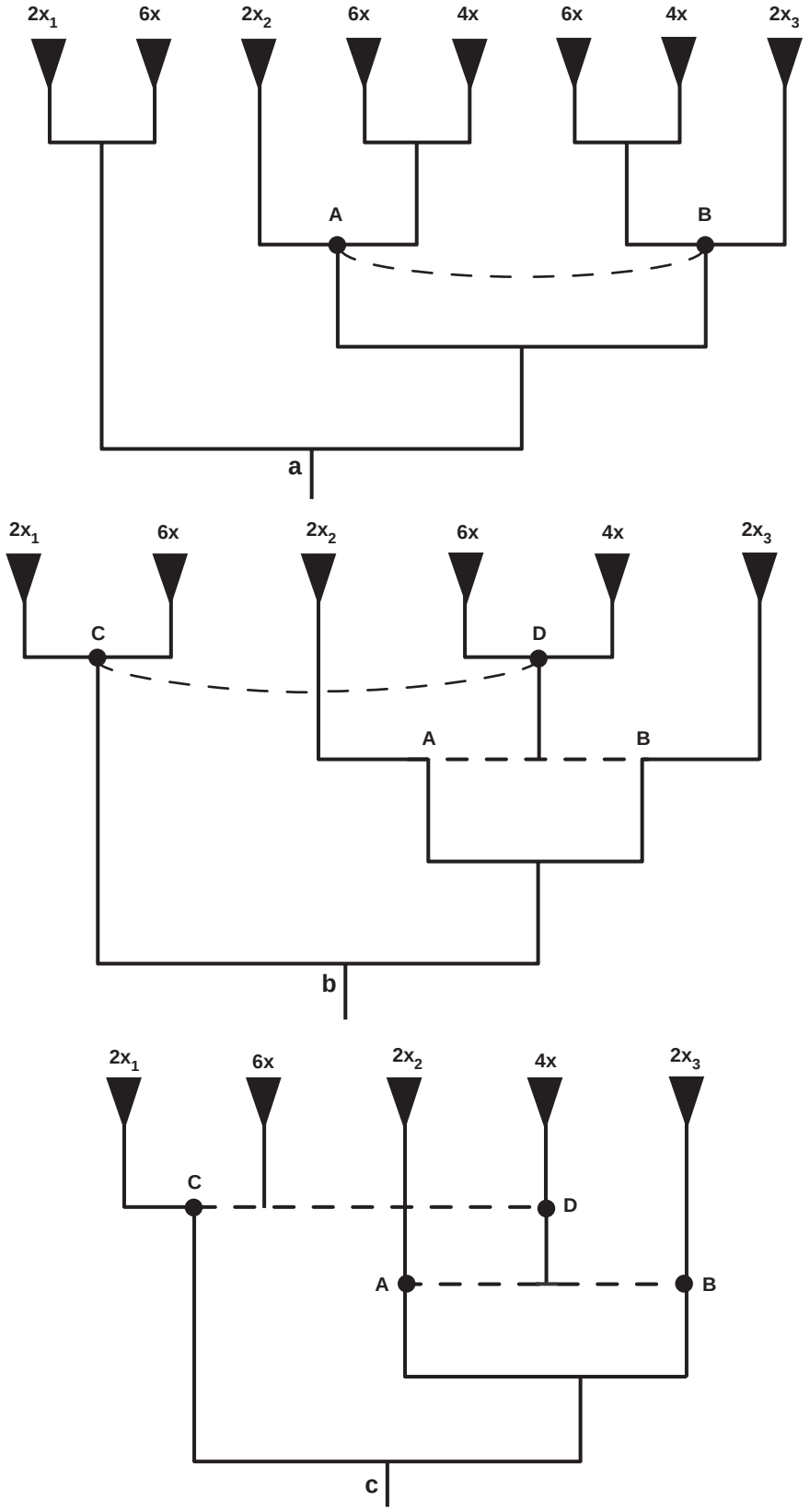

Fig. 4. Graphical representation of the method used for transforming a gene tree of allopolyploids into a reticulate organismal tree. In (a), the nodes A and B have been identified as connection points (lowermost lineage splits of primary polyploid with lower ploidy level). In (b), these are joined by a dashed line, and the clades including the primary polyploid are combined and moved onto the connection. Also, nodes C and D are indicated as the next points of connection as in (a) (lowermost lineage splits of secondary polyploid with lower ploidy level). In (c), nodes C and D have been joined by a dashed line and the hexaploid clade moved onto the connection, completing the reticulate tree.

duplicate loci. The latter may be due to either gene duplication or genome duplication, autopolyploidy. It is sometimes difficult to distinguish alleles from loci, but the variation between loci is normally considerably greater than the variation found among alleles. A Southern blot can usually show whether several loci are present in the taxon, but further sampling may also help to sort out relationships among multiple copies. For a review on gene level processes, acting on alleles and paralogues, that may lead to erroneous conclusions about taxon phylogeny, see Doyle and Davis (1998). If copies from the same species instead appear in different parts of the tree, it is first determined whether they occur in congruent clades. If this is the case, and the clades are sister groups, they represent different paralogues that have originated in a gene or genome duplication. If the clades, on the contrary, are found to be more closely related to copies from different species, the paralogues have originated in separate lineages and have been brought together in the common ancestor of the group, e.g., as a result of hybridisation. The same is true if the copies are not in congruent clades, but each one is found to be more closely related to taxa of lower ploidy level than to each other. Both cases indicate that allopolyploidy has taken place, in which case historical relationships among taxa are best represented by a reticulate phylogeny.

To construct a reticulate tree, start by selecting a species, or group of species, identified as being of potential allopolyploid origin according to the reasoning above. This taxon should be of the second-lowest ploidy level present in the gene tree. We call this a primary polyploid. To begin tracing the ancestry of this primary polyploid taxon, start at the root of the tree and follow the branches upwards to the point where the lineage of a copy from the primary polyploid first branches off from a lineage of lower ploidy level (Fig. 4a, node A). Note that both the branch leading to the primary polyploid and its sister group may include copies of higher ploidy levels. Thus, it is the split between the least inclusive clades including a primary polyploid copy, and that including a copy of lower ploidy level, that should be found. Return to the root of the tree and follow the branches upwards as before, to a second point where another copy of the primary polyploid branches off from a copy of lower ploidy level (Fig. 4a, node B). Connect these two points by a line, indicating a hybridisation event (Fig. 4a, dashed line). The two branches leading to the primary polyploid are then combined into a single one, and moved onto the connecting hybridisation line. Should the copies of a primary polyploid taxon be found in congruent groups that include copies of higher ploidy level, the entire groups are merged and moved onto the connecting line (Fig. 4b, node D). When all copies of the first primary polyploid taxon are treated, continue with any remaining taxa of the same ploidy level. Thereafter, proceed with taxa of the next ploidy level (secondary polyploids) as before. Find the point where the lineages of the copies of a secondary polyploid branch off from lower ploidy levels (Fig. 4b, nodes C and D). Connect these with a line and move the combined polyploid taxon onto the connecting branch (Fig. 4b, c). Repeat this process until all multiple copies have been replaced by taxa.

In some cases, as in the present study, representatives of ancestral lineages may be missing from the analysis. This makes the reconstruction of the taxon tree more difficult. However, in such cases the points of reticulation (i.e. where to attach the connecting hybridisation lines) may be inferred from the points of attachment of the putative allopolyploid terminal. For example, in the gene tree in Fig. 2, node E should be connected with node I, and the congruent clades G and B (including the primary polyploid) combined. To continue with the secondary polyploid, move to the next ploidy level split, the now combined node H/J, and connect it to node C. The last step is to combine this third congruent clade of copies from the secondary (hexaploid) taxa with the other two (nodes H/J/C). This leads to the conclusion that a diploid taxon donated the genome C, and this inferred lineage may be added to the species tree for clarity, as in Fig. 3.

We have tested this method on fairly complicated examples of hypothetical gene trees derived from known reticulate taxon trees, and it has recovered the correct taxon tree in all cases. Therefore, the described method may be of general relevance for solving these kinds of problems.

Acknowledgements

The authors would like to thank the Göteborg Botanic Garden for providing material of *Sieversia pentapetala*, Alan R. Batten for material of *Novosieversia glacialis*, and Nahid Heidari for help in the lab.

References

Alice, L.A., 1997. Molecular phylogenetic systematics of *Rubus* (Rosaceae). Ph.D. dissertation, University of Maine, Orono, Maine.

Bolle, F., 1933. Eine Übersicht über die Gattung *Geum* L. und die ihr nahestehenden Gattungen. Feddes Repert. Beih. 72, 1–119.

Cronn, R., Wendel, J.F., 2004. Cryptic trysts, genomic mergers, and plant speciation. New Phytol. 161, 133–142.

Doyle, J.J., Davis, J.I., 1998. Homology in molecular phylogenetics: a parsimony perspective. In: Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), Molecular Systematics of Plants II. Kluwer Academic Publishers, Massachusetts, pp. 101–131.

Doyle, J.J., Doyle, J.L., Rauscher, J.T., Brown, A.H.D., 2004. Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (*Glycine* subgenus Glycine). New Phytol. 161, 121–132.

Evans, R.C., Alice, L.A., Campbell, C.S., Kellogg, E.A., Dickinson, T.A., 2000. The granule-bound starch synthase (GBSSI) gene in the Rosaceae: multiple loci and phylogenetic utility. Mol. Phylogenet. Evol. 17, 388–400.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Felsenstein, J., 1993. PHYLIP, version 3.5. Computer program available at http://evolution.gs.washington.edu/phylip.html>

Gajewski, W., 1957. A cytogenetic study on the genus *Geum*. Monogr. Bot. 4, 3–414.

Gajewski, W., 1958. Evolution in the genus *Geum*. Evolution 13, 378–388.

Grant, V., 1981. Plant Speciation. Columbia University Press, New York, pp. 193–347.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating the human–ape split by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160–174.

Huelsenbeck, J.P., Ronquist, F.R., 2001. MRBAYES: Bayesian inference of phylogeny. BioInformatics 17, 754–755.

Larget, B., Simon, D., 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16, 750–759.

Mason-Gamer, R.J., 2004. Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. Syst. Biol. 53, 25–37.

Morgan, D.R., Soltis, D.E., Robertson, K.R., 1994. Systematic and evolutionary implications of *rbcL* sequence variation in Rosaceae. Am. J. Bot. 81, 890–903.

Nylander, J.A.A., 2003. BootPHYML. Computer program distributed by the author. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Rambaut, A., 1996. Se-Al: Sequence Alignment Editor, version 2.0a11. Computer program, available at http://evolve.zoo.ox.ac.uk/>

Sang, T., Zhang, D., 1999. Reconstructing hybrid speciation using sequences of low copy nuclear genes: hybrid origins of five *Paeonia* species based on *Adh* gene phylogenies. Syst. Bot. 24, 148–163.

Smedmark, J.E.E., Eriksson, T., 2002. Phylogenetic relationships of *Geum* (Rosaceae) and relatives inferred from the nrITS and *trnL–trnF* regions. Syst. Bot. 27, 303–317.

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.

Staden, R., 1996. The Staden sequence analysis package. Mol. Biotech. 5, 233–241.

Stebbins, G.L., 1971. Chromosomal Evolution in Higher Plants. Edward Arnold, London.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Yang, Z, 1996. Among-site variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11, 367–371.