2010-06-11

Infrageneric Relationships Within *Collomia* (Polemoniaceae)

Eric Stewart Green
Brigham Young University - Provo

Infrageneric Relationships Within *Collomia* (Polemoniaceae)

Eric S. Green

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Leigh A. Johnson, Chair
Keith A. Crandall
Joshua A. Udall

Department of Biology

Brigham Young University

August 2010

ABSTRACT

Infrageneric Relationships Within *Collomia* (Polemoniaceae)

Eric S. Green

Department of Biology

Master of Science

Traditional evolutionary models depict evolution as a bifurcating pattern with a single ancestor diverging to form two lineages. However, reticulate species resulting from hybridization and recombination have unique histories shared with two independent lineages, not one. Accounting for the genetic histories of reticulate species increases the power and ability to recover biologically meaningful relationships. The genus *Collomia* (Polemoniaceae) is used to explore issues of reticulation and the importance of accounting for gene histories in a phylogenetic analysis. The issue of reticulation within species trees is discussed with a multilabeled, network approach being explored to better represent the genus's evolutionary history.

Wherry's hypotheses regarding the relationships that exist within *Collomia* are addressed and the need for a new intrageneric section is recognized based on support from multiple, independent genes and morphology. Sections *Collomiastrum* and *Courtoisia* remain as previously circumscribed. Section *Collomia* is modified with removal of two species, *C. grandiflora* and *C. biflora* from the section and by erecting a new section, *Calyperona*. A morphological key is included for each section and their corresponding species, followed by sectional discussions. Finally, the evolution of lifecycle duration, seed morphology, and pollen morphology are discussed based on the phylogeny of the genus.

Keywords: allopolyploidy, character evolution, *Collomia*, infrageneric classification, introgression, *Navarretia*, Polemoniaceae, reticulation, taxonomy

ACKNOWLEDGMENTS

Leigh, thank you for taking a high school teacher under your wing and providing me with this wonderful opportunity. Though you wish you could have done more with me, allowing me to blindly push myself through the ins and outs of phylogenetics and its corresponding software have allowed me to learn more than you know. You have always been there when I got stuck or had questions. Thank you. You truly deserve more credit than is sometimes given.

I also wish to thank two professors that have inspired me through their passion for evolutionary biology – Drs. Keith Crandall and Mike Whiting. Thank you for teaching the future systematists at this wonderful university.

Finally, I would like to thank my wonderful family. Rachel, thank you for allowing me this opportunity. I could not have done this without your continued support and faith in me. You are amazing. Lauryn, Meggan, and Rylee – you girls are great and thank you for understanding when I have been busy with school. Now, how am I going to pay for this midlife crisis?

TABLE OF CONTENTS

LIST OF TABLES AND APPENDICES

LIST OF FIGURES

CHAPTER 1:

Trumpets and their Phylogenetic Coronation: Reconstructing Species-level Relationships and Historical Processes in *Collomia* (Polemoniaceae)

1. INTRODUCTION

Despite the widespread representation of phylogenetic histories using dichotomous trees, systematists have long recognized that evolutionary histories are not always dichotomous and that hybridization plays an important role alongside descent with modification in generating species diversity. In plants, chloroplast capture is well documented as a source of discordance between nuclear and plastid phylogenies (Okuyama et al., 2005; Soltis and Kuzoff, 1995); homoploid hybridization, introgression, and allopolyploidization are similarly recognized as important evolutionary processes that can contribute to discordance among nuclear genes (Kelly et al., 2010; Rieseberg and Brunsfeld, 1992; Wendel and Doyle, 2005). Because reticulation and subsequent processes such as recombination can affect overall tree topology (Vriesendorp and Bakker, 2005), elucidating the source of observed discordance among genetic data sets increases the ability to recover and depict meaningful evolutionary events in addition to phylogenetic relationships.

Cladistic approaches dealing with hybrids include appending known hybrids to phylogenetic trees produced from analyses that excluded the reticulate species (Hughes et al., 2002; Sang et al., 1995) or simply proceeding with a total evidence approach (Kluge, 1989; Petersen and Seberg, 2004). Neither of these approaches are fully satisfactory; the former method excludes data that may bear on relationships among the included taxa while the latter approach, by combining data that do not share a common history, violates the implicit assumption of phylogenetic analyses that data share a common history (de Queiroz et al., 1995) and creates homoplasy by forcing these phylogenetically independent data to share a common bifurcating history. A multilabeled tree approach (Brysting et al.,

2007; Huber and Moulton, 2006; Huber et al., 2006; Huson and Bryant, 2006; Linder and Rieseberg, 2004) overcomes both of these problems. Though presented initially as a means of effectively representing polyploid individuals, this approach should be equally effective for handling instances of recombination and gene tree - species tree discordance attributable to hybridization (Pririe et al., 2009).

The genus *Collomia* (Polemoniaceae) is an unlikely model for illustrating a multilabeled tree approach to reticulating phylogenetic relationships. Unlike some genera in Polemoniaceae such as *Phlox*, *Gilia*, and *Ipomopsis* where hybridization has long been recognized as important, diversification in *Collomia* has been considered strictly divergent (Grant, 1959; Wherry, 1944) and naturally occurring primary hybrids are unknown. Nevertheless, comparative DNA sequence studies with only partial sampling of *Collomia* demonstrate that two species are allotetraploids (Johnson and Johnson, 2006) and one of the perennial species has likely captured the chloroplast of an annual species (Johnson et al., 2008). Chloroplast DNA data also weakly show *Collomia* as paraphyletic (Johnson et al., 2008) —a surprising suggestion given several morphological traits, including chromosome number, calyx morphology, and explosive capsule dehiscence that appear synapomorphic for the genus.

In addition to postulating a divergent pattern of diversification in *Collomia*, Wherry (1944) also postulated that the broad-leaved perennial species were ancestral in this genus and that the annual, linear-leaved species were derived. Subsequent studies have characterized pollen exine sculpturing (Chuang et al., 1978), seed coat morphology (Hsiao and Chuang, 1981), and flavonoid/anthocyanin chemistry (Wilken et al., 1982) in the genus. These studies revealed discrepancies between observed patterns of variation and

the sectional classification of Wherry, but stopped short of revising the classification until more thorough cytogenetic and phylogenetic analyses were conducted. DNA sequence data provide such an opportunity. Given patterns of variation described above that were unexpected prior to comparative sequencing approaches, a study of phylogenetic relationships and character evolution in this genus is warranted using an approach that incorporates evidence of reticulation discovered through the phylogenetic reconstruction process.

Here, a variety of DNA sequence markers supplemented with morphological characters were used to reconstruct phylogenetic relationships in *Collomia*. Specifically, the objectives were to 1) assess the monophyly of *Collomia* in the context of its nearest relatives, *Allophyllum* and *Navarretia*; 2) assess the significance of congruence between DNA data partitions and localize the taxa and likely source (i.e., sampling error, lineage sorting, horizontal gene transfer, and so forth) of observed incongruencies; 3) use a multilabeled analysis approach to phylogenetic reconstruction for reconstructing the phylogenetic history of *Collomia* based on the multiple DNA datasets; and 4) use the phylogenetic framework constructed through this work to assess patterns of character evolution within the genus.

2. Materials and Methods

2.1 Taxon sampling

Thirty-three species were sampled for this study. All 15 species of *Collomia* were represented with a minimum of two populations for each species except *C. larsenii* and *C. rawsoniana*. *Collomia larsenii* is sometimes treated as a subspecies of *C. debilis*, which is well

sampled in this study, and *C. rawsoniana* has an extremely restricted distribution. Partial gene sampling of secondary populations for these two species showed no variation relative to the completely sampled populations. However, because the material of the secondary populations came from older herbarium tissue, not all of the nuclear genes were obtained and therefore these additional populations were not included in the analyses. Three (of five) species from the genus *Allophylum* and 15 (of 35) species of *Navarretia* were also sampled. *Allophyllum* is sister to *Collomia* + *Navarretia,* and these three genera form a well-supported clade in previous comparative DNA sequencing studies of Polemoniaceae (Johnson et al., 2008; Johnson et al., 1996). *Allophyllum* served as the outgroup for rooting purposes in all analyses and *Navarretia* was included to assess monophyly. The *Allophyllum* and *Navarretia* species sampled represent the taxonomic, morphological, and molecular diversity observed within these genera based on published studies (Spencer and Porter, 1997) and ongoing work in the Johnson lab.

2.2 DNA extraction, amplification, and sequencing

DNA was isolated from silica dried leaf samples using a modified CTAB protocol (Cullings, 1992; Doyle and Doyle, 1987). Several chloroplast and nuclear regions were amplified and sequenced that represent both linked (chloroplast) and putatively unlinked (nuclear) loci, a range of evolutionary rates, and different modes of inheritance (maternal versus biparental). These regions include seven chloroplast loci: 5' *trnK–matK* intron/intergenic spacer (*trnK*), 5' *matK* (*matK*), the *trnL–trnL–trnF* intron/intergenic spacer (*trnL*), *ycf6–psbM* intergenic spacer (*ycf6*), *psbM–trnD* intergenic spacer (*psbM*), *trnD–trnT* intergenic spacer (*trnD*), and the *trnS–G* intergenic spacer (*trnS*); the nuclear ribosomal ITS-1, 5.8s rDNA, and ITS-2 region (ITS); and portions of three low-copy nuclear genes primarily composed of introns but with some exons as

well, namely, isocitrate dehydrogenase (*idhA*), glyceraldehydes-3-phosphate dehydrogenase (*g3pdh*), and pistillata (*PI*).

DNA regions were amplified and sequenced using published locus specific primers for regions other than *PI* (DeMesure et al., 1995; Hamilton, 1999; Johnson and Johnson, 2006; Porter, 1996; Shaw et al., 2005; Strand et al., 1997; Taberlet et al., 1991; White et al., 1990), and primers developed in the Johnson lab by T. Weese (2004) specific for Polemoniaceae that amplify a 5' portion of the *PI* gene. The *PI* primers are PI-7F (5'-AGAGGAAAGATTGAGATAAAGAGG-3') and PI-450R (5'-TTCTCTTCCTCCARCATCATT-3') used for amplification and sequencing, with additional internal primers PI-900R (5'-ATCATTCTCTTTCTTGATCC-3') and PI-880F (5'-ATCCATGGACAGATCTGGTAA-3') used for sequencing if needed.

All PCR were conducted in 30 μL reaction volumes containing 3 μL 10× ammonium buffer with $MgCl_2$ included, 1.5 μL glycerol, 4.5 μL dNTP (1mM each), 0.9 μL of the specified forward and reverse primers (10 μM), 18.8 μL dH_2O, 0.3 μL of extracted DNA, and 0.1 μL of TAQ polymerase (5u/μL). The PCR profile employed consisted of 30 cycles of 1 min at 95°C, 1 min at 52°C (50° for PI), and 1 min at 72°. Amplified fragments were cleaned using Millipore plates and were sequenced (BigDye v.3, Applied Biosystems) and electrophoresed on an AB 3730xl automated sequencer in the DNASC at Brigham Young University. The majority of sequences were obtained cleanly, even from the low copy nuclear genes. However, when chromatograms gave evidence of sequence heterogeneity (e.g., in the case of heterozygosity or allopolyploidy), PCR products were cloned using TOPO-TA kits (Invitrogen Corp., Eugene, OR) and 3–8 colonies per cloned PCR were sequenced. Allelic divergence was low and in all cases multiple alleles within diploid species, when observed, coalesced within species; consequently, a

single allele was chosen arbitrarily for inclusion in the analysis. Some of the "allelic diversity" from clones was undoubtedly from polymerase errors and a consensus sequence generated across all variants was used to represent the species.

2.3 Sequence alignment and indel coding

Sequences were aligned using MUSCLE (Edgar, 2004) and adjusted manually using Se-Al (Edgar and Botzoglou, 2006; Rambaut, 2002), with attention given to mechanisms such as inversions, duplications, and recombination that may contribute to alignment ambiguities (Kelchner, 2000). Some segments were excluded from idhA (13 segments; 16 base pairs), g3pdh (2 segments; 91 base pairs) and PI (6 segments; 227 base pairs), such as single or dinucleotide strings of variable length, or less often, ambiguous alignment where hypervariability of length and nucleotides led to low confidence in homology assignments. Gap positions were treated as missing rather than as a fifth state, with indels subsequently coded using simple indel coding (Simmons and Ochoterena, 2000) as implemented in SeqState 1.41 (Muller, 2005).

2.4 Allopolyploidy and recombination detection

The parental types for each nuclear gene from the known allopolyploid species were identified by preliminary phylogenetic analyses. All sequences (regardless of ploidy level) from the nuclear regions were also assessed for recombination using Recombination Detection Program (RDP3beta) version 3.34 (Martin et al., 2005), with recombinant breakpoints identified by RDP3beta confirmed through visual comparison with putative parental types. "Ancestral" sequences were reconstructed from detected recombinants in two ways. First, for sequences

identified as recombinant between two taxa (complexity greater than this was not observed), portions of a sequence showing similarity to taxon-a were retained while replacing portions of a sequence showing similarity to taxon-b with '?', and vice versa. Second, portions of a sequence showing similarity to taxon-a in allele-1 were melded with portions of sequence in allele-2 also showing similarity to taxon-a (and so forth for taxon-b), with any gaps filled in with '?'. In this way, nearly complete "ancestral" sequences were reconstructed.

2.5 Sequence data partitions and incongruence

To assess similarity among genetic regions and distinguish between "soft" and "hard" incongruence (Johnson and Soltis, 1998; Seelanan et al., 1997), genomic regions were analyzed separately and across various combinations of datasets. Though some evidence suggests the incongruence length difference test (ILD; (Farris et al., 1995) can be too sensitive and incorrectly identifies incongruence when none exists (Yoder et al., 2001), this test was applied as a first approximation of overall data set homogeneity using PAUP*4.0b10 (Swofford, 2002) with 100 replicates each with 100 random addition sequence replicates and TBR.

If ILD deemed two data sets incongruent, the Wilcoxon signed-rank test, as adopted by Templeton and also implemented in PAUP*4.0b10 (SLP_T; Johnson and Soltis, 1998; Templeton, 1983), was used to localize incongruence to particular branches. Data partitions were analyzed using a “fast” bootstrap with 100,000 replications. Clades with bootstrap support greater than 70 were used to make constraint trees. Using SLP_T, the most parsimonious topologies from one data set were tested to see if they represent statistically supported suboptimal topologies for a second data set. A random number generator was used to select a single pairwise score when multiple

trees were recovered and Bonferroni corrections resulting in $\alpha = 0.05$ across all comparisons were made.

2.6 Combined matrix construction

The primary objectives for this study were met by assembling two primary matrices, hereafter referred to as the 'monophyly matrix' and the 'infrageneric matrix'. The monophyly matrix was constructed by concatenating five chloroplast regions (*trnK*, *trnL*, *trnS*, *trnD*, and *psbM*), ITS, and *PI* data for 12 species of *Collomia*, 15 species of *Navarretia*, and three species of *Allophyllum*. The two allopolyploid and single diploid *Collomia* species with strong evidence for reticulation within this genus were excluded to facilitate the formation of a single combined matrix without substantial incongruence between the chloroplast and nuclear data sets. We are confident based on our exploration of these data through analyses of incongruence, recombination, and so forth that our inferences of monophyly based on this sampling are sound.

The infrageneric analyses included all 15 *Collomia* species using multiple populations for most of these, with the three *Allophyllum* species retained for rooting purposes. Based on inferences drawn from analyses of the monophyly matrix, *Navarretia* species were excluded to focus on patterns within *Collomia* alone. The DNA portion of this matrix was expanded to include the *matK* and *yc6f* chloroplast regions and the nuclear *idhA* and *g3pdh* regions. Based on the results of the analyses of incongruence and recombination, a concatenated infrageneric matrix using principles for the construction of multilabeled trees (Brysting et al., 2007; Huber et al., 2006; Pririe et al., 2009) was constructed to accommodate all sampled taxa while incorporating the accumulated evidence for incongruence between data partitions and recombination within partitions.

2.7 Sequence analysis – parsimony methods

Parsimony analyses were performed on each individual DNA region assembled for both the monophyly and infrageneric analyses using PAUP*4b10. Settings included ACCTRAN optimization, amb- for zero length branches, equal weights applied to all characters, and TBR branch swapping with 100,000 random addition replicates. Branch support was evaluated via bootstrapping (1000 replications, each with 10,000 random addition sequence replicates).

Identical analyses were also conducted on the concatenated monophyly and infrageneric matrices. Additionally, partitioned Bremer support (decay) values were calculated using TreeRot v3.2 (Sorenson, 1999). These support values were used across the infrageneric matrix to determine the extent of support or disagreement the different data partitions contributed to the most parsimonious topology recovered by a multilabeled analysis of all DNA sequence data (Baker and DeSalle, 1997; Whitting, 2002).

2.8 Sequence analysis – likelihood and Bayesian methods

Using the multilabeled infrageneric matrix, likelihood analyses were performed using GARLI v 0.95 (Zwickl, 2006). A general model of sequence evolution was estimated in ModelTest v3.7 (Posada and Crandall, 1998) using the Akaike information criterion (AIC) . In addition to a likelihood analysis, a bootstrap analysis was conducted using 1000 bootstrap replications, each ending after 10,000 generations of no change from the optimal tree found during each replication.

Bayesian analyses were performed using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). Data with coded indels for each region was concatenated as a single partition. To avoid overparameterization (Sullivan and Joyce, 2005), partitions were established based on dataset

homogeneity as determined by ILD. Those datasets that were significantly different from each other were partitioned. ModelTest v3.7 using the AIC was again used to determine a model of sequence evolution for each gene region. Using partition-specific models of sequence evolution, data was analyzed using four replicate runs, each with 8 chains starting from a random tree. Each run was sampled every 1000 generations for 20,000,000 generations. Convergence was visually assessed using Tracer 1.5 (Rambaut and Drummond, 2009) and statistically assessed using the standard deviation of split frequencies using a 0.01 convergence threshold (Ronquist and Deans, 2010). Results were summarized with a majority rule consensus tree after discarding the first 5000 trees (25%) as burnin.

2.9 Morphological characters and character evolution

Twenty-six morphological characters were scored within *Collomia* for the infrageneric analyses (Appendices A, B). The morphological matrix for the infrageneric comparisons were assembled using MacClade 4.08 (Maddison and Maddison, 2005). Using the software Mesquite v.2.0 (Maddison and Maddison, 2009) the patterns of morphological character evolution were visualized in the most parsimonious tree derived from the multilabeled total evidence analysis. Both ACCTRAN and DELTRAN character optimization were applied on a character by character basis to ambiguities in the reconstruction of character evolution and maximize statements of homology across the entire topology (Agnarsson and Miller, 2008).

3. RESULTS

3.1 Sequence alignment metrics

The 'monophyly' matrix consisted of 3 genomic matrices (a combined cpDNA matrix, an ITS matrix, and a *PI* matrix) to test the hypothesis of monophyly of *Collomia* and *Navarretia*. The cpDNA alignment consisted of five regions. The *trnD* intergenic spacer had in 1141 characters and 31 indels (72 parsimony informative). The *psbM* intergenic spacer had 1277 characters and 44 indels (84 informative). The *trnL* intron/intergenic spacer had 1020 characters and 35 indels (69 informative). The *trnK* intron/intergenic spacer had 770 characters and 15 indels (47 informative). The *trnS-G* intergenic spacer had 869 characters and 62 indels (94 informative). While cpDNA regions were inherited as a single genomic unit, the two nuclear genes were not and were treated separately. Alignment and other phylogenetic metrics were calculated for the combined cpDNA and the individual nuclear genes (Table 1).

The 'infrageneric' matrix consisted of five genomic (combined cpDNA, ITS, *g3pdh, idhA, PI*) matrices and was used to investigate the intrageneric relationships within *Collomia*. The cpDNA alignment consisted of seven regions. The *trnD* intergenic spacer had in 1111 characters and 19 indels (63 parsimony informative). The *ycf6* intergenic spacer had 1186 characters and 12 indels (55 informative). The *psbM* intergenic spacer had 1132 characters and 16 indels (59 informative). The *matK* gene had 1095 characters (56 informative) and no indels. The *trnL* intron/intergenic spacer had 989 characters and 27 indels (65 informative). The *trnK* intron/intergenic spacer had 759 characters and 10 indels (37 informative). The *trnS-G* intergenic spacer had 857 characters and 40 indels (67 informative). Again, alignment and other phylogenetic metrics were calculated for the combined cpDNA and the individual nuclear genes used in this matrix (Table 1).

3.2 *Detection and treatment of recombination*

Only the *idhA* gene showed statistically significant evidence of recombination (Table 2). RDP detected recombination in both *C. biflora Sersic* types and in the *C. biflora 156* L-type. In all these, recombination occurred in the middle of parental strands with both a start and a stop point for the recombinant section. RDP did not detect recombination at the end of a sequence were there was start point but not an end point. We inferred recombination in *C. biflora 156* G type by manually reading the strands, calling the first location that switched from G- parental type to L- parental type as the beginning of the recombinant.

3.3 Analysis of incongruence

The ILD test demonstrated homogeneity between all cpDNA genes in both the monophyly and infrageneric analyses, but concatenated chloroplast datasets were not homogenous with the nuclear gene sequences in both analyses. In the monophyly analyses, the nuclear genes were homogenous with each other but asymmetrically homogenous with cpDNA (cpDNA homogenous with ITS but not *PI*). In the infrageneric matrix, the nuclear genes were asymmetrically homogenous with each other (Table 3). While ITS data was not homogenous with all other nuclear genes, *idhA* was homogenous with *PI* but not so with *g3pdh,* while *PI* was homogenous with *g3pdh*.

Data partitions in the monophyly matrix did not yield any topologies that were statistically incongruent. The topological congruency SLP_T identified 13 clades that were significantly incongruent from the infrageneric matrix (Table 4). In addition, the partitioned Bremer analysis conducted showed which nodes in the parsimony tree had conflicting signals between the data partitions (Tables 5, 6).

3.4 Infrageneric (mutlilabeled) matrix composition

Based on the results of the incongruence and recombination analyses, the final multilabeled matrix for recovering infrageneric relationships in *Collomia* was constructed. For *Allophyllum* and 11 *Collomia* species (*C. debilis, diversifolia, grandiflora, heterophylla, larsenii, linearis, macrocalyx, renacta, tenella*, *tinctoria,* and *tracyi*), chloroplast and nuclear sequences were simply concatenated in the manner typical of combined analyses. For *C. biflora* and *C. wilkenii*, the two allopolyploid species, two operational taxonomic units (OTUs) were included in the matrix for each sample. One corresponded to the maternal contribution to these allopolyploids and included chloroplast and nuclear sequences for all regions (except for nuclear ITS in *C. biflora* where no maternal copy was recovered in either direct sequencing or sequencing of clones and thus '?' was filled in at all nucleotide positions in ITS and corresponding coded indels). The second OTU included only the paternal contribution and thus contained sequence for all nuclear regions, but only '?' for the chloroplast regions (and corresponding indels). In the case of *C. biflora* and the nuclear *idhA* region, the reconstructed "ancestral" maternal and paternal sequences (see section 2. 5) were used in place of the actual sequenced regions. Though diploid, two OTUs were also included for each sample of *C. mazama*. One OTU was created for the chloroplast sequences and indels (with '?' used at all nucleotide and indel positions in the nuclear regions) and one OTU for the nuclear sequences and indels (with '?' used at all nucleotide and indel positions in the chloroplast regions). This enabled the phylogenetic signal in the chloroplast and nuclear data sets to be incorporated in the combined analysis without creating homoplasy by forcing the strongly conflicting signal between chloroplast and nuclear data sets for this taxon into a single OTU. For *C. rawsoniana*, the *g3pdh* nucleotides were replaced with '?' given that the placement of this taxon in the *g3pdh*

gene tree was significantly incongruent and anomalous with its placement in all other gene trees and its morphology. A second OTU for this sample to incorporate the *g3pdh C. rawsoniana* sequence was therefore not included.

3.4 Phylogenetic analyses

3.4.1 Monophyly assessment

Parsimony analyses of a concatenated cpDNA partition (five regions combined), ITS partition, and *PI* partition recovered a monophyletic *Collomia* and *Navarretia* with cpDNA and *PI* (Fig. 1a,b), but showed *Navarretia* alone monophyletic within a paraphyletic *Collomia* with ITS (Fig. 1c). Combined analysis of all data resolved *Collomia* and *Navarretia* as reciprocally monophyletic. Six most parsimonious trees were recovered from the combined analysis (Fig. 2). Nodal support based on partitioned Bremer support and bootstrap values were high in all analyses and support the monophyly of *Collomia* and *Navarretia* (Table 5).

ModelTest selected the GTR + I + Γ model of nucleotide substitution for the combined monophyly matrix used in the maximum likelihood analyses. The most likely topology based on these data had a -lnL value of 22437.78226. For the Bayesian analysis, three data partitions were used. A TVM + I + Γ model of nucleotide substitution was estimated for the concatenated cpDNA partition, an SYM + I + Γ model for the ITS partition, and a GTR + I + Γ model for the *PI* partition. After 20 million generations, all four runs converged on the same parameter space (standard deviation of splits frequencies was 0.0014151) with a mean -lnL value of 22829.3881. Given that the topologies recovered by parsimony, likelihood, and Bayesian methods were similar, likelihood bootstrap and Bayesian posterior probability values are presented on the parsimony topology (Table 5).

3.4.2 Infrageneric relationships

Separate parsimony analyses of the gene partitions individually (with the cpDNA regions concatenated into a single partition) provided varied inferences regarding relationships within *Collomia* (Fig. 3). In general, *C. diversifolia*, *C. heterophylla*, *C. tracyi*, and *C. tinctoria* diverged earlier than other species, but their relationship to each other and other *Collomia* varied between data partitions. A clade composed of the perennial species (i.e., section *Collomiastrum*; *C. debilis*, *C. larsenii*, *C. mazama*, and *C. rawsoniana*) was recovered by all nuclear regions except *g3pdh*, where *C. rawsoniana* was positioned as part of a polytomy that included several other multispecies clades. Chloroplast sequences, however, strongly placed the perennial species *C. mazama* within a clade of lanceolate-leaved annuals, at odds with all nuclear data partitions and morphology. Chloroplast sequences identified *C. grandiflora* and *C. tenella* (or their recent ancestors) as the maternal parent of the allopolyploids *C. biflora* and *C. wilkenii*, respectively. The nuclear sequences for these two species clustered with these maternal species or *C. linearis*, the paternal parent of both polyploids. Among the diploid species when multiple populations were sampled, most, but not all, populations coalesced within a species. Exceptions included one population of *C. linearis* that showed greater affinity to *C. macrocalyx* in the cpDNA tree, one population of *C. debilis* in the *PI* tree, and *C. tinctoria* in all but the ITS trees.

Analyses of a single concatenated multilabeled matrix (sections 2.5, 3.4) produced 12 parsimonious topologies, the strict consensus topology being reported (Fig. 4). Clades corresponding to sections *Courtoisia* and *Collomiastrum* were recovered (allowing for chloroplast capture in C. mazama), but section *Collomia* was shown to be paraphyletic with *C. tinctoria* and *C. tracyi* forming a lineage well separated from the core member of section

Collomia, and C. *grandiflora* as sister to the core of section *Collomia* but without statistical support (Fig. 4; Table 6).

ModelTest selected the GTR + I + Γ model of nucleotide substitution for the combined data set used in the maximum likelihood run in GARLI. The most likely topology based on the data had a -lnL value of 30759.50387. For the Bayesian analysis, five data partitions were used. A TVM + I + Γ model of nucleotide substitution was estimated for the concatenated cpDNA dataset, an SYM + Γ model for the ITS dataset, a TVM + Γ model for *g3pdh*, and a TVM + Γ model for *PI*. After 20 million generations, all four runs converged on the same parameter space (standard deviation of splits frequencies was 0.006698) with a mean -lnL value of 31336.0917. As with the monophyly analyses, the topology of the likelihood and Bayesian trees were similar to the parsimony trees and only their nodal support values were presented here (Fig. 4; Table 6). The only difference between the Bayesian tree from the likelihood and parsimony topologies was the placement of *C. grandiflora* (and allied *C. biflora* sequences) as sister to the perennial species of section *Collomiastrum*, but with a posterior probability value of only 50.

3.5 Character evolution within Collomia

Perennialism evolved once and is a synapomorphic character of the *Collomiastrum* clade. Seed morphology incorporates two characters – seed type and mucilage amount (Fig. 5a). Based on the topologies discussed, type 1 seeds, identified by having hexagonal epidermal cells with well-defined boundaries between the cells, and the absence of seed mucilage appeared as the pleisiomorphic seed state of *Collomiastrum*. Seed type 2 with inconspicuous cell boundaries evolved a second time, uniting *C. mazama* and *C. rawsoniana* within *Collomiastrum.*

Pollen morphology incorporates two characters – aperture distribution and sexsine sculpturing (Fig. 5b). Zonotreme apertures evolved twice in the topology, once as an autapomorphy within *C. diversifolia* and once synapomorphically for the clade consisting of all other species except for *C. heterophylla, tinctoria,* and *tracyi*. Irregular reticulate ridges evolved as an autapomorphy within *C. heterophylla*. Radiate ridges evolved once as a synapomorphy for the *C. tinctoria/tracyi* clade. Striated and striato-reticulate ridges fall out in multiple, well-supported clades across the topology.

4. DISCUSSION

4.1 Incongruency and reticulation

The ILD test, though prone to type I errors (Yoder et al., 2001), served as a basis to determine data combinability. For the monophyly matrix, the cpDNA and the nuclear DNA were not homogenous, as expected, because on differing modes of inheritance, gene histories, or high degrees of homoplasy (Dolphin et al., 2000). Based on the topological congruency test, however, the notion that one gene topology is significantly different than another failed to be rejected, suggesting the gene topologies all were suboptimal topologies of each other. All genomic datasets were combined in a total evidence approach to assess evolutionary signal and phylogeny. Because the homogeneity based on the ILD test could not be attributed to any biological agent, a simultaneous analysis of all data was the best approach to maximize the phylogenetic signal present in multiple, independent data, thus producing a well-supported, biologically plausible topology (Nixon and Carpenter, 1996).

The ILD test found many of the genomic data partitions used in infrageneric matrix as not being homogenous. The topological congruency test found clades that were statistically

incongruent. These clades were composed of the same six species, namely *C. linearis, C. grandiflora, C. tenella, C. biflora, C. wilkenii,* and *C. mazama*. Cross-referencing these species back to the partitioned Bremer values showed conflicting support of the partitioned genes based on the most parsimonious topology. The incongruencies and conflicting support detected were due to reticulation. *C. wilkenii* is an allopolyploid species whose parental lineages can be traced to *C. linearis* and *C. tenella* (Johnson and Johnson, 2006). *C. biflora* is also an allopolyploid with ties to *C. linearis* and *C. grandiflora*. Hard incongruence between the cpDNA and nuclear DNA/morphology suggest that *C. mazama* is the product of an introgression event with a species ancestral to several extant annual species (Johnson et al., 2008; Johnson and Johnson, 2006). Removal of reticulate species resulted in topologies that were suboptimal trees for all partitions (Nixon and Carpenter, 1996).

4.2 Handling reticulation

Simultaneous analyses of data with reticulate species can result in misleading phylogenies (de Queiroz et al., 1995). Previous studies involving reticulate species advocate the removal of said species from the analysis (Hughes et al., 2002). The removal of reticulate species from analyses, however, does not coincide with the phylogenetic purpose. Though simultaneous analyses maximize phylogenetic signal from independent datasets, care should be taken when compiling reticulate data into a matrix. Failure to account for reticulation in a dataset can lead to spurious results.

Plant reticulation results from multiple biological phenomena including hybridization and recombination. Allopolyploidy instantly creates a new species with the complete nuclear genome of two parental species and the organellar genome of the maternal species. The lack of lineage

independency needs to be addressed before a simultaneous analysis can be performed and a phylogeny inferred. Knowing *C. biflora* and *C. wilkenii* were allopolyploid species, the parental sequence types were identified and named them accordingly (i.e. *C. biflora*…L type or G type). Multiple labels were used for the same sample indicating the parental type and that type's corresponding genetic data (Huber and Moulton, 2006; Huber et al., 2006). Since allopolyploids have the chloroplast of the maternal type, the known paternal type had its chloroplast data replaced with (?) as it was not sequenced in the taxa. A multi-labeled approach accurately portrayed the relationship the allopolyploid has with each parental lineage.

Introgression, through repeated backcrossing of a hybrid with a parental species, blurs phylogenetic boundaries. The perennial species *C. mazama* has nuclear markers similar to other perennial species. However, chloroplast DNA alone infers a relationship between some annual species sister to the existing *Collomia* sectional clade. It is hypothesized that *C. mazama,* early in its evolutionary history, captured a chloroplast through introgression with a species ancestral to the annual species in the genus (Johnson et al., 2008), a claim statistically supported by the SLP_T and the partitioned Bremer support values. Similar to the allopolyploids, a multilabeled approach was used, labeling each *C. mazama* taxon twice, once for the chloroplast genome and once for the nuclear contribution. Such an approach should showed the relationship between the ancestral annual and perennial species as related to modern taxa.

Meiotic recombination is also a source of reticulation. While recombination occurs randomly, it is seldom detected between homologous chromosomes. In allopolyploid species, where each parental type contributes half of its nuclear genome, each parental haplotype block has a unique evolutionary history. In homeolog recombination, exchange between these unique haplotypes occur, resulting in chromatids with two evolutionary histories (Linder et al., 2003).

All nuclear genes were tested for recombination. Recombination was detected in *C. biflora* for the *idhA* gene. RDP3 estimated the recombination start and stop points which were used to remove the recombinant sections. Where recombination was detected between homeologs, the recombinant section was removed and included in with the other parental type, thus removing the reticulation between the two parental species types and creating a nearly complete "ancestral" sequence (Fig. 6). Accounting for the reticulate history of the species allowed for the retrieval of biologically meaningful relationships between taxa.

Population genetic studies use network approaches to show genetic relationships between different haplotypes. Should we not use similar approaches in phylogenetic studies to accurately portray the evolutionary history known about the species in question? Multifurcation programs that use multilabeled trees are beginning to be developed (Brysting et al., 2007; Huber and Moulton, 2006; Huber et al., 2006). PADRE (Lott et al., 2009) allows different topologies to be combined in a majority rule fashion to show a phylogenetic network between multilabeled gene trees. However, forming a majority-rule consensus tree from gene trees causes trees to lose resolution. Rather than form a consensus tree, we took the multilabeled total evidence tree and manually created a networked phylogeny for *Collomia* (Fig. 7). This topology represents the total amount of evidence currently known regarding the species present in this genus. Each of the allopolyploids (*C. biflora* and *C. wilkenii*) is grouped with their respective chloroplast type. Both of these species also have nuclear genes that also can be traced to an ancestral paternal source, namely *C. linearis*. By having one species linked to two other species, historical allopolyploidy can be accurately displayed on a single gene topology. Furthermore, based on the hypothesis of introgression and *C. mazama*, this historical event can also be accurately displayed. While an exact annual species within the section *Collomia* cannot be pinpointed, the line shows a

historical event that links modern-day *mazama* with some unknown annual. Because phylogenies and species trees aim to accurately represent and describe the relationships that exist between the species, multifurcating trees or even multilabeled trees should be used to account for the historical genomic processes of reticulation.

4.3 Monophyly of Collomia

Emphasizing calyx morphology, Greene (1887) laid the foundation for the modern concept of *Collomia*. Though initially not accepted universally, Greene's concept of *Collomia* has been followed at least since the work of Wherry (1944) and accepted unquestionably as a natural group for studies of character diversity and evolution (Grant, 1959; Wilken et al., 1982). Early DNA sequencing work (Johnson and Soltis, 1995; Porter, 1996; Steele and Vilgalys, 1994) suggested a close relationship between *Collomia* and *Navarretia*, a surprising result given the classification of Polemoniaceae commonly accepted at the time placed these genera in separate tribes (Grant, 1959). Subsequent molecular work with broader taxon sampling has confirmed that *Allophyllum* is sister to *Collomia* + *Navarretia* (Johnson et al., 1996), but not always with *Collomia* and *Navarretia* as reciprocally monophyletic (Johnson et al., 2008).

Here, a combined analysis of several chloroplast genes, nuclear ITS sequences, and partial nuclear PI sequences with nearly complete sampling of *Collomia* and a well sampled *Navarretia* recovered both genera as monophyletic. Bootstrap support from parsimony (86% and 94% for *Collomia* and *Navarretia*, respectively), likelihood (81% and 99%), and posterior probabilities from Bayesian analyses (100% for both) provided moderate to strong support for monophyly of both genera.

In these analyses, *C. mazama*, *C. biflora*, and *C. wilkenii* were excluded from *Collomia* given strong evidence in genetic data for reticulation in these species. Because all three species occupy relatively derived positions within *Collomia* (Fig. 3), their absence did not impact the inference for monophyly as confirmed through various permutations of the data (not shown). Similarly, the species sampled to represent *Navarretia* spanned the phylogenetic and morphological diversity of this genus, particularly with the inclusion of *N. sinistra*, a species which, until recently was included in *Gilia,* that occupied an early branching position with respect to the remainder of *Navarretia* in ongoing investigations in the Johnson lab.

Explicit tests of monophyly showed that the shortest trees that did not recover a monophyletic *Collomia* were not significantly worse than the shortest overall trees (3 additional steps, $p = 0.4913$). This, together with branch lengths, suggested a relatively narrow window of time during which *Collomia* and *Navarretia* both diverged from each other and began diversifying within themselves. Calyx morphology (plicate sinuses in Collomia vs. not plicate in *Navarretia*), base chromosome number ($x = 8$ in *Collomia* vs. $x = 9$ in *Navarretia*), and capsule dehiscence (explosive in *Collomia* vs. rupturing or none in *Navarretia*) agreed with the molecular data and provide morphological synapomorphies for *Collomia*.

4.4 Intrageneric relationships within Collomia

Wherry (1944), based solely on morphological comparisons, partitioned the genus into three sections that continue to be recognized by taxonomists (Chuang et al., 1978; Hsiao and Chuang, 1981). These three sections are *Collomiastrum*, *Collomia*, and *Courtoisia*. Section *Collomiastrum* comprises the four perennial species *C. debilis*, *C. larsenii*, *C. mazama*, and *C. rawsoniana*. These four species have relatively broad leaves and a rhizomatous habit. Section

Courtoisia contains the spreading, lobed-leaved annuals *C. heterophylla* and *C. diversifolia*, the only two species with greater than one seed per locule. Section *Collomia* circumscribes the remaining nine annual species, which are largely linear-leaved although occasional leaf lobes can be found. While Wherry's work has been widely accepted and supported morphologically, genetic markers conflicted with his proposal. When the species belonging to each section were located on the topology, Wherry's three-section genus classification was inadequate. While Wherry's original three sections still existed, only sections *Collomiastrum* and *Courtoisia* existed as monophyletic groups containing the species as proposed. Section *Collomia* was not monophyletic and since taxonomic names should imply information about the existence of monophyly, section *Collomia* needed to be divided into a modified section *Collomia* with a new section being suggested to encompass *C. tinctoria* and *C. tracyi* (name forthcoming) while *C. grandiflora* and *C. biflora* were removed from any sectional affiliation.

The perennial section *Collomiastrum*, consisting of *C. debilis*, *C. larsenii*, the nuclear portion of *C. mazama*, and *C. rawsoniana*, continued to be supported as a monophyletic clade. The intrasection Bremer support values totaled 30.0 with support from the chloroplast genome and nuclear genome. Overall, the MP and ML bootstrap values (100, 99) and Bayesian posterior probability (1.00) for the perennial node showed support. There are also morphological characters that unite the section including perennialism, striate pollen ridges, and a reduction of seed mucilage.

Section *Courtoisia* containing the spreading, lobed-leaved annuals *C. heterophylla* and *C. diversifolia* continued to be a strongly supported, monophyletic section. Bremer support value for the intrasection nodes was 158 with the chloroplast genome, nuclear genome, and morphological characters providing phylogenetic signal in support of the section.

Morphologically, multiple seeds per locule characterized the section. The MP bootstrap value (100), ML bootstrap value (100), and posterior probability (1.00) for the sectional node also supported the continuance of the section as postulated by Wherry.

Collomia diversifolia continues to be a source of topological incongruency. The chloroplast genome strongly supported the placement of *C. diversifolia* and *C. heterophylla* in a monophyletic group. Every nuclear gene had these two species topologically separated in some manner (Table 4, Fig. 3). However, only one gene used, *g3pd*, significantly was not a suboptimal topology of the chloroplast-based placement of the two species. The combination of all nuclear genes as well as the combination of ITS, *PI*, and *idhA* still did not produce statistical support for topological incongruency between the chloroplast topology. It remains noted that one nuclear gene is in statistical disagreement with the chloroplast topology and the remaining nuclear genes were only suboptimal agreements. Also, pollen morphology of *C. diversifolia* is the same as other annuals in section *Collomia.* It is possible that *C. diversifolia* is an ancestor of some ancient hybridization event with an unknown annual (see section 4.5).

Wherry's treatment of the section *Collomia* was paraphyletic. Since the goal of systematics is to form monophyletic groups, we propose a reevaluation of section *Collomia*, with the remaining nine species placed in the monophyletic groups shown and supported by the parsimony tree. First, section *Collomia* will continue to exist by name but should be revised to include *C. linearis, C. tenella, C. macrocalyx, C. renacta, C. wilkenii* and only half of the nuclear material of *C. biflora* and only the chloroplast genome of *C. mazama*. This revised section had a summed intrasection Bremer support value of 49.24 with support from genomic and morphological characteristics. The MP bootstrap value (100), ML bootstrap value (99), and posterior probability (1.00) for the sectional node strongly supported this reevaluated section.

The species in this section are linear-leafed annuals and contains the type species for the genus, *C. linearis*. As noted by Wilken (1977), *C. linearis* has a very plastic morphology. As such, there are no definitive morphological characters that are unique across all members of the section. There are distinct features that can be used to identify species within the section. Unknown is the placement of the chloroplast of *C. mazama*. Though nestled within the section, parsimony trees placed the species as sister to *C. tenella* or as a polytomy within the clade. It remains unclear as to who the maternal donor of the chloroplast is.

While parsimony and likelihood trees placed *C. grandiflora* within the above prescribed section, nodal support values obtained from bootstrapping are low. Bayesian analysis placed *C. grandiflora* basal to the perennial species in section *Collomiastrum*. This is not a preposterous notion given *C. grandiflora*'s pollen morphology matches that of the perennials and since we propose that perennialism was a derived trait it is possible that *C. grandiflora* is the most closely related. Because of the lack of support, we advocate the exclusion of *C. grandiflora* from section *Collomia*. Topologically, this section has two species found in a monophyletic clade, *C. biflora* and *C. grandiflora*. Though there were two species in the clade, *C. biflora* is an allopolyploid descendant of *C. grandiflora* and *C. linearis*. Analysis of the nuclear genes alone places each of the parental chromosomal types in their respective sections, *C. biflora* L-type falls within section *Collomia* with *C. linearis* while *C. biflora* G-type falls alongside *C. grandiflora* while cpDNA alone indicates *C. grandiflora* to be the maternal ancestor. Because *C. biflora* shares genomic and morphological characteristics with both of its parents, the one species that firmly establishes the recognition of this section is *C. grandiflora*. Being a monotypic section, there is no reason to formally erect a section, synonimizing it with the species itself. Like the previously discussed sections, this clade had a summed intrasection Bremer support value of 52 with support from the

genomic and morphological data, with bootstrap (100), likelihood (100), and posterior probability (1.00) providing strong nodal support.

We propose that a new section needs to be recognized to best explain the relationships within the genus. Historically placed in section *Collomia* (Chuang et al., 1978; Grant, 1959; Hsiao and Chuang, 1981; Wherry, 1944), the linear-leafed annuals *C. tracyi* and *C. tinctoria* form a monophyletic group basal to others in section *Collomia*. Though Hsiao and Chuang recognized differences between these species and others in section *Collomia*, no formal reorganization was proposed, lacking further evidence. With genetic support confirming the morphological differences observed, we propose the formation of a new section for *C. tinctoria* and *C. tracyi* (name forthcoming; Green, submitted). This new section had a summed intrasection Bremer support value of 88 and was supported by all data currently available. This section has two synapomorphic morphological characters states that defined it. First, both species in this section have radiated ridges on the surface of their pollen grains. Secondly, the aristate apex of the calyx lobes can identify both species in this section. In addition, the bootstrap value (100), ML bootstrap value (100), and posterior probability (1.00) provided sectional node support to the genomic and morphological data.

Because the parameters of taxon sampling, taxon weighting, base frequencies, substitution rates, etc. were tested using bootstrap, likelihood, and Bayesian methods, all of which converged on the same parsimonious topology, the topology is stable and we can be confident in the intrageneric relationships. Though a stable topology is a good estimate of the species tree, current phylogenetic methods require species tree to be bifurcating. However, when conditions of recombination, introgression, and allopolyploidy arise, the use of a bifurcating tree does not relate the species true phylogeny. The use of these bifurcating methods will only show

the most supported line of data, be it from the chloroplast or the nuclear genome. However, when a species is known to have genetic information from two different sources, where on a species tree would it lie? Would it be fair to choose one placement over another given its historical background? We believe the network approach is an adequate method to demonstrate the historical processes resulting in reticulation, and the relationships between these species with others in the genus.

4.5 Character evolution

Wherry (1944) and Grant (1959) postulated a likely pattern of diversification in *Collomia* based on the morphological data in hand at the time. Both authors postulated that the perennial *Collomia* are ancestral to annual *Collomia*. Regardless of how this character is optimized on the trees reconstructed here (Fig. 3), the perennial lifespan is derived within *Collomia* rather than a pliesomorphy for this genus. *Allophyllum*, *Navarretia*, and *Collomia* are themselves derived within tribe Gilieae (Johnson et al., 2008), and the entire tribe with the exception of *Collomia* section *Collomiastrum* is composed of annual species. Thus, while the first Polemoniaceae were likely perennial plants, the expression of this trait takes many forms (i.e., woody shrubs, monocarpic rossetes, rhizomatous herbs) and is undoubtedly a homoplasious character at the family level (Barrett et al., 1997). Within *Collomia*, the perennial habit evolved once, but there is variation in the expression of this character among the four species. *Collomia rawsoniana* and *C. mazama* grow in relatively mesic sites, conducive to a rhizomatous habit. Above ground stems are often associated in patches, without a discernable caudex. In contrast, *C. larsenii* and *C. debilis* frequent relatively barren slopes of talus, rubble, or volcanic scree. Above ground stems can be traced through the ruble to a subterranean caudex.

Changes in seed morphology in *Collomia* co-occur with changes to the perennial habit (Fig. 5). Whereas many Polemoniaceae and all species of tribe Gilieae produce copious mucilage often associate with spiracles upon wetting, mucilage and spiracle production is scanty in C. rawsoniana and C. mazama and absent in C. debilis and C. larsenii. The development of seed mucilage and spiracles has been studied in *Collomia* (Hsiao and Chuang, 1981), but their function is largely conjectural. Certainly the stickiness produced enable the seeds to adhere to surfaces, and have likely played a role in the dispersal of species, including the long-distance dispersal of several Polemoniaceae to southern South America (Carlquist et al., 1984), but it is unlikely that dispersal is the primary selective agent for mucilage. A role in seed germination may be important, and a shift to a perennial habit would place less emphasis on seedling recruitment for maintaining populations. The adhesion properties of mucilage may even be detrimental in *C. debilis* and *C. larsenii*, where a seed stuck to the surface of a rock will likely never be able to extend its radical deep enough into the talus substrate to enable establishment. Nonsticky seeds, on the other hand, would be carried below a talus or scree surface by natural processes to a point more conducive for seedling establishment. A sister relationship between *C. debilis* and *C. larsenii*, as discussed previously, would simplify the view of character evolution among the perennials species (Fig 5).

In *Gilieae*, zonotreme pollen grains are apparently ancestral, and ubiquitous in *Saltugilia* and *Gilia* with *Lathrocasis* having anonotreme apertures (Johnson et al., 1996). *Allophyllum* and *Navarretia* are exclusively pantotreme, and the earliest branching *Collomia* also possess pantotreme apertures (Fig 5). Zonotreme apertures in *Collomia* are thus a reversion to a more wide spread state elsewhere in *Gilieae*. Within *Collomia*, the zonotreme apertures appear to have evolved twice, once in the ancestor shared by all species excusive of *C. tinctoria*, *C. tracyi*, *C.*

heterophylla, and *C. diversifolia*, and again in *C. diversifolia*. Alternatively, if *C. diversifolia* is the product of diploid homoploid hybridization or introgressed with an ancestral *Collomia* with zonotreme pollen as the nuclear data ambiguously suggest, a single origin of zonotreme grains in Collomia could be postulated with the disjunct distribution of this state providing a relictual clue to this species origins.

Pollen ridges also have a complex pattern of evolution. *Collomia* pollen grains have one of four ridge types depending on the species. Multiple equally optimal routes exist to explain pollen ridge evolution (Fig 5), but only one is explained here. A state change from pertectate to irregular reticulate ridges occurred at the genus node giving *C. heterophylla* irregular reticulate ridges. An autapomorphic state change occurred during the diversification of the *C. tinctoria/tracyi* clade giving these species radiate ridges. A third state change occurred as a synapomorphy for the *C. grandiflora* and *Collomiastrum* clades giving them striated ridges. Finally, striato-reticulate ridges evolved independently twice during the history of the genus, once as an apomorphy for *C. diversifolia* and once as a synapomorphy for section *Collomia*.

While the data suggests the above scenarios, the data also suggests a possible ancient hybridization event with *C. diversifolia* and some other annual species belonging to or ancestral to section *Collomia*. Nodal supports (parsimony bootstrap, ML bootstrap, Bayesian posterior distribution) highly support the node linking *C. diversifolia* and *C. heterophylla* in a clade. Morphologically they are the only *Collomia* species to have more than one seed per locule. However, the pollen aperture and ridge type of *C. diversifolia* is different from *C. heterophylla*. Those of *C. diversifolia* match those of section *Collomia*. Furthermore, the partitioned Bremer support values show a disagreement between the chloroplast and the nuclear data at the node and within the section uniting *C. diversifolia* and *heterophylla*. The chloroplast genome supports the

inclusion of *C. diversifolia* within *Courtoisia* whereas the nuclear genome conflicts with this placement. There is both morphological and genetic support suggesting a possible ancestral hybridization event with *C. diversifolia* and another annual species belonging to or ancestral to section *Collomia*. If such is the case, each pollen ridge type evolved once during the generic history.

REFERENCES

Agnarsson, I., Miller, J.A., 2008. Is ACCTRAN better than DELTRAN? Cladistics 24, 1-7.

Baker, R.H., DeSalle, R., 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46, 654-673.

Barrett, S.C.H., Harder, L.D., Worley, A.C., 1997. The comparitive biology of pollination and mating in flowering plants. In: Silvertown, J., Franco, M., Harper, J.L. (Eds.), Plant Life Histories. The Royal Society, Cambridge, pp. 57-76.

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.

Carlquist, S., Eckhart, V.M., Michener, D.C., 1984. Wood anatomy of Polemoniaceae. Aliso 10, 547-572.

Chuang, T.I., Hsieh, W.C., Wilken, D.H., 1978. Contribution of pollen morphology to systematics of Collomia (Polemoniaceae). Am. J. Bot. 65, 450-458.

Cullings, K.W., 1992. Design and testing of a plant-specific PCR primer for ecological adn evolutionary studies. Mol. Ecol. 1, 233-240.

de Queiroz, A., Donoghue, M.J., Kim, J., 1995. Separate versus combined analysis of phylogenetic evidence. Annu. Rev. Ecol. Syst. 26, 657-681.

DeMesure, B., Sodzi, N., Petit, R.J., 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol. Ecol. 4, 129-131.

Dolphin, K., Belshaw, R., Orme, C.D.L., Quicke, D.L.J., 2000. Noise and incongruence: interpreting results of the incongruence length difference test. Mol. Phylogen. Evol. 17, 401-406.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissues. Phytochem. Bull. 19, 11-15.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792-1797.

Edgar, R.C., Botzoglou, S., 2006. Multiple sequence alignment. Curr. Opin. Struct. Biol. 16, 1-6.

Farris, J.S., Kallersjo, M., Kluge, A.G., Bult, C., 1995. Testing significance of incongruence. Cladistics 10.

Grant, V., 1959. Natural history of the phlox family. Martinus Nijhoff, The Hague, Netherlands.

Greene, E.L., 1887. American Polemoniaceae. Pittonia: a series of papers relating to botany and botanists. Cubery & Co., San Fransisco (CA), pp. 120-139.

Hamilton, M.B., 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. Mol. Ecol. 8, 521-523.

Hsiao, Y., Chuang, T.I., 1981. Seed-coat morphology and anatomy in Collomia (Polemoniaceae). Am. J. Bot. 68, 1155-1164.

Huber, K.T., Moulton, V., 2006. Phylogenetic networks from multi-labelled trees. J. Math. Biol. 52, 613-632.

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.

Hughes, C.E., Bailey, C.D., Harris, S.A., 2002. Divergent and reticulate species relationships in *Leucaena* (Fabaceae) inferred from multiple data sources: insights into polyploid origins and nrDNA polymorphism. Am. J. Bot. 89, 1057-1073.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254-267.

Johnson, L.A., Chan, L.M., Weese, T.L., Busby, L.D., McMurry, S., 2008. Nuclear and cpDNA sequences combined provide strong inference of higher phylogenetic relationships in the phlox family (Polemoniaceae). Mol. Phylogenet. Evol. 48, 997-1012.

Johnson, L.A., Johnson, R.L., 2006. Morphological delimitation and molecular evidence for allopolyploidy in *Collomia wilkenii* (Polemoniaceae), a new species from northern Nevada. Syst. Bot. 31, 349-360.

Johnson, L.A., Schultz, J.L., Soltis, D.E., Soltis, P.S., 1996. Monophyly and generic relationships of Polemoniaceae based on matK sequences. Am. J. Bot. 83, 1207-1224.

Johnson, L.A., Soltis, D.E., 1995. Phylogenetic inference in Saxifragaceae sensu stricto and Gilia (Polemoniaceae) using matK sequences. Annals of the Missouri Botanical Garden 82, 149-175.

Johnson, L.A., Soltis, D.E., 1998. Assessing congruence: empirical examples from molecular data. In: Molecular systematics of plants II: DNA sequencing. Kluwer Academic Publishers, Boston. MA.

Kelchner, S.A., 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. Ann. Mo. Bot. Gard. 87, 482-498.

Kelly, L.J., Leitch, A.R., Clarkson, J.J., Hunter, R.B., Knapp, S., Chase, M.W., 2010. Intragenic recombination events and evidence for hybrid speciation in *Nicotiana* (Solonaceae). Mol. Biol. Evol. 27, 781-799.

Kluge, A.G., 1989. A concern for evidence and a phylogenetic hypothesis for relationships among Epicrates (Boidae, Serpentes). Syst. Biol. 38, 7-25.

Linder, C.R., Moret, B.M.E., Nakhleh, L., Warnow, T., 2003. Network (reticulate) evolution: biology, models, and algorithms. Proceedings of The Ninth Pacific Symposium on Biocomputing (PSB).

Linder, C.R., Rieseberg, L.H., 2004. Reconstructing patterns of reticulate evolution in plants. Am. J. Bot. 91, 1700-1708.

Lott, M., Spillner, A., Huber, K.T., Moulton, V., 2009. PADRE: a package for analyzing and displaying reticulate evolution. Bioinformatics 25, 1199-1200.

Maddison, D.R., Maddison, W.P., 2005. MacClade 4: analysis of phylogeny and character evolution, version 4.08. Sinauer Associations, Sunderland. MA.

Maddison, W.P., Maddison, D.R., 2009. Mesquite: a modular system for evolutionary analysis. Version 2.72 http://mesquiteproject.org.

Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260-262.

Muller, K., 2005. SeqState: primer design and sequence statistics for phylogenetic DNA datasets. Appl. Bioinf. 4, 65-69.

Nixon, K.C., Carpenter, J.M., 1996. On simultaneous analysis. Cladistics 12, 221-241.

Okuyama, Y., Fujii, N., Wakabayashi, M., Kawakita, A., Ito, M., Watanabe, M., Murakami, N., Kato, M., 2005. Nonuniform concerted evolution and chloroplast caputure: heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian Mitella (Saxifragaceae). Mol. Biol. Evol. 22, 285-296.

Petersen, G., Seberg, O., 2004. On the origin of the tetraploid species *Hordeum capense* and *H. secalinum* (Poaceae). Syst. Bot. 29, 862-873.

Porter, J.M., 1996. Phylogeny of Polemoniaceae based on nuclear ribosomal internal transcribed spacer DNA sequences. Aliso 15, 57-77.

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

Pririe, M.D., Humphreys, A.M., Barker, N.P., Linder, H.P., 2009. Reticulation, data combination, and inferring evolutionary history: an example from Dathonioideae (Poaceae). Syst. Biol. 58, 612-628.

Rambaut, A., 2002. Se-Al sequence alignment editor version 2.0. University of Oxford, Oxford. United Kingdom.

Rambaut, A., Drummond, A.J., 2009. Tracer v1.5: MCMC trace analysis tool.

Rieseberg, L.H., Brunsfeld, S.J., 1992. Molecular evidence and plant introgression. In: Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), Plant molecular systematics. Chapman and Hall, New York (NY), pp. 151-176.

Ronquist, F., Deans, A.R., 2010. Bayesian phylogenetics and its influence on insect systematics. Annu. Rev. Entomol. 55, 189-206.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Sang, T., Crawford, D.J., Stuessy, T.F., 1995. Documentation of reticulate evolution in peonies (Paeonia) using internal transcribed spacer sequences of nuclear ribosomal DNA -- implications for biogeography and concerted evolution. Proceedings of the National Academy of Sciences 92, 6813-6817.

Seelanan, T., Schnabel, H., Wendel, J.F., 1997. Congruence and consensus in the cotton tribe. Syst. Bot. 22, 259-290.

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Lie, W., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E., Small, R.L., 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. 92, 142-166.

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

Soltis, D.E., Kuzoff, R.K., 1995. Discordance between nuclear and chloroplast phylogenies in the Heuchera group (Saxifragaceae). Evolution 49, 727-742.

Sorenson, M.D., 1999. TreeRot version 2. Boston University, Boston. MA.

Spencer, S.C., Porter, J.M., 1997. Evolutionary diversification and adaptation to novel environments in Navarretia (Polemoniaceae). Syst. Bot. 22, 649-668.

Steele, K.P., Vilgalys, R., 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene matK. Syst. Bot. 19, 126-142.

Strand, A.E., Leebens-Mack, J., Milligan, B.G., 1997. Nuclear DNA-based markers for plant evolutionary biology. Mol. Ecol. 6, 113-118.

Sullivan, J., Joyce, P., 2005. Model selection in phylogenetics. Annu. Rev. Ecol. Syst. 36, 445-466.

Swofford, D.L., 2002. PAUP* phylogenetic analysis using parsimony (* and other methods), version 4.0b10. Sinauer Associations, Sunderland. MA.

Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17, 1105-1110.

Templeton, A.R., 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. Evolution 37, 221-244.

Vriesendorp, B., Bakker, F.T., 2005. Reconstructing patterns of reticulate evolution in angiosperms: what can we do? Taxon 54, 593-604.

Weese, T.L., 2004. Phylogenetic analyses in the phlox family (Polemoniaceae). Biology. Brigham Young University, Provo, p. 149.

Wendel, J.F., Doyle, J.J., 2005. Polyploidy and evolution in plants. In: Henry, R.J. (Ed.), Plant diversity and evolution: genotypic and phenotypic variation in higher plants. CABI, Oxfordshire (UK), pp. 97-117.

Wherry, E.T., 1944. Review of the genera Collomia and Gymnosteris. Am. Midl. Nat. 31, 216-231.

White, T.J., Bruns, T., Lee, S., Taylor, J.W., 1990. Amplification and direct sequencing of fungal ribosomal genes for phylogenetics. In: Innis, M., Gelfand, D., Sninsky, J., White, T. (Eds.), PCR protocols: a guide to methods and applications. Academic Press, San Diego, pp. 315-322.

Whitting, M.F., 2002. Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. Zool. Scr. 31, 93-104.

Wilken, D.H., 1977. Local differentiation for phenotypic plasticity in the annual Collomia linearis (Polemoniaceae). Syst. Bot. 2, 99-108.

Wilken, D.H., Smith, D.M., Harborne, J.B., Glennie, C.W., 1982. Flavonoid and anthocyanin patterns and the systematic relationships in *Collomia*. Biochem. Syst. Ecol. 10, 239-243.

Yoder, A.D., Irwin, J.A., Payseur, B.A., 2001. Failure of the ILD to determine data combinability for slow Loris phylogeny. Syst. Biol. 50, 408-424.

Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. University of Texas at Austin, Austin. TX, p. 125.

CHAPTER 2:

Infrageneric classification of *Collomia* revisited (Polemoniaceae)

INTRODUCTION

The genus *Collomia* (Polemoniaceae) was erected by Nuttall (1818) for Polemoniaceae species having bracteated inflorescences, alternate, entire leaves, and mucilaginous seeds. As plant discovery continued and more species of Polemoniaceae were described, the generic boundaries of *Collomia* blurred. Gray's emphasis on single key characters, such as unequal stamens or single-seeded locules as a delimiting feature for *Collomia*, led first to the inclusion of many unrelated species in this genus (Gray, 1878), followed by abandoning the genus altogether and subsuming it members in *Gilia* (Gray, 1882). Greene, emphasizing calyx morphology, argued for reestablishing *Collomia* at the generic rank (Greene, 1887) and presented the first treatment of this genus circumscribing all species then known that are recognized within *Collomia* today. In *Collomia*, the five, diamond-shaped sepals are joined into a calyx tube by a narrow, chartaceous membrane with the connation extending roughly half the length of the sepal. Where adjacent lobes separate, this chartaceous membrane forms a spout-like projection that distinguishes *Collomia* from other Polemoniaceae. Wherry (1944) reviewed *Collomia* nomenclature and proposed three morphology-based sections to which the 12 then-accepted species belonged: *Collomiastrum, Courtoisia,* and *Collomia*. This classification was subsequently accepted by Grant (1959). As outlined by Wherry, section *Collomiastrum* includes perennial species with seeds lacking spiricles or only sparsely spirilliferous when wet. Section *Courtoisia* includes lobed-leaved annuals with multi-seeded locules and a spreading habit. Section *Collomia* includes the remaining annual species, which are largely linear-leaved although occasional leaf lobes can be found. Currently, the genus consists of 15 species; 14 are distributed natively primarily throughout the western United States and Canada, with a single species endemic to the southwestern flanks of South America. Using Wherry's morphological

circumscriptions, the three species described since the publication of his classification have been placed within appropriate sections (Table 7). While Wherry's infrageneric classification has served as the basis for sampling and discussion in virtually all subsequent studies of this genus, multiple, independent genetic markers and morphological surveys suggest a revised infrageneric classification is needed.

The revised infrageneric classification of *Collomia* we present here is based on a phylogenetic hypothesis derived from a combined analysis of morphological and DNA sequence data for all 15 species (Figs. 4, 7). While we retain Wherry's original three sections, only sections *Collomiastrum* and *Courtoisia* are maintained as he proposed. Wherry's section *Collomia* is not monophyletic. Given that taxonomies are often used as proxies for phylogenies, we remove *C. tinctoria* and *C. tracyi* from section *Collomia* and erect a new section for these two species based on accumulated evidence, and also remove *C. grandiflora* and *C. biflora* from section *Collomia*. Here we present a key to the sections and species of *Collomia* and a revised, section-level classification of the genus.

KEY TO SECTIONS AND SPECIES OF *COLLOMIA*

1a. Plant perennial; stems arising from rhizomes or deep taproots and subsurface crowns ..... 2 (1. sect. *Collomiastrum*)

2a. Plant prostrate or mound-forming; above-ground stems sprawling and generally less than 1.5 dm long; generally of talus slopes..... 3

3a. Principal leaves (3--)5--7 or more irregularly lobed with some or all of the primary lobes again divided; corolla mostly 12--15 mm long.....*C. larsenii*

3b. Principal leaves entire to 3--5(--7) toothed or lobed with the lobes entire;

Corolla mostly 15--35 mm long.....*C. debilis*

2b. Plant erect, above ground stems generally greater than 1.5 dm long; of rocky, forest meadow, or riparian habitats..... 4

4a. Leaves lanceolate; blade ± 4--5 × longer than wide, entire to few toothed apically; corolla blue to violet; endemic to southern Oregon..... *C. mazama*

4b. leaves elliptic; blade ± 2 × longer than wide, with irregular toothed margin; flowers deep orange; endemic to Madera and Mariposa Co., California..... *C. rawsoniana*

1b. Plant annual; stems arising from taproots and surface crowns..... 5

5a. Leaves elliptic and, at least near the base, usually 3-toothed or deeply pinnate lobed, blades mostly 3--4 times longer than broad or less; locules 2--3 seeded.....6 (2. sect. *Courtoisia*)

6a. Principal leaves 1--2 × pinnate lobed (infrequently entire); corolla tube yellow to light pink, lobes light pink to white; of varied substrates, California to Idaho..... *C. heterophylla*

6b. Leaves 3-toothed below, entire above; corolla tube purple, lobes deep pink; of serpentine substrates in North Coast Ranges, California..... *C. diversifolia*

5b. Leaves narrowly lanceolate and entire, (occasionally few toothed in some species), blades mostly 5--20 × longer than broad; locule 1-seeded..... 7

7a. Internode trichomes uniformly stipitate glandular; Calyx lobes ± equal, aristate; pollen pantoporate.....8 (3. Sect. *Calyperona*)

8a. Corolla 8--14 mm long; stamens equally inserted in upper throat.....

C. tinctoria

8b. Corolla 12--22 mm long, stamens unequally inserted in corolla tube.....

C. tracyi

7b. Internode trichomes, at least at mid plant, including short, white, eglandular retrorse hairs (except in *C. tenella*, which is uniformly glandular, but then flowers borne singly rather than in clusters or heads); calyx lobes ± equal and acute, somewhat unequal and acuminate, or strongly unequal with needle-like tips greater than half the lobe length; pollen zonocolporate..... 9

9a. Corolla lobes red, salmon, or yellowish fading to white, or some or all flowers cleistogamous..... 10

10a. Corolla greater than 15 mm in length, salmon or yellowish fading to white, or some or all flowers cleistogamous and less than 4 mm long; gametic chromosome number $n = 8$; indigenous to North Amer..... *C. grandiflora*

10b. Corolla generally 10--15 mm long, lobes red, (infrequently yellow); gametic chromosome number $n = 16$; indigenous to South America..... *C. biflora*

9b. Corolla lobes pink, blue, lavender, or white..... 11 (4. Sect. *Collomia*)

11a. Calyx lobes ± equal in fruit and less than 4 mm long..... 12

12a. Corolla glandular externally; longest filaments < 1mm; anthers included in corolla throat..... *C. linearis*

12b. Corolla generally glabrous externally (infreq. glandular in *C.*

wilkenii); longest filaments 1--2 mm; anthers of longest stamens exserted to mid corolla lobe..... 13

13a. Flowers borne singly; corollas 4--6 mm; stamens inserted equally; gametic chromosome number *n* = 8..... *C. tenella*

13b. Flowers borne in clusters of 2--6; corollas (7--) 9--10 mm; stamens inserted subequally to unequally; gametic chromosome number *n* =16..... *C. wilkenii*

11b. Calyx lobes somewhat to strongly unequal in fruit with the longest greater than 4 mm long.....14

14a. Calyx lobes strongly unequal in fruit, the lobes accuminate to needle-like for at least half their length, terminating in a (sometimes bent) pointed tip; corolla lobes blue when fresh, tube sparsely glandular; some pubescence throughout plant body glandular..... *C. macrocalyx*

14b. Calyx lobes somewhat unequal in fruit, the lobes narrowly acute (but rounded-tipped under magnification); corolla lobes pink when fresh, tube glabrous; pubescence throughout plant body eglandular..... *C. renacta*

SECTIONAL CIRCUMSCRIPTIONS

1. Collomia Nutt. sect. **Collomiastrum** Brand, Pflanzenr. 4 (250): 52. 1907. *Collomiastrum* (Brand) S. L. Welsh, A Utah Flora ed. 3. 480. 2003. TYPE: *Collomia debilis* (S. Watson) Greene.

Section *Collomiastrum* includes *Collomia debilis, C. larsenii, C. mazama,* and *C. rawsoniana.* It is a strongly supported monophyletic group differentiated from other *Collomia* species not only by their duration and rhizomatous habit, but also by subtle differences in pollen morphology (zonotreme apertures with striated ridges, though C. grandiflora shares this morphology; Chuang et al., 1978), seed coat variation (type 1 seeds with hexagonal epidermal cells with distinct cell boundaries, and the reduction or loss of mucilage; Hsiao and Chuang, 1981), and apparent self-incompatibility. Chemically, this group is characterized, relative to other *Collomia,* by the absence of K-3-arabinosylgalactoside and Q-5-glucoside (though the latter is also absent in the South American species; Wilken et al., 1982). Despite these differences, the members of *Collomiastrum* share the unique calyx morphology, base chromosome number, and explosively dehiscent capsule characteristic of this genus. Phylogenetically, *Collomiastrum* is embedded within the genus, rather than sister to all other *Collomia*. Given that tribe Gilieae exclusive of these species is itself composed solely of annual species, this indicates the perennial habit was derived after the early diversification of the genus rather than being ancestral, and also argues for the retention of the perennial species within *Collomia*, rather than elevating this section to generic status, as proposed by Welsh (Welsh et al., 2003).

2. Collomia sect. **Courtoisia** (Reichenbach) Wherry. E. in Am. Midl. Nat. 31: 216. *Courtoisia* Reichenbach 1829. Gilia sect. Courtoisia (Reichenbach) A. Gray 1882. TYPE: *Collomia heterophylla* Hooker.

Courtoisia includes *Collomia diversifolia* and *C. heterophylla.* These two species share a similar open, well-branched habit and a distinguishing synapomorphy of multi-ovulate locules; all other *Collomia* possess a single ovule per locule. Chloroplast DNA agrees with ovule number in circumscribing these species in a single, monophyletic group. Nuclear DNA sequences neither strongly support nor strongly contradict the phylogenetic placement of these species as sister taxa, an outcome that is not altogether surprising given that the pollen morphology of *C. heterophylla* is distinct in *Collomia* (spheroidal, pantoporate with irregularly reticulate ridges) whereas the grains of *C. diversifolia* are similar to those possessed by species in section *Collomia* (speroidal, zonocolporate with striato-reticulate ridges; Chuang et al., 1978). Perhaps *C. diversifolia*, early in the diversification of *Collomia*, introgressed with the ancestor of *C. heterophylla* (or in the other direction, with an ancestral member of section *Collomia*), or possibly the species originated via homoploid hybrid speciation. Present data are inconclusive for distinguishing among these alternative hypotheses.

3. Collomia Nutt. sect. **Calyperona** E. S. Green & L. A. Johnson, sect. nov. TYPE: *Collomia tinctoria* Kellogg.

Section nova ab Collomia section Collomia differt calycibus aristatus nec acutatus - acuminatus y pollinis granum pantoporate nec zonocolporate.

Section *Calyperona* includes *Collomia tinctoria* and *C. tracyi*. Though possessing long, lanceolate leaves typical of species in section *Collomia*, *Calyperona* are readily distinguished from species in this and other sections of *Collomia* by a combination of morphological characters. In gross morphology, their calyx lobes are aristate rather than acute and their nearly salverform corollas differ from the funnelform corollas common in sections *Collomia* and *Collomiastrum*, but approached by the typically smaller flowers of section *Courtoisia*. The pantoporate pollen grains with deep radiating lirae are unique not only in *Collomia*, but within Polemoniaceae. DNA sequence data show these species as well removed phylogenetically from section *Collomia* (Figs. 4, 7). The section name is derived from the Greek calycis, in reference to the calyx, and perone, "pin; anything pointed or piercing."

4. Collomia sect. **Collomia**, – Gray, A. in Fl. N.A. (ed. 2) 1(2) and 2(1) Suppl.: 407 1886. Gilia section Collomia (A. Gray) A. Gray 1886. Type: *Collomia linearis* Nuttall.

Section *Collomia* is here defined to include *C. linearis, C. macrocalyx, C. renacta, C. tenella,* and *C. wilkenii.* Gray's (1870) *Collomia* section *Eucollomia* included a diversity of *Collomia* and non-*Collomia* species, but excluded many non-*Collomia* species he included in *Collomia* section *Phloganthea*. In synonymzing *Collomia* with *Gilia*, Gray's *Gilia* section *Collomia* approached the modern concept of this section although it still included one non-*Collomia* species. Though section *Collomia* is defined largely by lacking the features that characterize the other sections, the species share a linear leaved habit and zonocolporate ovate pollen grains with striate exine (Chuang et al., 1978). Of the included species, *C. tenella* is unusual in *Collomia* for having a single flower per inflorescence. It has an open branching habit

possessed by *Collomia* sect. *Calyperona*, and shares with those species the absence of white, retrorse, eglandular hairs on the internodes, and several anthocyanins, such as A-7-glucoside, P-3-galactoside, P-5-glucoside, and Q-3-rhamnosylglucosylglucoside not found in other members of section *Collomia* (Wilken et al., 1982). However, pollen morphology is consistent with DNA sequence variation, strongly placing this species in section *Collomia*. One of two known allopolyploid *Collomia* species, *C. wilkenii*, is the product of hybridization between *C. tenella* and *C. linearis* (of section *Collomia*) in the relatively recent, though not necessarily historical, past (Johnson and Johnson, 2006).

5. Unplaced species

Collomia grandiflora and *C. biflora* are here excluded from any section. *Collomia grandiflora* has historically resided in section *Collomia* and is phylogenetically close (Figs. 4, 7), but not part of the well-supported clade defined here as section *Collomia*. Total evidence tenuously places it parsimoniously within section *Collomia*, but with little support (bootstrap value <50) while Bayesian analyses have *C. grandiflora* sister to *Collomiastrum* (posterior probability =50). The ambiguous placement with DNA for this species is paralleled to a degree with morphology, where *C. grandiflora* shares the striate reticulate exine of the perennial species (Chuang et al., 1978) and was suggested by those authors to be intermediate between the annuals of section *Collomia* and the perennials of section *Collomiastrum*. *Collomia grandiflora* is the only *Collomia* species with both chasmogamous and cliestogamous flowers. Leaving *C. grandiflora* unplaced is thus consistent with available evidence, draws attention to the

distinctiveness of this species relative to other species of section *Collomia*, and avoids erecting a monotypic section, redundant taxonomically with this species alone.

Collomia biflora is the second allopolyploid species in *Collomia*, with the ancestors of *Collomia grandiflora* and *C. linearis* participating as the maternal and paternal genome donors, respectively (Green & Johnson, submitted). With one parent clearly placed in section *Collomia* and one parent excluded from any section, we here also exclude *C. biflora* from any section given its phylogenetic chimerism.

REFERENCES

Chuang, T.I., Hsieh, W.C., Wilken, D.H., 1978. Contribution of pollen morphology to systematics of Collomia (Polemoniaceae). Am. J. Bot. 65, 450-458.

Grant, V., 1959. Natural history of the phlox family. Martinus Nijhoff, The Hague, Netherlands.

Gray, A., 1870. Revision of the North American Polemoniaceae. Proc. Amer. Acad. Arts Sci. 8, 247-282.

-------, 1878. Syn. Flora N. A. 2, 134.

-------, 1882. Contributions to North American botany. Proc. Amer. Acad. Arts Sci. 17, 223.

Greene, E.L., 1887. American Polemoniaceae. Pittonia: a series of papers relating to botany and botanists. Cubery & Co., San Fransisco (CA), pp. 120-139.

Hsiao, Y., Chuang, T.I., 1981. Seed-coat morphology and anatomy in Collomia (Polemoniaceae). Am. J. Bot. 68, 1155-1164.

Johnson, L.A., Johnson, R.L., 2006. Morphological delimitation and molecular evidence for allopolyploidy in *Collomia wilkenii* (Polemoniaceae), a new species from northern Nevada. Syst. Bot. 31, 349-360.

Nuttall, T., 1818. Gen. N. Am. Pl. 1, 126.

Welsh, S.L., Atwood, N.D., Goodrich, S., Higgins, L.C., 2003. A Utah Flora. Brigham Young University, Provo, UT.

Wherry, E.T., 1944. Review of the genera Collomia and Gymnosteris. Am. Midl. Nat. 31, 216-231.

Wilken, D.H., Smith, D.M., Harborne, J.B., Glennie, C.W., 1982. Flavonoid and anthocyanin patterns and the systematic relationships in *Collomia*. Biochem. Syst. Ecol. 10, 239-243.

Acknowledgements: J. Mark Porter provided helpful insights regarding nomenclatural issues and specimens loaned from Rancho Santa Ana Botanic Garden (RSA and POM) were consulted for this study. This research was supported by NSF grant DEB-0344837.

TABLES

Table 1: Sequence alignment metrics for "monophyly" and "infrageneric" matrices.

Comparison	Monophyly Matrix and Components			
	cpDNA	ITS	*PI*	Combined
Length of gene alignment	5077	637	1349	7063
Number of indels	187	23	95	305
Total number of characters	5264	660	1444	7368
Number potential informative characters	366	85	173	624
Number of most parsimonious trees	3	1180	19	6
MPT length for matrix	1911	263	741	1895
CI (autapomorphies retained)	0.77	0.66	0.73	0.77
RI	0.82	0.74	0.75	0.83

	Infrageneric Matrix and Components					
	cpDNA	ITS	*idhA*	*g3pdh*	*PI*	Combined
Length of gene alignment	7129	632	1311	1230	1343	11955
Number of indels	124	10	44	67	65	310
Total number of characters	7253	642	1355	1297	1408	11955
Number potential informative characters	402	80	210	227	224	1141
Number of most parsimonious trees	2	5	48	3	32	12
MPT length for matrix	584	135	382	444	417	2043
CI (autapomorphies retained)	0.91	0.81	0.84	0.77	0.84	0.80
RI	0.98	0.93	0.91	0.89	0.94	0.92

Table 2: Recombination sites detected by RDP3 in the IDH gene in C. biflora, the probable parental sources, and the method (p-value) used to detect recombination

Col_biflora_Sers_Gtype	major parent = Col_grandiflora_04151 minor parent = Col_linearis_04104	
	Recombination start point (estimated) = 40 Recombination end point (estimated) = 930 891 bp of C. linearis	
	Methods of detection	MaxChi ($p = 3.086 \times 10^{-03}$)
		Chimaera ($p = 4.293 \times 10^{-03}$)
		SiScan ($p = 1.700 \times 10^{-08}$)
		3Seq ($p = 6.818 \times 10^{-04}$)
Col_biflora_Sers_Ltype	major parent = Col_linearis_92045 minor parent = Col_grandiflora_04151	
	Recombination start point (estimated) = 443 Recombination end point (estimated) = 886 444 bp of C. grandiflora	
	Methods of detection	MaxChi ($p = 1.981 \times 10^{-02}$)
		3Seq ($p = 5.925 \times 10^{-02}$)
Col_biflora_156_Gtype	major parent = Col_grandiflora_04051 minor parent = Col_linearis_92045	
	Recombination start point (estimated) = 1116 Recombination end point (estimated) = 1311 196 bp of C. linearis	
	Method of detection	Manually looked at alignment and marked points where 2 types differed
Col_biflora_156_Ltype	major parent = Col_linearis_92045 minor parent = Col_grandiflora_04051	
	Recombination start point (estimated) = 472 Recombination end point (estimated) = 886 415 bp of C. linearis	
	Methods of detection	MaxChi ($p = 1.981 \times 10^{-02}$)
		3Seq ($p = 5.925 \times 10^{-02}$)

Table 3: a) P-values of ILD partition homogeneity test testing congruency between data sets in matrix 1 testing monophyly. b) P-values of ILD partition homogeneity test testing congruency between data sets in matrix 2 testing intrageneric relationships. Tests done with 100 replicates and TBR implementing the partition homogeneity test in PAUP*

a)			b)				
cpDNA			cpDNA				
ITS	0.25		ITS	0.01*			
PI	0.01*	0.65	IDH	0.01*	0.01*		
			GPD	0.01*	0.01*	0.01*	
			PI	0.01*	0.01*	0.80	0.18

* Values of $\alpha < 0.05$ denote statistical significance between the two data sets and the hypothesis of data set congruence is rejected.

Table 4: Number of additional steps / probability values of SLP_T pairwise comparison between each gene partition parsimony tree and a constraint tree. Random number generator used to determine which pairwise score (if multiple comparisons made) reported, Bonferroni corrections made.

Constraint	ITS	GPD	IDH	PI	cpDNA
((b,ln,w,rn,mc),all)	4 / 0.1025	0 / 1.0000	1 / 0.7389	3 / 0.3173	80 / < 0.0001*
((g,b),all)	23 / <0.0001*	0 / 1.0000	1 / 0.7055	3 / 0.3173	23 / <0.0001*
((ln,b,w),all)	2 / 0.3173	0 / 1.0000	1 / 0.7055	1 / 0.6547	90 / <0.0001*
((b,g,ln,mc,rn,tn,w),all)	0 / 1.0000	1 / 0.7389	1 / 0.7389	2 / 0.2059	21 / <0.0001*
((db,la,mz),all)	1 / 0.5637	0 / 1.0000	0 / 1.0000	3 / 0.3173	40 / <0.0001*
((db,la,rw),all)	0 / 1.0000	10 / 0.0075*	1 / 0.7389	3 / 0.3173	0 / 1.0000
((db,la,rw,mz),all)	0 / 1.0000	2 / 0.5637	0 / 1.0000	3 / 0.3173	21 / <0.0001*
(((db,la),(rw,mz)),all)	N/A	49 / 0.0011*	N/A	0 / 0.6547	39 / <0.0001*
((ln,tn,w,rn,mc,mz),all)	8 / 0.0209*	24 / 0.0001*	20 / <0.0001*	26 / <0.0001*	0 / 1.0000
(dv,(all))	2 / 0.3173	0 / 1.0000	3 / 0.5485	5 / 0.1655	38 / <0.0001*
((dv,h),all)	2 / 0.3173	6 / 0.0399*	4 / 0.2850	3 / 0.2568	0 / 1.0000
(h,(all))	0 / 1.0000	7 / 0.0196*	3 / 0.5485	3 / 0.0833	39 / <0.0001*
(((ti,tr),h),all)	2 / 0.3173	10 / 0.0039*	11 / 0.0477*	0 / 1.0000	47 / <0.0001*

* Values of $\alpha < 0.05$ denote statistical significance between the two topologies the hypothesis of topological congruency is rejected.

b= biflora, db= debilis, dv= diversifolia, g= grandiflora, h= heterophylla, la= larsenii, ln= linearis, mc= macrocalyx, mz= mazama, rn= renacta, rw= rawsoniana, ti= tinctoria, tn= tenella, tr= tracyi, w= wilkenii

Table 5: Nodal support values for parsimony topology in Fig. 2 testing monophyly of *Collomia*. Columns list non-parametric parsimony bootstrap values, maximum likelihood parsimony values, Bayesian posterior probabilities, Bremer support values, and partitioned Bremer support values as calculated for the simultaneous analysis. Bootstrap support values result from 1000 replicates and the ML and Bayesian values are based specified models of evolution.

Node	Boot	ML	Bayes	Total Bremer	Partitioned Bremer Support Values						
					cp2	cp8	trnK	trnL	trnS	ITS	PI
1	100	100	100	41	9	4	8	2	8	11	-1
2	95	98	100	5.99	0.83	2.83	1.33	1	-0.17	1	-0.83
3	100	100	100	19	3	6	3	2	0	3	2
4	100	100	100	27	2	6.5	0.5	7	3	3	5
5	100	100	100	13	1	2	0	0	4	0	6
6	100	100	100	33	0	5.5	0.5	4	3	4.5	15.5
7	51	<50	<50	1	0	0	-0.5	-1.5	-1.5	0.5	4
8	52	<50	<50	1	0	0	-0.5	-1.5	-1.5	0.5	4
9	56	60	90	1	0	0	0.5	0	-0.5	0.5	0.5
10	61	65	97	1	0	-1	-1	0	0	1	2
11	84	75	100	4	-0.5	-1	-0.5	-0.5	1	1.5	4
12	99	100	100	11	1.75	0.25	1.25	3.5	2.25	1.5	0.5
13	99	100	100	12.99	0.33	0.33	1.33	-2	4.33	5	3.67
14	100	99	100	21	0	1	-1	-1	13	3	6
15	100	100	100	24	10	3	3	5	5	0	-2
16	100	100	100	37.99	5.83	2.83	6.33	3	6.83	4.5	8.67
17	<50	52	75.3	0	0.5	0	0	0	0.5	-0.5	-0.5
18	86	85	100	1.99	-1.17	0.33	-0.67	-1	-0.17	2.5	2.17
19	100	100	100	15.99	0.33	6.33	0.33	4	5.33	1	-1.33
20	<50	<50	<50	0	-0.17	-0.17	0.33	0	-0.17	0	0.18
21	69	73	98	1	0	0	0	0	-1	0	2
22	100	100	100	19	3.17	2.33	2.33	-0.33	1.83	3.17	6.5
23	100	100	100	34.99	5.33	6.33	3.33	3	7.33	2	7.67
24	71	<50	<50	2	2	2	0	0	1	0	-3
25	92	81	100	3.99	-0.67	-0.67	0.33	0	1.33	1	2.67
26	100	100	100	114	19.5	19.17	12	21	11.83	26.67	3.83
27	99	96	100	2	-2	-1	-1	0	2	0	4

Table 6: Nodal support values for parsimony topology in Fig. 4. Columns list parametric parsimony bootstrap values, maximum likelihood bootstrap values, Bayesian posterior probabilities, Bremer support values, and partitioned Bremer support values as calculated for the simultaneous analysis. Bootstrap support values result from 1000 replicates and the ML and Bayesian values are based on specified models of evolution.

Node	Pars Boot	ML Boot	Bayes PP	Total Bremer	Partitioned Bremer Support Values				
					cpDNA	ITS	GPD	IDH	PI
1	100	100	1.00	41	1	2	0	22	16
2	100	100	1.00	19	0	-0.5	1.5	7	11
3	100	100	1.00	34	2	4	17	-1	12
4	86	90	0.99	2	1	0	1	0	0
5	100	100	1.00	62	5	7	22	15	13
6	100	100	1.00	37.05	40.68	0	-4.3	0	0.67
7	100	100	1.00	54	2.5	7	10.5	25	9
8	100	100	1.00	22	4	0	-1	9	10
9	60	65	1.00	2	-1	0	0	-10	13
10	88	85	1.00	6	0	7	0	2	-3
11	100	100	1.00	64	45	3	6	1	9
12	100	99	1.00	16	0	0	0	4	12
13	71	<50	0.64	2	0	0	0	0	2
14	99	99	1.00	5	3	1	1	0	0
15	94	94	1.00	3	0	0	0	0	3
16	82	62	0.79	2	0	0	0	0	2
17	71	<50	0.64	2	0	0	0	0	2
18	100	99	1.00	30	15	3	1	4	7
19	100	100	1.00	12	7	0	0	5	0
20	97	97	1.00	7	-3	0	3	3	4
21	100	100	1.00	33	17	0	11	3	2
22	100	100	1.00	42	41.93	0.1	0.07	-0.1	0
23	82	81	0.99	1	0	0	0	1	0
24	100	99	1.00	7.58	7.48	0	0.1	0	0
25	91	79	0.99	1	0	0	0	1	0
26	100	100	1.00	12	12	0	0	0	0
27	100	100	1.00	10	10	0	0	0	0
28	67	64	1.00	1	-1	0	0	2	0
29	55	66	0.96	1.33	0.33	0	0	1	0
30	75	71	0.94	1.33	0.33	0	0	1	0
31	99	99	1.00	14	0	2	12	-1	1
32	100	99	1.00	22	20.33	1	0.67	0	0
33	<50	<50	<0.50	1	0.5	0	1	0	-0.5
34	100	100	1.00	59	37	1	5	12	4
35	91	95	1.00	7.02	6.68	-1	3.67	4	-6.33
36	100	100	1.00	191	103	26	31	18	13
37	74	62	1.00	4	-1	-1	5	9	-8

Table 7: Wherry's original sections with placement known *Collomia* species within respective section

Section Collomiastrum	**Section Collomia**	**Section Courtoisia**
C. debilis	*C. biflora*	*C. diversifolia*
C. larsenii	*C. grandiflora*	*C. heterophylla*
C. mazama	*C. linearis*	
C. rawsoniana	*C. macrocalyx*	
	C. renacta (identified 1986)	
	C. tenella	
	C. tinctoria	
	C. tracyi (identified 1948)	
	C. wilkenii (identified 2006)	

Appendix A: Morphological characters for *Collomia,* and *Allophyllum*. The pleisiomorphic state (0) is the state at the outgroup node in the most parsimonious tree of the combined analysis. States are unordered.

1. Calyx accretion: 0 = absent, 1 = present (sepal lobes fused together)
2. Calyx sinus plication: 0 = absent; 1 = present (spout-like projection from sinus where sepal lobes meet)
3. Capsule explosively dehiscent: 0 = no; 1 = yes
4. Chromosome base number: 0 = x=8; 1 = x=9
5. Seeds per locule: 0 = 1; 1 = > 1
6. Lifecycle duration: 0 = annual; 1 = perennial
7. Flowers cleistogamous: 0 = no; 1 = yes
8. Pollen pore distribution: 0 = pantotreme (occur in circular pattern around surface); 1 = zonotreme (occur along equator)
9. Pollen color: 0 = white, 1 = blue
10. Pollen ridge type: 0 = striato-reticulate, 1 = striated, 2 = irregularly reticulate, 3 = radiate
11. Seed mucilage amount: 0 = copious, 1 = scant, 2 = absent
12. Seed type: 0 = type 2 (longitudinally ridged, irregular depressions, inconspicuous cell boundaries); 1 = type 1 (hexagonal epidermal cells, distinct cell boundaries)
13. Calyx apical shape: 0 = acute, 1 = aristate, 2 = acuminate
14. Calyx lobe length equal: 0 = yes, 1 = no
15. Stem hairs retrorse eglandular : 0 = absent, 1 = present (claw-like hairs without glands)
16. Stem hairs glandular : 0 = absent, 1 = present
17. Corollar hairs: 0 = absent, 1 = present all over, 2 = present around throat only
18. Filament insertion: 0 = equally in throat, 1 = unequally in throat
19. Filament length: 0 = equal, 1 = unequal
20. Stamen exerted: 0 = no, 1 = yes
21. Stigma position: 0 = included, 1 = exerted
22. Stems: 0 = simple, 1 = forked, 2 = highly branched
23. Habit: 0 = prostrate/spreading, 1 = erect
24. Rhizomatous: 0 = no, 1 = yes
25. Calyx margin color: 0 = green, 1 = solid purple, 2 = splotchy purple and green
26. Leaf shape: 0 = linear-lanceolate, 1 = broad, 2 = lobed, toothed, or dissected

Appendix B: Morphological character states for coded characters

Taxon/Character	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
C._heterophylla	1	1	1	1	1	0	0	0	1	2	0	0	0	0	0	1	1	1	1	1	0	2	1	0	0	2
C._diversifolia	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	1	1	1	0	0	0	2	1	0	0	2
C._tinctoria	1	1	1	?	0	0	0	0	0	3	0	0	1	0	0	1	2	0	1	1	1	2	1	0	2	0
C._tracyi	1	1	1	?	0	0	0	0	0	3	0	0	1	0	0	1	0	1	1	0	0	1	1	0	2	0
C._rawsoniana	1	1	1	1	0	1	1	1	0	1	1	0	0	0	0	1	1	0	1	1	1	0	1	1	0	1
C._mazama	1	1	1	1	0	1	1	1	0	1	1	0	0	0	0	1	1	0	1	1	1	0	1	1	2	1
C._debilis	1	1	1	1	0	1	1	1	0	1	2	1	0	0	0	1	1	0	0	1	1	0	0	1	2	1
C._larsenii	1	1	1	1	0	1	1	1	0	1	2	1	0	0	0	1	1	0	1	0	0	0	0	1	2	1
C._linearis	1	1	1	1	0	0	0	1	1	0	0	0	0	0	1	0	1	1	0	0	0	0	1	0	0	0
C._tenella	1	1	1	?	0	0	0	1	1	0	0	0	0	0	0	1	0	0	?	1	1	2	1	0	1	0
C._grandiflora	1	1	1	1	0	0	(0,1)	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0	1	0	0	0
C._renacta	1	1	1	?	0	0	0	1	0	0	0	0	2	1	1	0	0	1	0	0	1	0	0	0	2	0
C._macrocalyx	1	1	1	1	0	0	0	1	0	0	0	0	2	1	1	0	1	1	0	0	0	2	1	0	0	0
C._biflora	1	1	1	1	0	0	0	1	1	0	0	0	0	0	1	0	1	1	0	0	0	0	1	0	0	0
C._wilkenii	1	1	1	1	0	0	0	1	0	0	0	0	0	0	1	1	0	1	1	1	0	2	1	0	1	0
A._divaricatum	0	0	0	(0,1)	1	0	0	0	0	4	0	0	0	0	?	?	?	?	?	?	?	?	?	?	?	?
A._gilioides	0	0	0	0	1	0	0	0	0	4	0	0	0	0	?	?	?	?	?	?	?	?	?	?	?	?
A._integrifolium	0	0	0	(0,1)	(0,1)	0	0	0	0	4	0	0	0	0	?	?	?	?	?	?	?	?	?	?	?	?

FIGURES

Figure 1. Representative trees from individual parsimony (ACCTRAN optimization) analyses of combined chloroplast, ITS1-5.8S-ITS2, and *PI* regions. a) One of 3 most parsimonious cpDNA trees. b) One of 1180 most parsimonious ITS trees. c) One of 19 most parsimonious *PI* trees.

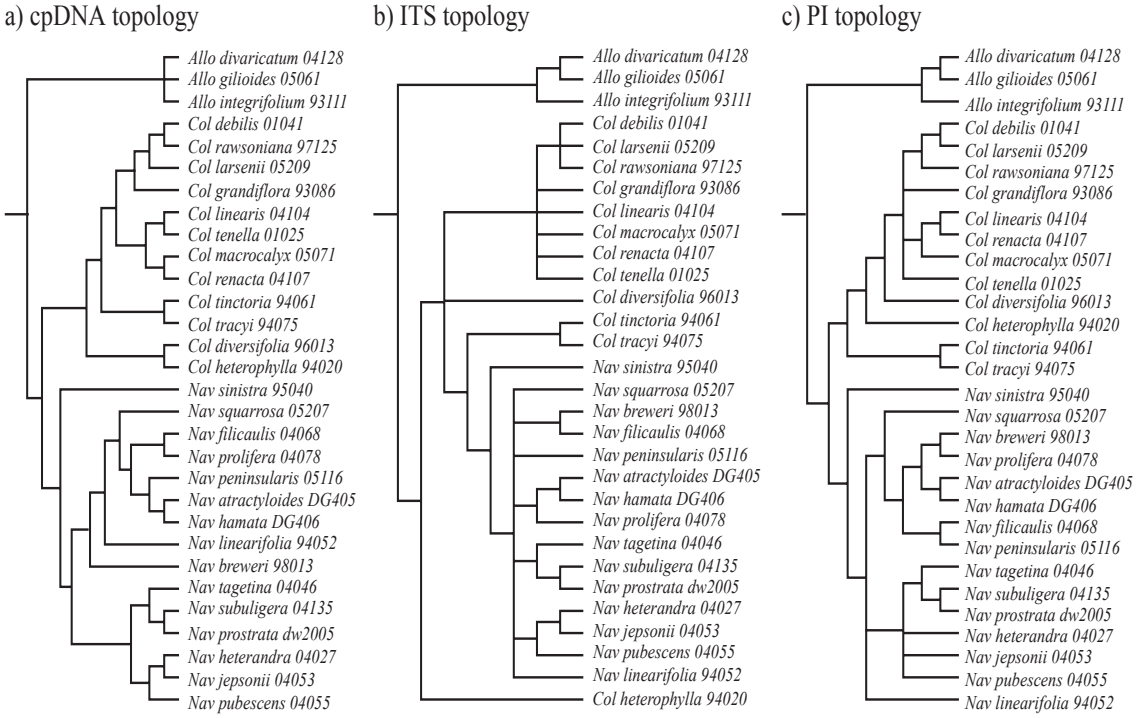

Figure 2. Representative tree from simultaneous parsimony (ACCTRAN optimization) analysis of regions. Nodal labels correspond with Table 5. Bootstrap and posterior probability values shown above branches for corresponding node (parsimony/ likelihood/ Bayesian).

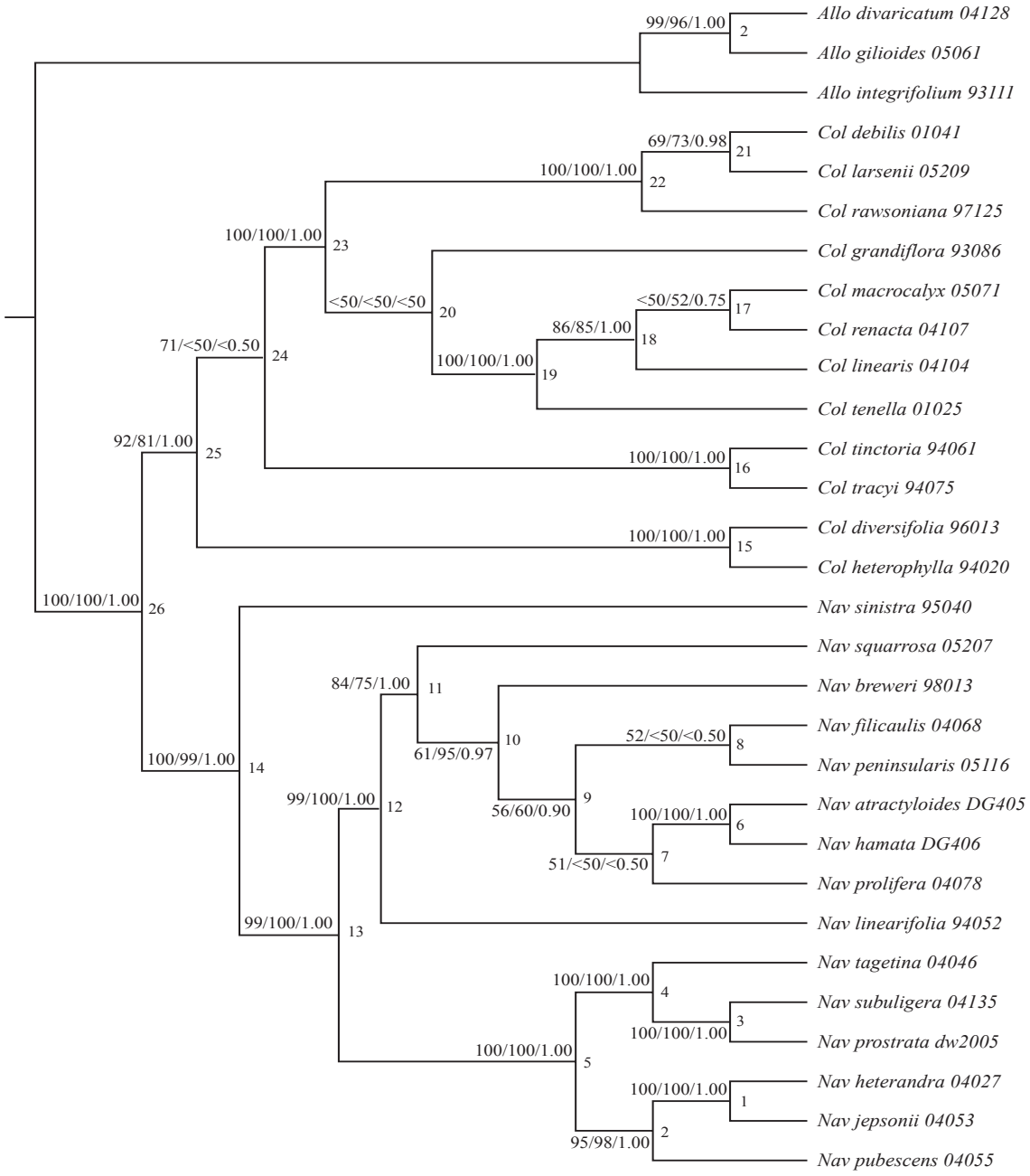

Figure 3. Representative trees from individual parsimony (ACCTRAN optimization) analyses of combined chloroplast, ITS1-5.8S-ITS2, *idhA*, *g3pdh*, *PI* regions, and morphology. a) One of 2 most parsimonious cpDNA trees. b) One of 5 most parsimonious ITS trees. c) One of 48 most parsimonious *idhA* trees. d) One of 3 most parsimonious *g3pdh* trees. f) A semi-strict consensus of 155 most parsimonious morphology trees.

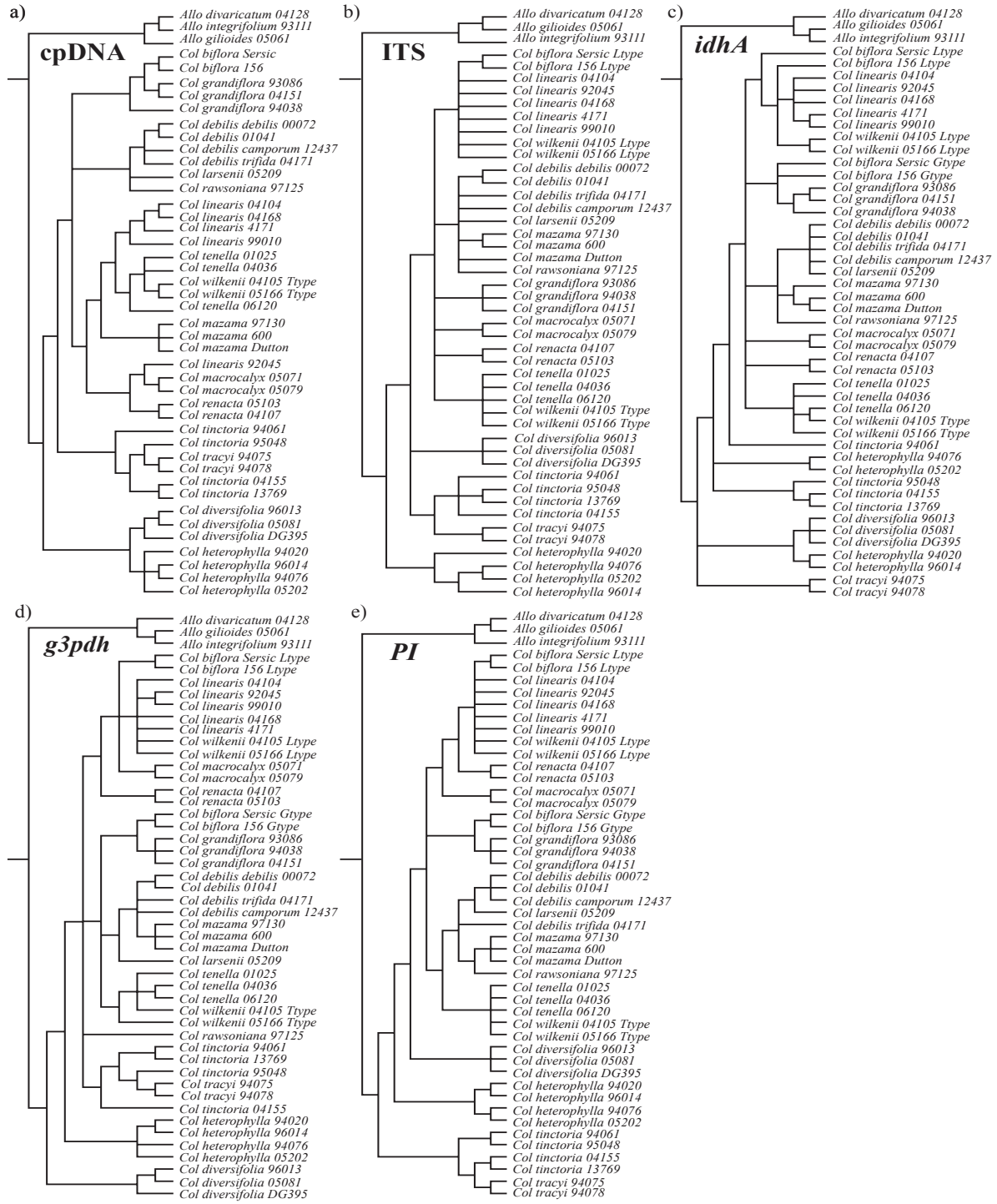

Figure 4. Strict consensus of 12 simultaneous parsimony (ACCTRAN optimization) analysis of all loci. Nodal labels correspond with Table 6. Bootstrap and posterior probability values shown above branches for corresponding node (parsimony/ likelihood/ Bayesian). The dashed line represents branches not found in all parsimonious trees.

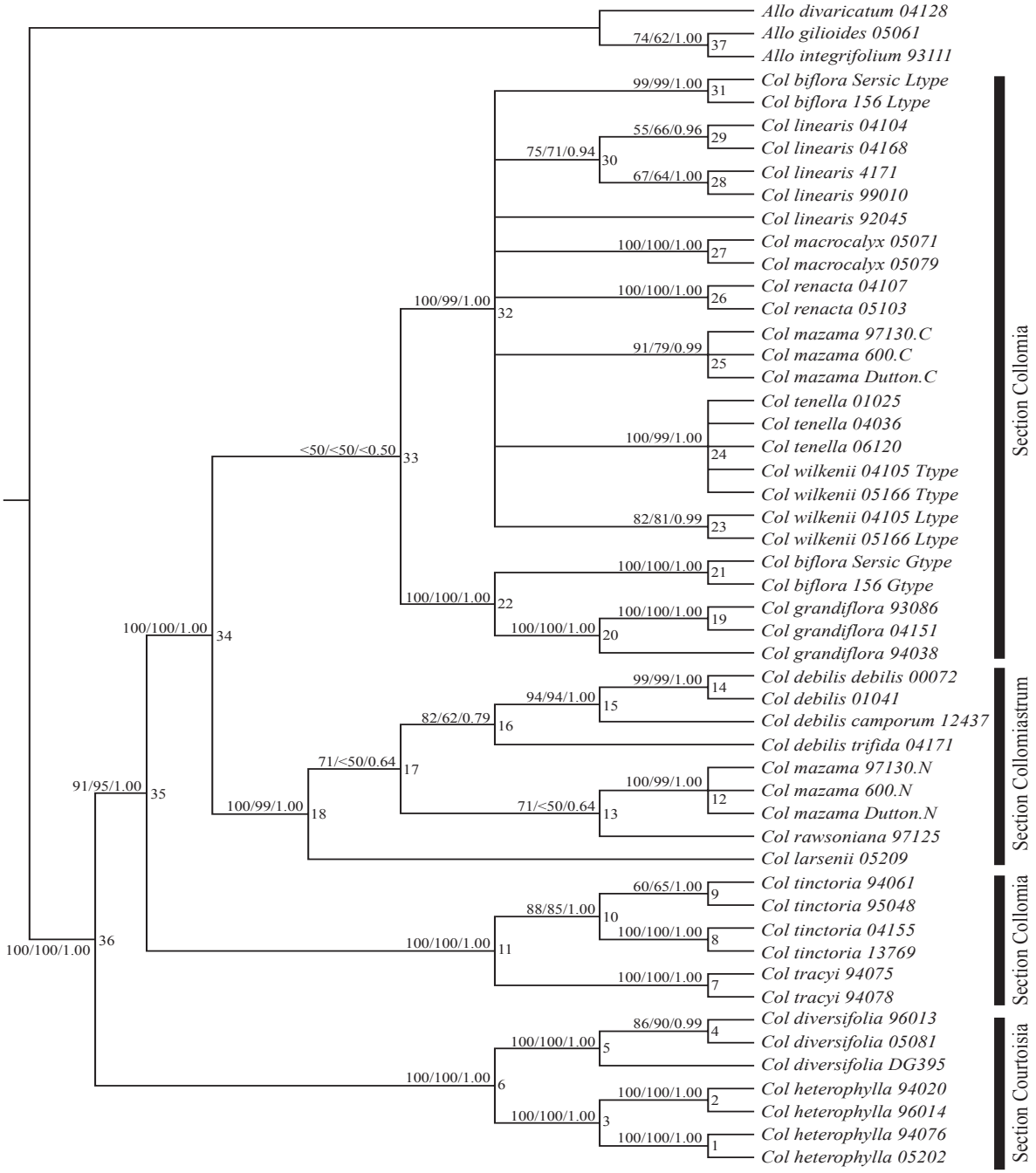

Figure 5. Evolution of morphological characters in genus *Collomia*. Topologies based on multilabeled parsimony analysis, with only paternal allopolyploid contributor and nuclear genes for *C. mazama* mapped. a) lifecycle duration and seed morphology. b) pollen grain morphology.

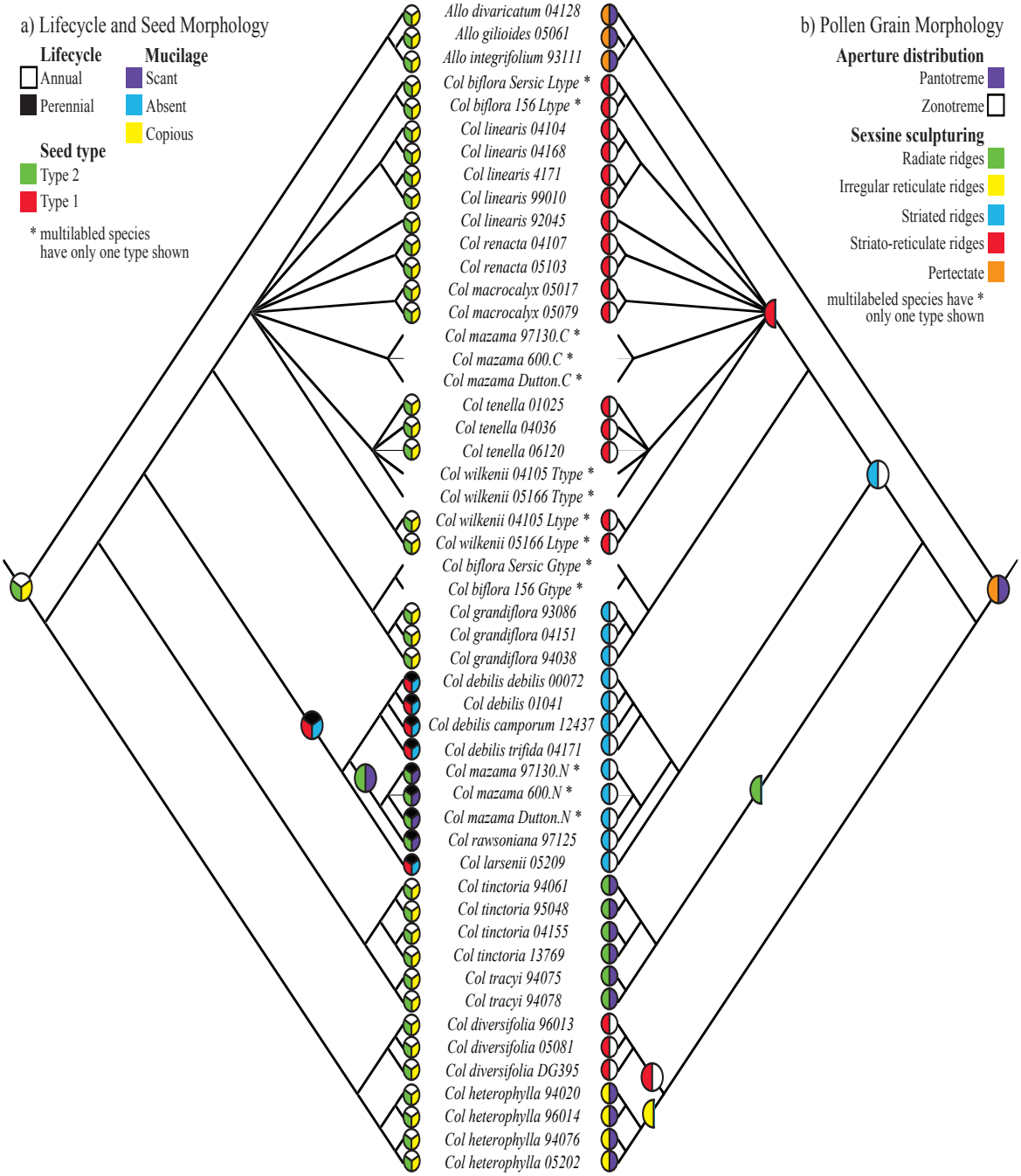

Figure 6. With the detection of recombination between the homeologs, the recombinant pieces where located and removed from the *C. biflora* gene and reassembled using the together with their respective parts to make a parental *idhA* gene and the *C. biflora* gene with the recombinant segment missing. *idhA* gene phylogeny places both the pieced together gene and the *C. biflora gene* in its respective location on the topology relative to the parental type.

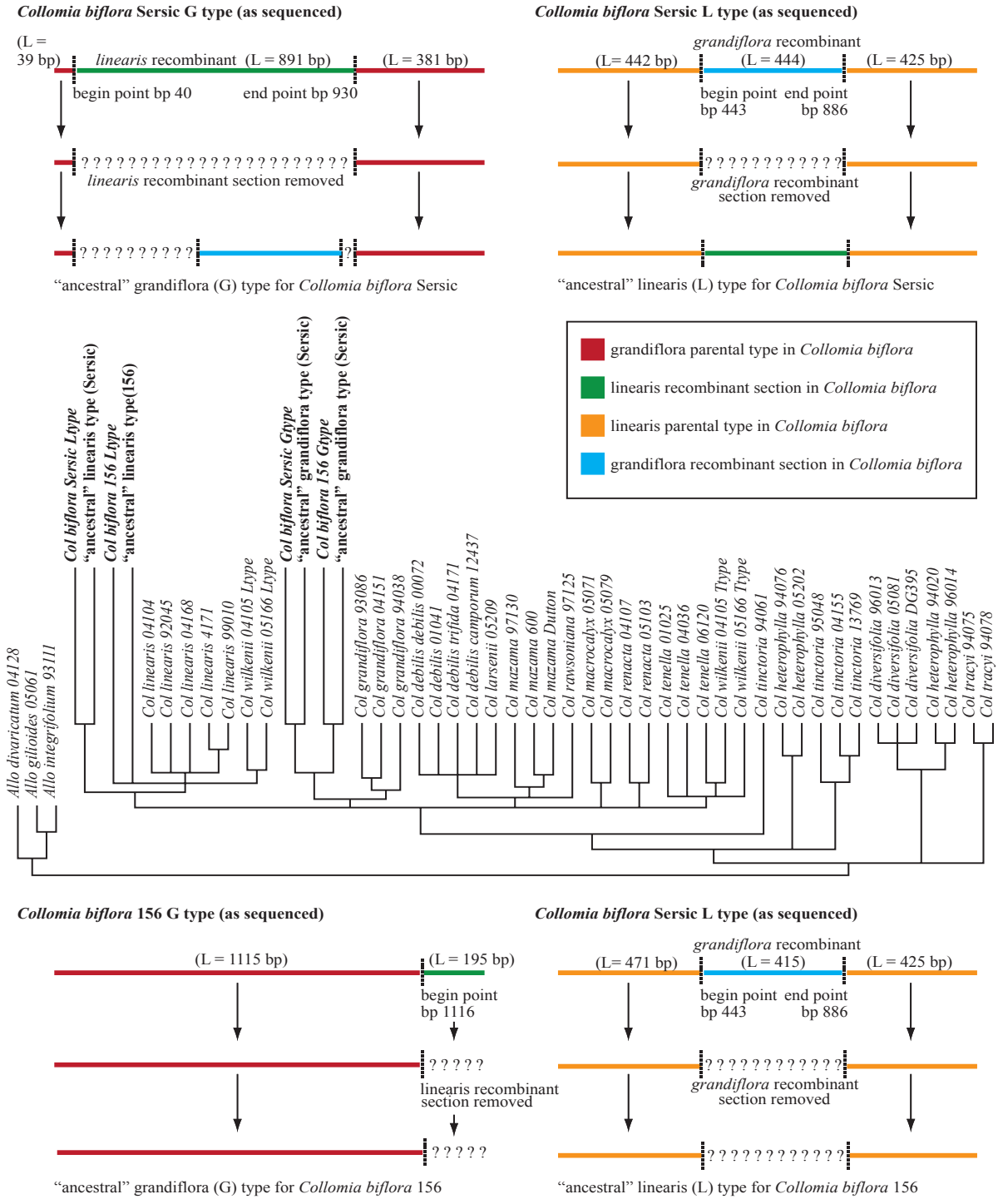

Figure 7. A species network manually constructed from multilabeled simultaneous parsimony analysis of all genomic and morphological data illustrating biologically meaningful relationships between reticulate species and the parental types in *Collomia*. Different colors represent different genomic contributions to the reticulate species.