Phylogenetic relationships, systematics, character-associated diversification, and chloroplast genome evolution in *Asarum* (Aristolochiaceae).

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Brandon Tyler Sinn, B.S.

Graduate Program in Evolution, Ecology and Organismal Biology

The Ohio State University

2015

Dissertation Committee:

Dr. John V. Freudenstein, Advisor

Dr. Marymegan Daly

Dr. Johannes S. Klompen
Copyright by
Brandon Tyler Sinn
2015
The flowering plant genus *Asarum* (Aristolochiaceae) comprises approximately 110 species in north temperate forests worldwide. This dissertation represents the most comprehensive assessment of the diversification and genomic evolution in the group.

Chapter one is a multi-locus phylogeny upon which an evaluation of the evolution of outcrossing and a new intra-generic classification scheme are founded. Chapter two uses these same loci to investigate rates of diversification and their association with morphological features to investigate the potential impact of key morphological or life-history innovations in the genus. Chapter three is a description of a new species discovered during the course of this work, along with multivariate investigations that distinguish it from similar species. Lastly, chapter four is a description of the first plastomes sequenced from this family of plants and hypotheses as to how they have come to be found in their present states.
Dedication

To those who have created new nodes in the tokogeny that has ultimately produced me.

From Sir Thomas Wyatt to Nan and Pop, you have all inspired me in some way.
Acknowledgments

My decision to change my major to the Biological Sciences from Middle Childhood Education to the Biological Sciences, under the direction of Carl Chuey, during my sophomore year at Youngstown State University has proven to be the most impactful choice of my life to date. If I would not have changed course, I absolutely would not be writing this document, be married to an amazing women born a quarter-turn of the world away (the beautiful Ieva Roznere), met inspiring and eccentric people from all walks of life or have completed an absolutely life- and mind-altering intellectual apprenticeship with the most kind, intelligent people that one could ever dream of meeting; I have come to view them as my second family.

My nuclear family is ultimately responsible for this dissertation, since my station in life and sphere of interests are due to their open-mindedness, nurturing, and their own selfish desire to satiate my seemingly unending questioning of the natural world. My first recollection of the Theory of Evolution is not reading Darwin’s “abstract” while in college, but rather it is reading an issue of Zoobooks magazine entitled “Life Begins” with my mother; I know of no other student so fortunate. My extended family has gone to great lengths to nurture my curiosity of the mechanisms of the natural world, and without their open-mindedness I would not only be without a terminal degree, but also be far less questioning and aware of the world around me.

iv
It is difficult to convey how appreciative and fortunate I am and have been to have John Freudenstein as my mentor. Upon applying to study under him I had absolutely no knowledge of phylogenetics, but he was willing to take the risk of accepting me -- this risk has come to shape my entire future. It is absolutely impossible to repay him equally within my lifetime, but I hope to be able to bear some academic grandchildren who will make him proud.

Marymegan (Meg) Daly and Hans Klompen have been extremely accommodating and supportive of my work. Meg has been especially encouraging, even in regards to freely expressing my sense of humor, and has become an amazing friend.

Michael Broe, Jeffrey Rose, Ryan Folk, and Craig Barrett have been incredibly supportive lab mates. It is not an exaggeration to say that I think of them as brothers.

Many graduate students of other labs, namely Erica Szeyller, Abigail Reft, Elijah Talamas, Monica Farfan, Nick Skomrock, and Thomas and Cait Evans, have supported me in various ways, and their friendship has greatly buffered the bumps that I have come upon while in Columbus.

My work would not have been possible without generous funding provided by the Doctoral Dissertation Improvement Grant program of the United States National Science Foundation (award #1406732), two William Chambers Coker Fellowships from Highlands Biological Station (2011 & 2013), the Botanical Society of America, the American Society of Plant Taxonomists, Alumni Grants for Graduate Research and Scholarship from the Ohio State University Graduate School, and a Janice Carson
Beatley Grant from the Ohio State University Herbarium. Larry Kelly, Gary Kauffman, and the USDA National Arboretum helped greatly with securing material.

I have been extremely fortunate to have befriended many undergraduate students who have also helped to keep me both young at heart and empathetic to the struggles and stress of those in their first four years of elective education. It has been extremely enriching to have the many students whom I have met during my two summers at Stone Laboratory (staff and faculty included), and those who have made my time in the herbarium so memorable (especially Brent Macolley, Blair Perry, Kate Rose, Charlotte Royer, Kyle McElroy, and Caroline Delevich), in my life – I hope that I will be able to acknowledge some of you as colleagues at the end of my career. In particular, I am fortunate to count Samantha Primer (especially for her “power pose” in the back of conference halls during my national presentations over the years, for critically reading and providing feedback on what has become Chapter 2 of this dissertation, and for accompanying Ieva and me on an amazing backpacking trip to the Northern Rockies just prior this incredibly stressful semester), Dylan Sedmak (for his inspiring determination and persistence), Sebastián Mejía (for farmer market runs), Marissa Lauer (for so many de-stressing, espresso-fueled lunches), and Olivia Cleek (a clone of a younger me) as friends.

Lastly, I acknowledge the support and love of Paul Larson, my academic life partner. You have been there when I lay dying, and I trust when this happens ultimately you will again cuddle me. May our bromance live on in fables.
Vita

2003.................................................................Lordstown Local High School

2008.................................................................B.S. Biology, Youngstown State University

2008 to present ..............................................Graduate Teaching/Research Associate,
Department of Evolution, Ecology and
Organismal Biology, The Ohio State
University

Publications

2015 Sinn, B. T., L. M. Kelly, and J. V. Freudenstein. Phylogenetic relationships in
Asarum: partitioning-dependent topologies and a revised classification. American

2015 Sinn, B. T., L. M. Kelly, and J. V. Freudenstein. Putative floral brood-site
mimicry, loss of autonomous selfing, and reduced vegetative growth are
significantly correlated with increased diversification in Asarum
(Aristolochiaceae). Molecular Phylogenetics and Evolution 89: 194-204.

2015 B. T. Sinn. Asarum chueyi (Aristolochiaceae), a new species from the foothills of
the Blue Ridge Mountains of Tennessee and Virginia, USA. Phytotaxa 224(1):
85-95.

2014 Renne, I. J., B. T. Sinn, G. W. Shook, D. M. Sedlacko, J. R. Dull, D. Villarreal,
and J. L. Hierro. Eavesdropping in plants: delayed germination via biochemical

Fields of Study

Major Field: Evolution, Ecology and Organismal Biology
### Table of Contents

Abstract ........................................................................................................................................ ii

Dedication ..................................................................................................................................... iii

Acknowledgments ........................................................................................................................ iv

Vita ............................................................................................................................................... vii

Publications .................................................................................................................................. vii

Fields of Study .............................................................................................................................. vii

List of Figures ............................................................................................................................... x

List of Tables .................................................................................................................................. xvii

Chapter 1: Phylogenetic relationships in *Asarum* (Aristolochiaceae): effect of data partitioning and a revised classification ................................................................................................................................. 1

Chapter 2. Putative floral brood-site mimicry, loss of autonomous selfing, and reduced vegetative growth are significantly correlated with increased diversification in *Asarum* (Aristolochiaceae) ................................................................................................................................. 48

Chapter 3. *Asarum chueyi* (Aristolochiaceae), a new species from the foothills of the Blue Ridge Mountains of Tennessee and Virginia, USA. ................................................................. 77

Chapter 4: The first reported duplication of the Small Single Copy region in angiosperms: rearrangement, tRNA loss, IR boundary expansion, extreme

viii
regional AT-rich expansion, and episodic selection-regime shifts in *Asarum*
plastomes............................................................................................................. 101

Literature Cited ................................................................................................. 158

Appendix A. Herbarium and GenBank voucher information for material used in
Chapter 1............................................................................................................ 187

Appendix B. Herbarium and GenBank voucher information for material used in
Chapter 2. Accessions denoted with an asterisk were used in our RAxML
analysis; all accessions were used for BCI........................................................ 194

Appendix C. A new combination for *Asarum sorriei*................................. 221

Appendix D. Primers designed for scaffolding of *Asarum* plastome contigs........ 222
List of Figures

Figure 1. Phylogenetic trees collapsed to showcase sectional relationships inferred through the analysis of combined matrices by analysis type. (A) Bayesian, (B) Maximum likelihood, and (C) combined parsimony analysis. A and B are PartitionFinder-optimized. Note the topological congruence between B and C. 14

Figure 2. ML global node support (A) decreased in our PartitionFinder-optimized analysis but increased in our marker-partitioned analysis of plastid data, relative to analysis of singly-partitioned matrices, while global support of nodes resulting from BI only decreased with partitioning of the same data (B). Global support is the sum of node support values for all recovered nodes. Standard error bars are depicted. ................................................................................................................ 22

Figure 3. Neither ML (A) nor BI (B) global node support responded strongly to increased partitioning of our combined matrices, with the exception of our PartitionFinder matrix where global support decreased under both analysis methods. Global support is the sum of node support values for all recovered nodes. Standard error bars are depicted. ............................................................................................................. 23

Figure 4. PartitionFinder-optimized ML cladogram of combined plastid and nuclear data with proportional branch lengths annotated with our classification of Asarum. An ML phylogram for the same dataset is shown in the upper left corner. Maximum likelihood bootstrap, MP (combined plastid and nuclear), and BI of the same
matrix are plotted on nodes, respectively. All BI posterior probabilities were rounded to the nearest 100th place, though values of 0.99 were unchanged........ 27

Figure 5. The presence of autonomous self-pollination (square symbol) and the absence of autonomous self-pollination (star symbol) plotted on collapsed representations of the PartitionFinder-optimized ML cladogram of combined plastid and nuclear data (A), non- and genome-partitioned ML and BI trees (B), and the ML and MP cladograms inferred using only rDNA data (C). Note that *Saruma henryi* is marked with a question mark to underscore the lack of floral comparability between it and *Asarum* species. ................................................................. 40

Figure 6. The likelihood of ML (A) and posterior probability of BI (B) trees improved in our PartitionFinder and marker-partitioned analyses. Akaike Information Criterion values decreased with increased matrix partitioning under ML (A), but BIC values were only impacted by marker partitioning, and then only negatively. Partition number of each matrix analyzed is listed on the x axis. Standard error bars are depicted. ................................................................. 43

Figure 7. ML (left) and BCI (right) trees used in this study; increased diversification is significantly correlated with decreased vegetative growth and characters of putative floral brood-site mimicry. Character states tested here are shown to the left of each taxon (legend in upper left). Change in taxon position in our BCI topology relative to position in the ML tree is denoted by lines. Clades are colored according to clade: Geotaenium = yellow, Euasarum = blue, Asiasarum = purple, Braun = red, Longistylis = black, and Hexastylis = green. Values above branches
represent ML bootstrap and posterior probabilities greater than 50 and 0.95, respectively. .............................................................. 56

Figure 8. Photographs of the flowers of *Asarum canadense* (Euasarum clade, B – D), *A. delavayi* (E – G), *A. arifolium* var. *arifolium* (H – J), and *A. shuttleworthii* var. *shuttleworthii* (K – M) of the Heterotropa clade showing the floral orifice (B, E, H, K), the floral interior (C, F, I, L), and the calyx tube (D, G, J, M) .............. 66

Figure 9. Example of an *A. chueyi* (1195) calyx tube silhouette, with the sepal lobes cropped, used for Elliptical Fourier Analysis. The proximal and distal portions of the flower are to the left and right, respectively. ....................................................... 81

Figure 10. Ordination represented by coordinates 1 and 2 resulting from Nonmetric Multidimensional Scaling of Elliptical Fourier shape descriptors of calyx tube shape of *A. contractum* (+), *A. chueyi* (circles), *A. heterophyllum* (triangles) and *A. minus* (squares). The biplot depicts the impact of stamen filament morphology on data point spread. ........................................................................ 82

Figure 11. Illustrations of *A. chueyi* showing A) a whole plant (prepared from the holotype; scale represents 6.5 cm), and B) a flower with a single sepal removed to reveal the reproductive column, C) a dissected sepal, and D) an entire flower (each prepared from field-preserved flowers from Tennessee and Virginia; scale represents 2.3 cm). Illustrations prepared by Ryan A. Folk. ...................... 83

Figure 12. UPGMA phenogram of stamen filament characters and EFA shape descriptors of calyx tube shape of the flowers of *A. contractum* (red), *A. chueyi* (black), *A. heterophyllum* (green), and *A. minus* (pink). ................................................................. 85
Figure 13. Photographs of *A. chueyi* flowers from Tennessee (A-C) and Virginia (D and E) populations. Typical flower with one sepal removed (A, D), entire flower (B, E) and floral orifice (C). Note that the thecae have not dehisced in A. Scale is not maintained......................................................................................................................... 88

Figure 14. *Saruma henryi* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome ........................................... 112

Figure 15. *Asarum minus* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome. ................................. 113

Figure 16. *Asarum sieboldii* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome. ................................. 114

Figure 17. *Asarum epigynum* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome. ................................. 115

Figure 18. *Asarum canadense* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are
annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome. ...................... 116

Figure 19. *Asarum megacalyx* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome. ...................... 117

Figure 20. *Asarum delavayi* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome. ........................................ 118

Figure 21. Gains (black bars) and losses (white bars) associated with Inverted Repeat-Large Single Copy (IR-LSC) and Inverted Repeat-Small Single Copy (IR-SSC) expansions plotted on a Maximum Likelihood cladogram inferred through the analysis of nuclear rDNA; bootstrap values are plotted along branches. Note the contrast between the rapid expansion and subsequent erosion of the IR-LSC boundary and the more gradual, and rarely lost, duplication of SSC through shifting of the IR-SSC boundary........................................ 123

Figure 22. Gains (black bars) and losses (white bars) associated with Inverted Repeat-Large Single Copy (IR-LSC) and Inverted Repeat-Small Single Copy (IR-SSC) expansions plotted on a Maximum Likelihood cladogram inferred through the analysis of our plastome and combined datasets; bootstrap values for each are plotted along branches, respectively. Note the contrast between the rapid
expansion and subsequent erosion of the IR-LSC boundary and the more gradual, and rarely lost, duplication of SSC through shifting of the IR-SSC boundary...

Figure 23. Select genetic spacer regions characterized alongside Maximum Likelihood phylogenies which have been inferred from the analysis of nuclear rDNA. The magnitudes of AT-biased expansion (A-C), GC-biased compression (D-E), and GC-biased enrichment (F & G) appear to increase in concert with the derivation of lineages. We propose that AT-biased expansion is evidence of either the relaxation of genomic proofreading or a reduction in proofreading fidelity, while both GC-biased compression and enrichment are the result of overall genome GC content balancing in response to these AT-rich expansions.

Figure 24. Alignment of the IR_B boundary of magnoliid plastomes; note the marked expansion of IR_B post loss of trnH-GUG, and subsequent contraction that coincides with the duplication of the SSC. Black portions of sequence are indicative of a lack of comparability between aligned portions.

Figure 25. Alignment of the 5’ flanking region of trnE-UUC through the 5’ flank of psbM in 11 magnoliid species representing four orders. The 5’ flank of trnE-UUC in Asarum species is preceded upstream by rpl16 while this same gene is preceded upstream by trnT-GGU in all other published magnoliid plastomes. Note the contrast between the level of conservation of the tRNA genes and their spacer regions with the spacer regions outside of the trnE-UUC – trnD-GUC syntenic block; this may suggest that these genes migrated as single gene cluster prior to
the psbM-matK inversion, which is one possible reconstruction of this inversion.

Figure 26. Possible structures based on stereotypical plastome circularization suggested by assemblies of Asaroideae plastomes.
List of Tables

Table 1. Primers used for PCR amplification of plastid and nuclear DNA. ....................... 6

Table 2. Table showing our eight matrices, the methods used to analyze each, and key to topologies (Appendix A, S1-S19). Our total concatenated matrix, from which the matrices below are derived, can be found at TreeBASE (#16869). ........................ 8

Table 3. Table of color-coded support measures for section-level relationships by analysis method and matrix partitioning scheme. Green = BI 0.95-1.0, ML 75-100, P 75-100; Orange = BI 0.85-0.94, ML 50-74, P 50-74; Red = BI < 0.75, ML < 50, MP < 50; Blank = not recovered. All BI support values were rounded to the nearest 100th place, though values of 0.99 were not rounded to 1.0. Cells marked with a Δ denote a change in recovered clades due only to partitioning scheme difference, due either to clade loss or gain, or to support values dropping below commonly recognized cutoffs. Partition schemes suggested by PartitionFinder are denoted with an asterisk. .............................................................. 19

Table 4. PartitionFinder and marker partitioning scheme matrices better fit models of evolution and produced trees of greater likelihood in ML analyses. PartitionFinder partitioned and marker partitioned matrices produced Bayesian trees of greater likelihood than those of less partitioned datasets. PartitionFinder (Likelihood, AIC, BIC) and RAxML (ML tree likelihood) scores for combined matrices partitioned singly, by genomic compartment, by PartitionFinder best scheme, and
by marker are shown. The best scores for each criterion are denoted with an asterisk. .............................................................. 44

Table 5. BiSSE model parameters, likelihoods, and likelihood ratio test results of constrained and unconstrained models of character state distributions tested on our *BEAST BCI species tree. ................................................................. 59

Table 6. BiSSE model parameters, likelihoods, and likelihood ratio test results of constrained and unconstrained models of character state distributions tested on our ML phylogram. .................................................................................. 62

Table 7. Accessions analyzed by EFA, NMDS, and UPGMA. ............................... 80

Table 8. Key to models of $\omega$ constraint tested in maximum likelihood analysis of selection regime. ........................................................................................................ 109

Table 9. Comparison of length and percent GC, gene, and non-coding sequence for each plastome sequenced. ........................................................................................................ 111

Table 10. Results of maximum likelihood tests of selection regime in PAML 4 using the branch selection and the F61 codon models on the nuclear rDNA cistron-based phylogram. Significance of parameter change on model improvement was determined via likelihood ratio tests. Constraint schemes are described in Table 8. ........................................................................................................................................ 132

Table 11. Results of maximum likelihood tests of selection regime in PAML 4 using the branch selection and the F61 codon models on the combined data phylogram. Significance of parameter change on model improvement was determined via likelihood ratio tests. Constraint schemes are described in Table 8. ................. 134
Table 12. Total number of perfect repeats identified using Phobos in plastomes from across the angiosperm tree of life compared with those newly described here. 140
Chapter 1: Phylogenetic relationships in *Asarum* (Aristolochiaceae): effect of data partitioning and a revised classification.

**Introduction**

*Asarum* L. is of north temperate distribution with approximately 100 species described (Cheng and Yang, 1983; Whittemore and Gaddy, 1997; Whittemore et al., 1997; Kelly, 2001; Shumei et al., 2003; Sugawara, 2006; Shi et al., 2008; Lu and Wang, 2009), the minority of which autonomously self-pollinate via exaggerative stamen movement and a lack of spatial separation between the thecae and the stigmas (non-herkogamy). Work by Kelly (1997; 1998) supported herkogamous, non-autonomously self-pollinating clades as derived relative to those which autonomously self; this is of particular interest to those investigating plant reproductive biology due to the paucity of other such examples elsewhere in the angiosperms (Takebayashi and Morrell, 2001). The center of diversity of the group is in Asia, where approximately 65 species are distributed across China, Taiwan, Japan, Vietnam, and the Korean Peninsula. The remaining species are distributed in North America (15) and Europe (1). The majority of Asian species outside of mainland China are poorly known. This group of plants has received a modest amount of attention due to their taxonomic difficulty (Blomquist, 1957; Gaddy, 1987; Yamaji et al., 2006, 2007; Weakley, 2012), broad disjunctions between morphologically
similar species (Xiang and Soltis, 2001; Xiang et al., 2004), and medicinal and horticultural value (Han et al., 2008; Nivot et al., 2008; Quang et al., 2012; Kopyt’ko et al., 2013; Ku et al., 2013a, b; Li et al., 2013; Li et al., 2013; Perumalsamy et al., 2013; Williamson et al., 2013). Several genera have been segregated from *Asarum* (i.e., *Asiasarum* F.Maek., *Geotaenium* F.Maek., *Heterotropa* Morren & Decne., *Hexastylis* Raf., *Japonasarum* Nakai), but in recent treatments these have been treated as infrageneric taxa (Shumei et al., 2003; Sugawara, 2006). Minimal taxonomic sampling and regionally focused floristic treatments have contributed to taxonomic disagreement.

Kelly (1997, 1998) tested the monophyly of *Asarum sensu lato* using both morphology and the Internal Transcribed Spacer (ITS) region of rDNA and found strong congruence between signal in these data types. His was the first work to provide phylogenetic evidence that *Asarum s.l.* is monophyletic and that the Appalachian species (section *Hexastylis*) are more closely related to Asian species than they are to other North American members of *Asarum*. The analysis of ITS sequence data (Kelly, 1998) resulted in increased resolution within section *Hexastylis* in both the ITS-only and combined trees, relative to morphological-only analyses (Kelly, 1997). However, section *Hexastylis* was recovered as paraphyletic in all trees inferred from matrices that included molecular data, with Blomquist’s (1957) informal *Arifolia* + *Speciosa* groups sister to his *Virginica* + the remainder of subgenus *Heterotropa*. Niedenberger (2010) conducted a molecular phylogenetic study that used three plastid regions (*matK, rpl32-trnL, and trnQ-rps16*).

Bayesian (BI; Rannala and Yang, 1996), Maximum Likelihood (ML; Felsenstein, 1981), and Maximum Parsimony (MP; Camin and Sokal, 1965) trees contained varying levels of
resolution. The reported phylogeny is not congruent with that recovered by Kelly (1997, 1998) in that section *Hexastylis* was recovered as monophyletic and section *Heterotropa* as paraphyletic.

Despite the horticultural value of *Asarum* (many species are being increasingly grown for their ornate flowers), as well as mounting evidence of the medicinal utility of some species (Kopyt’ko et al., 2013), infrageneric relationships and character evolution remain unclear. By analyzing the most inclusive sampling to date from *Asarum s.l.*, as well as incorporating rDNA and plastid data, we hope to clarify infrageneric relationships that will allow investigators to interpret biogeographical, vegetative, and floral character transitions in this understudied group of plants. Specific questions related to the phylogeny of *Asarum* addressed here include: Do *Asarum s.l.* segregate genera form monophyletic groups? What are the relationships between and within these groups?

Phylogenetic relationships within *Asarum* hold important implications for floral evolution in the group and elsewhere. Floral characters that have been implicated as attracting fungus gnats through floral mimicry of basidiomycete sporocarps (Vogel, 1978; Lu, 1982; Sugawara, 1988; Leins and Erbar, 2010) have been identified in more than half of the species, and contrary to many angiosperm lineages (Takebayshi and Morrell, 2001), work to date suggests that the loss of autonomous self-pollination is a derived condition in the genus (Kelly, 1997; 1998). Here we ask: Are floral traits such as autonomous self-pollination and those involved in putative fungal mimicry (i.e. calyx tube sculpturing and floral orifice constriction) synapomorphies of clades within *Asarum*?
In order to better generate robust phylogenetic hypotheses on which to base *Asarum* classification and from which to identify putative synapomorphies and to guide future investigations of character-associated evolution, we have sequenced several coding and intergenic spacer regions from both the plastid and nuclear genomes. While much has been discussed regarding the impact of dataset partitioning on model fit and tree likelihood, comparatively little analysis of the effects of partitioning scheme on resulting topologies and the classification schemes based on them has been presented (Fenn et al., 2008; Xi et al., 2012). Partitioning of a dataset better accounts for marker heterogeneity by estimating model parameters separately for each genetic region analyzed. This is crucial for parametric approaches such as ML and BI, where evolutionary model rates and frequencies must be estimated accurately for models to be appropriately parameterized (McGuire et al., 2007); this can be especially difficult when multiple loci with different histories or rates of evolution are used in a single analysis. Here we evaluate the effects of common a priori and information theory-informed partitioning schemes of our matrices on topology. Specifically, we evaluate the extent to which matrix partitioning affects our results.

**Materials and Methods**

**DNA Extraction and Sequence Generation**

Species from all *Asarum* segregate genera were included in our analysis, with care taken to include the designated type species of each. Fresh leaf material was field
collected and stored at -80°C. Total DNA extractions were conducted using the CTAB method (Doyle and Doyle, 1987) modified to use 67 mg of leaf tissue, which was disrupted using a bead mill and stored at -20°C.

Eight DNA regions were amplified: \textit{rpoB-trnC}^{GCA}, \textit{rps16-trnK}, \textit{trnL}^{UAA} \text{exon} – \textit{trnF}^{GAA}; \textit{trnT}^{UGU} – \textit{trnL}^{UAA}; \textit{trnL}^{UAA} – \textit{trnL}^{UAA} \text{exon}, \textit{trnS}^{UGA}-\textit{trnFM}^{CAU}, \text{and} \textit{ycf1} 3850-5310 from the plastome and ITS1 – partial 26S from the nuclear genome. ITS1 and 2 sequences were obtained from GenBank for \textit{A. splendens} (F.Maek.) C.Y.Cheng & C.S.Yang (isolate #58430). All taxa, sequences amplified from each, and Genbank accession numbers are given in Appendix A; primers used are listed in Table 1. PCR amplifications for all plastid markers were carried out using the \textit{rpl16} program of Shaw et al. (2005). The PCR program for all rDNA sequences was as follows: initial template denaturation at 95°C for three minutes followed by 34 cycles of denaturation at 95°C for 30 seconds, primer annealing at 55°C for one minute, and primer extension at 72°C for one minute. The program was ended with an additional five minutes of primer elongation at 72°C and held at 4°C.
Table 1. Primers used for PCR amplification of plastid and nuclear DNA.

<table>
<thead>
<tr>
<th>Region</th>
<th>Primer Name</th>
<th>Primer Sequence (5′-3′)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITS1-26S</td>
<td>268rev</td>
<td>GCATTCCCAAACAACCCGAC</td>
<td>Kuzoff et al. 1998</td>
</tr>
<tr>
<td></td>
<td>ITS1</td>
<td>TCCGTAGGTGAACCTGCGC</td>
<td>White et al. 1990</td>
</tr>
<tr>
<td>rpoB-trnC&lt;sup&gt;GCA&lt;/sup&gt;</td>
<td>rpoB</td>
<td>CKACAAAYCCYTCRAATTG</td>
<td>Shaw et al. 2005</td>
</tr>
<tr>
<td></td>
<td>trnC&lt;sup&gt;GCA&lt;/sup&gt;R</td>
<td>CACCCRGATTYGAACTGCGG</td>
<td>Shaw et al. 2005</td>
</tr>
<tr>
<td>trnS&lt;sup&gt;UGA&lt;/sup&gt;-trnFM&lt;sup&gt;CAU&lt;/sup&gt;</td>
<td>trnS&lt;sup&gt;UGA&lt;/sup&gt;</td>
<td>GAGAGAGAGGAGACGACCC</td>
<td>Shaw et al. 2005</td>
</tr>
<tr>
<td></td>
<td>trnFM&lt;sup&gt;CAU&lt;/sup&gt;</td>
<td>CATAACCTTGAGGTACGGG</td>
<td>Shaw et al. 2005</td>
</tr>
<tr>
<td>3′rps16-5′trnK</td>
<td>rps16x2f2</td>
<td>AAAATGGGTTTTATGATCC</td>
<td>Shaw et al. 2007</td>
</tr>
<tr>
<td></td>
<td>trnK&lt;sup&gt;UUU&lt;/sup&gt;x1</td>
<td>TTAAAAGCCGGATGACTCTACC</td>
<td>Shaw et al. 2007</td>
</tr>
<tr>
<td>trnT&lt;sup&gt;UGU&lt;/sup&gt;-trnL&lt;sup&gt;UAA&lt;/sup&gt;</td>
<td>tabA</td>
<td>CATTACAATTCGGATGCTCT</td>
<td>Taberlet et al. 1991</td>
</tr>
<tr>
<td></td>
<td>tabB</td>
<td>TCTACCGATTTCGGACTATCC</td>
<td>Taberlet et al. 1991</td>
</tr>
<tr>
<td>trnL&lt;sup&gt;UAA&lt;/sup&gt;-trnL&lt;sup&gt;UAA&lt;/sup&gt;</td>
<td>tabC</td>
<td>CGAAATCGGATGACTACG</td>
<td>Taberlet et al. 1991</td>
</tr>
<tr>
<td>exon</td>
<td>tabD</td>
<td>GGGGATAGGAGGACTGAC</td>
<td>Taberlet et al. 1991</td>
</tr>
<tr>
<td>trnL&lt;sup&gt;UAA&lt;/sup&gt;exon –</td>
<td>tabE</td>
<td>GGTTCAGTCCCTCTATCC</td>
<td>Taberlet &lt;i&gt;et al.&lt;/i&gt; 1991</td>
</tr>
<tr>
<td>trnF&lt;sup&gt;GAA&lt;/sup&gt;</td>
<td>tabF</td>
<td>ATTTGAAGTGGACGAG</td>
<td>Taberlet &lt;i&gt;et al.&lt;/i&gt; 1991</td>
</tr>
<tr>
<td>ycf1</td>
<td>MagnoliidF3850yc</td>
<td>AGATGGTCTAGATTAGTACGGGCAAAA</td>
<td>Neubig &amp; Abbott 2010</td>
</tr>
<tr>
<td></td>
<td>fl</td>
<td>TGGCG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Magnoliid5310Ryc</td>
<td>TAGGAGATGGAAATTTTCAG</td>
<td>Neubig &amp; Abbott 2010</td>
</tr>
</tbody>
</table>
Amplicons were sequenced by Beckman-Coulter Genomics (Danvers, MA, USA) and inspected and assembled in Geneious versions 5.3-6.1 (Drummond et al., 2011). Sequence assemblies were trimmed using an $\alpha$ of 0.05, and all were reference guided using the consensus read implied by an assembly constructed of the highest quality read from each locus. Sequences of poor quality were proofread and edited by hand using context provided by the reference sequence. Alignments were generated for each locus using MUSCLE version 3.8 (Edgar, 2004) through Geneious 6.1 with default parameters with up to 100 iterations of refinement; alignments were then concatenated.

**Matrix construction and phylogenetic analyses**

Twenty phylogenetic analyses were conducted (Appendices S1-S19; legend in Table 2), which varied in both matrix partitioning and constitution (Table 2). Matrices were composed of plastid and rDNA, only plastid, or only rDNA sequences.
Table 2. Table showing our eight matrices, the methods used to analyze each, and key to topologies (Appendix A, S1-S19) for the resultant topology. Our total concatenated matrix, from which the matrices below are derived, can be found at TreeBASE (#16869).

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Partitions</th>
<th>Model Selection</th>
<th>Source</th>
<th>Length (bp)</th>
<th>Phylogenetic Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>jModelTest2</td>
<td>Plastid + Nuclear</td>
<td>9313</td>
<td>BEAST S1</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>jModelTest2</td>
<td>Plastid</td>
<td>8435</td>
<td>S4</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>jModelTest2</td>
<td>Plastid + Nuclear</td>
<td>9313</td>
<td>S6</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>jModelTest2</td>
<td>Nuclear</td>
<td>878</td>
<td>S8</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>jModelTest2</td>
<td>Plastid</td>
<td>8435</td>
<td>S11</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>PartitionFinder</td>
<td>Plastid + Nuclear</td>
<td>9313</td>
<td>S14</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>PartitionFinder</td>
<td>Plastid</td>
<td>8438</td>
<td>S16</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>jModelTest2</td>
<td>Plastid + Nuclear</td>
<td>9313</td>
<td>S18</td>
</tr>
</tbody>
</table>
Model selection and partitioning of matrices

jModelTest 2 (Darriba et al., 2012) was used for nucleotide substitution model selection of models that can be implemented in RAxML, using the Akaike Information Criterion (AIC; Akaike, 1974) to determine the model of best fit, and the Generalized Time Reversible Model (GTR; Tavaré, 1986) was preferred for all loci, whether tested singly or concatenated. GTR + I was used in all model-driven analyses. The same model was used in all parametric analyses to allow for maximum comparability between RAxML and BEAST.

In order to more fully explore the phylogenetic signal in our data, our matrices were assembled four ways: unpartitioned, partitioned by genomic compartment, marker-partitioned, and a PartitionFinder, RAxML-optimized, scheme. PartitionFinder v1.1.0 (Lanfear et al., 2012) was used to select AIC- and BIC-optimal partitioning of markers. All PartitionFinder search types were used to evaluate 4140 possible schemes (the approximate Bell number of 8). The optimal partitioning and model scheme was identified as a four-partition matrix (partition1: rpoB-trnC\textsuperscript{GCA}, rps16-trnK, trnL\textsuperscript{UAA}exon – trnF\textsuperscript{GAA}; partition 2: trnT\textsuperscript{UGU} – trnL\textsuperscript{UAA}; partition 3: trnL\textsuperscript{UAA}– trnL\textsuperscript{UAA} exon, trnS\textsuperscript{UGA} – trnFM\textsuperscript{CAU}, ycf1 3850-5310; partition 4: rDNA). Although the plastome is a single chromosome, plastid markers were partitioned in order to fully explore their relative impacts on phylogenetic reconstruction and to improve model-fit in light of regional substitution-rate heterogeneity.
Bayesian Inference – Bayesian inference was conducted using BEAST version 1.7.5 (Drummond et al., 2012) and the BEAGLE API was implemented for multicore processing (Ayres et al., 2011). BEAST XML files were constructed using BEAUti (Drummond et al., 2012). All partitions had unlinked substitution models and clocks. A Yule process (Yule, 1925) was used for branching and tree prior estimation (Gernhard, 2008) using a uniform prior initially set at 3700 and bounded by 0 to 1E100. Lognormal Uncorrelated clock models were run for all analyses, with rates estimated from a lognormal prior distribution with an initial value and mean of 1. Default priors were used for all other parameters. The MCMC chain was run for 50,000,000 generations or until a minimum effective sample size (ESS) of 200 was achieved for all estimated values. A 10% burn-in was used for each analysis. Parameters were logged every 1000 generations. Tracer version 1.7 (Rambaut and Drummond, 2008) was used to monitor parameter estimations as well as ESS during analysis. Maximum clade credibility trees were constructed using Tree Annotator (Drummond et al., 2012), with a burn-in of 25% of the total trees generated and node annotation posterior probability cutoff set to > 0.95. Figtree version 1.4 (Rambaut, 2012) and TreeGraph2 version 2.0.5 (Stöver and Müller, 2010) were used to visualize and annotate trees.

Maximum likelihood analyses – All Maximum Likelihood analyses were conducted using the high performance computing Pthreads version of RAxML version 7.4.2 (Stamatakis, 2006; Ott et al., 2007) through RAxML GUI version 1.3 (Silvestro, 2012). We used the GTRGAMMAI model which includes Γ rate heterogeneity, estimates the proportion of
invariant sites, and uses six discrete rate categories in the Q matrix. Tree searches were carried out using the ML + rapid bootstrap (Stamatakis et al., 2008) option with 1000 bootstrap replicates.

*Maximum parsimony analyses* – TNT version 1.1 (Goloboff et al., 2008) was used for all parsimony analyses. Twenty replicate tree searches were conducted using the driven search method set to 99. Ratchet (Nixon, 1999), Sectorial Search, Drift, and Tree Fusing (Goloboff, 1999) were all used and left at default values. Ratchet up-weighting and down-weighting probabilities were perturbed as a method to break from localized optima. Jackknife resampling (Farris et al., 1996), using the default parameters of the traditional search (Wagner starting tree with ten rounds of branch swapping, tree-bisection-reconnection with ten trees saved per replicate, and replacing existing trees) with 1000 replicates was used to generate clade support values. Strict consensus trees were calculated.

**Results**

*rDNA sequences*

ITS1 through the first 229 base pairs of the 5’ end of 26S rDNA was successfully amplified from 49 of 58 accessions. The rDNA aligned matrix totaled 878 base pairs in length (TreeBASE submission #16869), with missing data, including alignment gaps, comprising 17.9% of the matrix. ITS1 and 2 sequence data alone were used for *A.*
harrwagii S.Watson, *A. heterophyllum* Ashe (1058), *A. kumageanum* Masam., and *A. speciosum*. Aligned ITS1 and 2 totaled 645 base pairs in length. Average pairwise identity between sequences was 91.8%, with 59.5% of sites identical.

**Plastid sequences**

Seven plastid regions were amplified yielding a matrix whose aligned length was 8435 bp. Alignment of plastid sequences resulted in many more gaps than alignment of rDNA; missing data and alignment gaps accounted for 37.6% of the matrix. Plastid markers differed greatly in variability; 89.4% of sites in *ycf1* 3850-5310 were identical, compared to only 24.9% of sites in *trnT^UGU–trnL^UAA*. The most variable region recovered, *trnT^UGU–trnL^UAA*, expanded to 2112 bases post-alignment from just 912. Our plastid regions averaged 92.7% pairwise identity and 61.6% identical sites, making this genomic compartment only slightly less variable than the nuclear partition.

**Phylogenetic relationships and topological incongruence in Asarum s.l.**

Infrageneric relationships in *Asarum* were generally found to consist of two large monophyletic groups corresponding to the subgenera *Heterotropa* and *Asarum* sensu Kelly (1997, 1998). The monophyly of *Asarum* subgenus *Asarum sensu* Kelly (1998) was recovered by ML analyses of both PartitionFinder and marker-partitioned combined matrices with support values of 91 and 94, respectively. Combined plastid and nuclear BI and ML analyses of matrices 1 and 8 recovered subgenus *Geotaenium* as sister to section *Asarum*. However, topologies recovered by combined nuclear and chloroplast ML
PartitionFinder-optimized and marker-partitioned analyses were in agreement with our combined MP analysis in the placement of subgenus *Geotaenium* as sister to the remainder of *Asarum sensu lato* (Figure 1). The sister relationship of subgenus *Geotaenium* to section *Asarum* dropped out of ML trees of combined plastid and nuclear data when matrices were divided into greater than two partitions. The monophyly of subgenus *Heterotropa sensu* Kelly (1998) was supported by chloroplast and rDNA data in the majority of analyses; only BI of plastid matrices resulted in the breakdown of monophyly of subgenera *Asarum* and *Heterotropa* due to the recovery of a paraphyletic subgenus *Asarum*. 
Figure 1. Phylogenetic trees collapsed to showcase sectional relationships inferred through the analysis of combined matrices by analysis type. (A) Bayesian, (B) Maximum likelihood, and (C) combined parsimony analysis. A and B are PartitionFinder-optimized. Note the topological congruence between B and C.
Tree topology was largely congruent at the subgeneric level among MP, ML, and BI analyses of identical matrices; however, discordance was strongly evident among trees inferred from plastid- or rDNA-only datasets. Trees resulting from the analysis of our combined nuclear and plastid matrices largely strengthened patterns seen in the plastid matrices, especially in regard to the sister relationships of sections *Hexastylis* and *Longistylis* and between sections *Asiasarum* and *Heterotropa*.

Asarum sensu stricto – We recover a monophyletic group consisting of all North American (NA) species of *Asarum* subgenus *Asarum*. One of the most consistent and highly supported relationships identified here is the position of *A. caulescens* within *Asarum sensu stricto*. This species is supported as sister to Asian and NA *Asarum* excluding subgenus *Geotaenium* and *A. pulchellum + A. himalaicum* in all analyses with the exception of the BI analysis of rDNA, where the relationship is extremely weakly supported (posterior probability = 0.25) as sister to *A. pulchellum*; when this tree is collapsed using a posterior cutoff of 0.95, this taxon falls into a polytomy with subgenus *Geotaenium* and *A. himalaicum* at the base of *Asarum s.s*.

The most contentious and inconsistently supported relationship found throughout our analyses is that of a sister relationship of *Geotaenium* to a clade consisting of *A. pulchellum + A. himalaicum* which are then sister to the remainder as *Asarum sensu lato*. *Asarum pulchellum* and *A. himalaicum* formed a highly supported clade in all analyses of
plastid or combined data, but formed a basal grade with *A. caulescens* Maxim. when rDNA was analyzed alone.

Subgenus *Geotaenium* was a wildcard in the analysis of nearly all matrices, where it was a member of a basal polytomy using matrix 4, and sister to subgenus *Asarum* in matrices 1 and 5. *Geotaenium* as sister to *Asarum s.s.* was recovered, but not supported, in all BI combined analyses (all posterior probabilities < 0.85); this relationship was weakly supported in our single (bootstrap = 62) and genomic compartment-partitioned (bootstrap = 48) ML analyses. However, subgenus *Geotaenium* was weakly recovered as sister to the remainder of *Asarum s.l.* by ML analysis of our PartitionFinder optimized (bootstrap = 26), marker partitioned (bootstrap = 30) matrices, as well as our combined MP analysis (jackknife < 50). Maximum Likelihood and MP analyses of rDNA also recovered this relationship (bootstrap = 40; jackknife < 50), although BI placed subgenus *Geotaenium* as sister to section *Asarum* and section *Asiasarum*. However, plastid sequences analyzed alone failed to identify this relationship under all partitioning schemes and analysis methods, with the placement of subgenus *Geotaenium* as sister to *Asarum s.s.* to the exclusion of *A. pulchellum* and *A. himalaicum*.

**Heterotropa –** Disagreement is strong within subgenus *Heterotropa*, and plastid and rDNA data suggest vastly different relationships within the subgenus, although four major clades were present in all analyses. When the *A. muramatsui* Makino + *A. asaroides* Makino clade, which was found in all trees, was recovered as sister to the clade containing *A. minamitanianum* Hatus., *A. hexalobum* (F.Maek.) F.Maek. var. *perfectum*
F. Maek. was recovered as sister to both clades. All combined matrices place *A. forbesii* Maxim. as sister to the remainder of section *Heterotropa*, whereas plastid and rDNA matrices recover this species either as sister to *A. minamitanianum* or to *A. hatsushimae* F. Maek. ex Hatus. & Yamahata and *A. gelasinum* Hatus. & Yamahata, respectively.

The Chinese subgenus *Heterotropa* species *A. delavayi* Franch., *A. splendens*, and *A. maximum* Hemsl. are recovered as a monophyletic group, though the identity of its sister clade changes dramatically depending on the matrix analyzed. This clade is recovered as sister to section *Hexastylis* in all analyses of combined and plastid matrices, though it is supported as sister to the remainder of section *Heterotropa* species in all analyses of the rDNA matrix.

The Asian *Asarum s.s.* species *Asarum pulchellum* and *A. himalaicum* are recovered, but not supported, as sister to subgenus *Heterotropa* in trees resulting from BI of plastid matrices (posterior probabilities of 0.67 to 0.89).

**Asiasarum** – The vast majority of methods and partitioning schemes recover section *Asiasarum* as sister to section *Heterotropa*, excluding *A. maximum*, *A. delavayi*, and *A. splendens*; section *Asiasarum* is sister to the remainder of subgenus *Heterotropa sensu* Kelly in rDNA trees.

**Hexastylis** – Section *Hexastylis* is recovered as a monophyletic group when analyzing plastid and combined matrices, but paraphyletic with respect to the remainder of subgenus *Heterotropa* when rDNA data are interpreted in a Parsimony or Maximum
Likelihood framework. Our BEAST analysis of rDNA lends moderate support (posterior probability = 0.92) to a monophyletic section *Hexastylis*, although the remainder of the tree agrees with ML and MP analyses of the same matrix.

_Effects of data partitioning_

Both ML and BI recovered trees that gained or lost support for clades due to partitioning of our data; however, topologies resulting from the analysis of our PartitionFinder-optimized matrices strongly resembled trees resulting from analysis of our marker-partitioned datasets (Table 3).
Table 3. Table of color-coded support measures for section-level relationships by analysis method and matrix partitioning scheme. Green = BI 0.95-1.0, ML 75-100, P 75-100; Orange = BI 0.85-0.94, ML 50-74, P 50-74; Red = BI < 0.75, ML < 50, MP < 50; Blank = not recovered. All BI support values were rounded to the nearest 100th place, though values of 0.99 were not rounded to 1.0. Cells marked with a ∆ denote a change in recovered clades due only to partitioning scheme difference, due either to clade loss or gain, or to support values dropping below commonly recognized cutoffs. Partition schemes suggested by PartitionFinder are denoted with an asterisk.
<table>
<thead>
<tr>
<th>Phylogenetic Method</th>
<th>Data Source</th>
<th>Matrix Name</th>
<th>Partitioning Scheme</th>
<th>Asarum minus A. pulchellum</th>
<th>Asarum + A. pulchellum</th>
<th>Asarum + Geotaenium minus A. himalaicum</th>
<th>Asarum + Geotaenium + A. pulchellum</th>
<th>Asarum + Asiasarum minus A. himalaicum</th>
<th>Asarum + Asiasarum + A. pulchellum</th>
<th>Heterotropa + A. pulchellum</th>
<th>Heterotropa + Asiasarum</th>
<th>Heterotropa + Longistylis</th>
<th>Hexastylis</th>
<th>Hexastylis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parsimony</td>
<td>Nuclear</td>
<td>4</td>
<td>1</td>
<td>86</td>
<td>89</td>
<td></td>
<td></td>
<td>91</td>
<td>53</td>
<td>&gt;50</td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Plastid</td>
<td>5</td>
<td>1</td>
<td>98</td>
<td>53</td>
<td></td>
<td></td>
<td>93</td>
<td>53</td>
<td>&gt;50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>1</td>
<td>1</td>
<td>99</td>
<td>65</td>
<td></td>
<td></td>
<td>96</td>
<td>&gt;50</td>
<td>&gt;50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum Likelihood</td>
<td>Nuclear</td>
<td>4</td>
<td>1</td>
<td>88</td>
<td>91</td>
<td></td>
<td></td>
<td>95</td>
<td>53</td>
<td>&gt;50</td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Plastid</td>
<td>5</td>
<td>1</td>
<td>100</td>
<td>77</td>
<td></td>
<td></td>
<td>99</td>
<td>89</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plastid</td>
<td>7</td>
<td>3*</td>
<td>100</td>
<td>68</td>
<td></td>
<td></td>
<td>99</td>
<td>89</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plastid</td>
<td>2</td>
<td>7</td>
<td>100</td>
<td>68</td>
<td></td>
<td></td>
<td>99</td>
<td>91</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>94</td>
<td>62</td>
<td></td>
<td>99</td>
<td>59</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>8</td>
<td>2</td>
<td>100</td>
<td>90</td>
<td>48 Δ</td>
<td></td>
<td>99</td>
<td>74</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>6</td>
<td>4*</td>
<td>100</td>
<td>91</td>
<td>Δ</td>
<td></td>
<td>99</td>
<td>79</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>3</td>
<td>8</td>
<td>100</td>
<td>94</td>
<td>Δ</td>
<td></td>
<td>99</td>
<td>79</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayesian Inference</td>
<td>Nuclear</td>
<td>4</td>
<td>1</td>
<td>1.0</td>
<td>0.93</td>
<td>0.44</td>
<td>0.99</td>
<td>0.92</td>
<td>0.99</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plastid</td>
<td>5</td>
<td>1</td>
<td>1.0</td>
<td>0.67</td>
<td>1.0</td>
<td>0.67</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plastid</td>
<td>7</td>
<td>3*</td>
<td>0.99</td>
<td>0.89</td>
<td>1.0</td>
<td>0.89</td>
<td>1.0</td>
<td>0.99</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plastid</td>
<td>2</td>
<td>7</td>
<td>1.0</td>
<td>0.54</td>
<td>1.0</td>
<td>0.54</td>
<td>1.0</td>
<td>0.99</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>1</td>
<td>1</td>
<td>1.0</td>
<td>1.0</td>
<td>0.85</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>8</td>
<td>2</td>
<td>0.99</td>
<td>0.92</td>
<td>0.41</td>
<td>0.92</td>
<td>0.99</td>
<td>0.99</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>6</td>
<td>4*</td>
<td>1.0</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>1.0</td>
<td>0.99</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>3</td>
<td>8</td>
<td>1.0</td>
<td>0.99</td>
<td>0.71</td>
<td>0.99</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
With the exception of our PartitionFinder scheme, cumulative support increased as partition number increased in an ML framework but decreased under BI analyses of our chloroplast matrices (Figure 2). Maximum Likelihood analysis of our marker-partitioned plastid matrix resulted in the highest cumulative support, while the greatest cumulative support for a tree inferred using BI of plastid data was recovered through the analysis of our unpartitioned matrix. However, cumulative support response to partitioning did not significantly differ between ML and BI when analyzing combined matrices (Figure 3), although many topological differences were evident.
Figure 2. ML global node support (A) decreased in our PartitionFinder-optimized analysis but increased in our marker-partitioned analysis of plastid data, relative to analysis of singly-partitioned matrices, while global support of nodes resulting from BI only decreased with partitioning of the same data (B). Global support is the sum of node support values for all recovered nodes. Standard error bars are depicted.
Figure 3. Neither ML (A) nor BI (B) global node support responded strongly to increased partitioning of our combined matrices, with the exception of our PartitionFinder matrix where global support decreased under both analysis methods. Global support is the sum of node support values for all recovered nodes. Standard error bars are depicted.
Clade-specific effects on node support due to partitioning scheme were especially evident when using the PartitionFinder-suggested scheme in a Bayesian framework; the placement of wildcard taxa became increasingly unstable, and sometimes unexpectedly highly supported. Nearly all bootstrap values for nodes within subgenus *Heterotropa* decreased with increased partition number. All support values in the *A. yakusimense* clade of subgenus *Heterotropa* decreased or remained unchanged with increased partition number in both ML and BI analyses. Conversely, of the eight nodes in subgenus *Asarum*, only three saw a decrease in the support values of either ML or BI. Posterior probabilities within subgenus *Heterotropa* varied more greatly (differences from 0.1 to 0.69) than those of the ML bootstrap (differences from 1 to 6). Of the support values for 33 nodes in the ML tree produced from matrix 2, 15 values increased, eight decreased, and ten remained unchanged, relative to matrix five. In the BI tree inferred from the same matrix an opposing trend was seen; seven nodes increased, 12 decreased and 14 remained unchanged. The magnitude of change in these values varied between ML and BI trees. Increases in ML bootstrap values ranged from one to 12 and those of BI posterior probabilities ranged from 0.04 to 0.26. Decreases in support values ranged from one to nine for ML bootstrap and 0.02 to 0.45 for BI posterior probabilities. Some of these changes were enough to shift a clade from above to below 50 for ML bootstrap or to a posterior probability of less than 0.95.
Discussion

*Revised taxonomic treatment and reconciliation with previous work*

Infrageneric relationships and species circumscription have long been a problem in *Asarum s.l.*; abundant floral shape, color, size, and ornament differences, as well as cytological and phytochemical variation have led to the recognition of as many as six segregate genera (Rafinesque, 1825; Maekawa 1936, 1953, 1978; Tanaka, 1935; Gregory, 1956; Blomquist, 1957; Ono, 1960; Maekawa and Ono, 1965; Fujita, 1966; Yuasa and Maekawa, 1976; Gaddy, 1987; Whittemore and Gaddy, 1997).

Linnaeus published four species of *Asarum* in his *Species Plantarum* (1753) including *Asarum (Hexastylis) virginicum*. Michaux (1803) followed Linnaeus’ treatment of *Asarum* with the addition of *A. arifolium*. This broad interpretation of *Asarum* survived for another 22 years, but once genera began to be segregated from *Asarum*, an inclusive view of the circumscription of *Asarum* would not be proposed again until 200 years after the publication of *Species Plantarum*. Araki (1937, 1953), Cheng and Yang (1983), and Kelly (1997, 1998, 2001) all recognized *Asarum* in a broad sense, with two subgenera and multiple sections. Contention as to the relationships among *Asarum* segregate genera persisted throughout the 20th century, and molecular phylogenetic investigations show much promise for resolving recalcitrant taxonomic affinities in *Asarum*.

Our trees, though admittedly lacking resolution of some key nodes, do much to resolve relationships within the genus and are strongly congruent with morphology-based
classifications (Araki, 1953; Cheng and Yang, 1983; Kelly, 1997) and a previous rDNA phylogeny (Kelly, 1998). Although our plastid-only, unpartitioned, and twice-partitioned combined plastid and nuclear ML analyses are highly congruent with all of our Bayesian inferred trees, our taxonomic treatment is based on the relationships recovered in our greater-than-twice-partitioned combined ML analyses and our combined MP tree. The improved model fit and increased likelihoods that result from increased partitioning of our combined matrix under an ML approach, independently supported by MP, suggest that if a stable classification is desired, it should be derived from these trees. The taxonomy proposed here is based on our PartitionFinder-optimized combined plastid and nuclear ML and MP trees (Figure 4), and the monophyly of the sections and subgenera indicated here is supported in the majority of our trees. We propose a revised taxonomic treatment (Appendix C) naming new ranks at the levels of subgenus and section that are predicated on the presence of both molecular and morphological characters, and we identify putative morphological synapomorphies for each below.
Figure 4. PartitionFinder-optimized ML cladogram of combined plastid and nuclear data with proportional branch lengths annotated with our classification of *Asarum*. An ML phylogram for the same dataset is shown in the upper left corner. Maximum likelihood bootstrap, MP (combined plastid and nuclear), and BI of the same matrix are plotted on nodes, respectively. All BI posterior probabilities were rounded to the nearest 100th place, though values of 0.99 were unchanged.
Subgenus Asarum section Asarum – Species traditionally placed within the genera *Asarum*, *Japonasarum*, and *Geotaenium* are generally characterized by long internodes, paired or single deciduous, evergreen, or marcescent leaves, and flowers with pilose, acuminate non-connate sepals.

*Japonasarum* Nakai (1936) was circumscribed to segregate *A. caulescens*, whose sepal lobes are reflexed but without elongate sepal extensions. These characteristics are quite plastic and are found to varying degrees in other taxa within section *Asarum* as treated here. Cheng and Yang (1983) treated *A. pulchellum* as a member of *Japonasarum*, though the basis of this decision is unclear, since the sole species named in *Japonasarum* was *A. caulescens* (Nakai, 1936); its placement within *Geotaenium* by Maekawa (1953) was contested by Cheng and Yang (1983). Based on chromosome counts, morphological similarity, and our phylogenies we follow Kelly (1998) and do not place this species in *Geotaenium*.

We recognize species that have deciduous leaves, bear flowers with actinomorphic calyces made up of non-connate sepals that do not form a tube, and have a chromosome number of 2n = 26 as belonging to section *Asarum*.

Subgenus Geotaenium section Geotaenium – Maekawa (1953) named *Geotaenium* for *Asarum* species with connate styles, weakly zygomorphic flowers, and sculpture-less calyx tube interiors. Three species, *A. epigynum*, *A. geophilum* Hemsl., and *A. yunnanense* T.Sugaw., Ogisu & C.Y.Cheng, have been recognized within this genus, all
with chromosome counts of 2n = 12 (Shumei et al., 2003; Sugawara, 2006; Shi et al., 2008).

Our analyses were unable to unanimously resolve the position of *A. epigynum* as sister to the remainder of *Asarum* or to subgenus *Asarum*, and thus based on our PartitionFinder-optimized ML tree and our combined MP tree, we recognize *Geotaenium* at the subgeneric level. Only a single species represents *Geotaenium* in our analyses, and thus we could not evaluate the monophyly of the group.

We recognize species that have deciduous leaves, calyces composed of non-connate sepals that crowd to form a slightly zygomorphic pseudo-tube, and a chromosome number of 2n = 12 as belonging to section *Geotaenium*.

*Subgenus* Heterotropa *section* Asiasarum – Maekawa (1936) circumscribed *Asiasarum* to segregate species with stamen movement, free styles, and a slightly zygomorphic, short calyx tube of connate sepals that is characterized by ascending calyx lobes and an interior with lamellate sculpturing.

The four species in section *Asiasarum* bear seemingly equal morphological similarity to those of both subgenera *Asarum* and *Heterotropa* due to the presence of paired deciduous, broadly reniform leaves and stamen movement coupled with herkogamous flowers with calyx sculpturing. However, analysis of plastid and combined matrices recovers *A. sieboldii* as sister to, or within, section *Heterotropa*.

We recognize species that have deciduous leaves, exhibit stamen movement, bear flowers held above ground level with a 90° bend in the peduncle with abaxial
longitudinally sculpturing of the calyces from the proximal end of the sepal to the flower orifice, and a chromosome number of 2n = 26, as belonging to section *Asiasarum*.

*Subgenus* Heterotropa *section* Heterotropa –Morren and Decaisne (1834) recognized *Heterotropa* due to the presence of a ring-like annulus (hereafter referred to as an orifice ring) that separates the calyx lobes from the remainder of the calyx tube, the presence of reticulations or raised lines on the interior of the sepals, and free styles and actinomorphic flowers. Kelly (1997, 1998) found support for the monophyly of section *Heterotropa*, but Niedenberger (2010), using plastid sequences, found the group to be paraphyletic with respect to section *Asiasarum*.

Our topologies suggest that the orifice ring has likely arisen twice within *Asarum*, once in section *Heterotropa* and again in the newly identified sister group to section *Hexastylis*, section *Longistylis*. An alternative, but less parsimonious explanation is that it may have arisen once in the branch leading to subgenus *Heterotropa* and was subsequently lost in both sections *Asiasarum* and *Hexastylis*. None of the aforementioned features represent synapomorphies for this section. Morphology is highly diverse within section *Heterotropa*, and denser taxonomic sampling and additional phylogenetic resolution are needed before relationships within this section can be further interpreted.

We recognize species that have evergreen leaves, lack stamen movement, bear flowers with abaxial sculpturing of high relief that extend from the proximal end of the sepal to the flower orifice, and a chromosome number of 2n = 24, as belonging to section *Heterotropa*. 
Subgenus Heterotropa section Hexastylis – Rafinesque (1825) provided a reinterpretation of Asarum in which he segregated *A. virginicum* as the genus Hexastylis. Hexastylis was not adopted by Ashe (1897), who named four new species. Small (1903) was the first flora to use Hexastylis. Blomquist (1957) published the only comprehensive revision of Hexastylis, though he overlooked or was not aware of the work of Araki (1937, 1953) or Gregory (1956). He recognized three groupings, though informally; these overlap entirely with Araki’s (1953) subsections Plana Araki (Speciosa and Arifolia) and Cancellata Araki (Virginica).

Gaddy (1987) continued to recognize Hexastylis at the generic level, citing a difference in the number and morphology of chromosome pairs (Soltis, 1984), and the lack of a well sampled, world-wide study of *Asarum s.l.* as reasons to recognize this genus. Maekawa (1953) noted that sections Hexastylis, Asiasarum, and Heterotropa are closely allied, especially in relation to reproductive organ morphology. Araki (1937) recognized section Hexastylis, but later (1953) considered this section to be synonymous with Braun’s (1861) section Ceratasarum, which included *Asarum variegatum* A.Braun & C.D.Bouché, an Asian species (Kelly, 2001). We choose to follow the earlier Araki (1937) sense of section Hexastylis.

Blomquist’s (1957) Virginica and Arifolia groups are recovered here as distinct entities in analyses of all matrices, although their constituents and relationship to each other differ depending on the matrix and method used. All members of Blomquist’s informal Virginica group, with the exception of *A. lewissii*, form a monophyletic grouping
in all analyses. *Asarum speciosum*, which formed a clade with *A. arifolium* in Kelly’s (1998) investigation, is here found to be sister to *A. lewisii + A. arifolium*. The inclusion of *A. lewisii* in the *Arifolia* group was also recovered by Niedenberger (2010), a relationship that is extremely interesting due to its implications for floral and vegetative evolution within section *Hexastylis*. The placement of this species results in the blurring of Blomquist’s (1957) groups. Future analyses of more informative loci are needed to investigate the possibility of floral convergence between *A. lewisii* and the *Virginica* group. The placement of *A. lewisii* within the *Arifolia* group may suggest that the highly constricted and less basidiomycete-like flowers of *A. arifolium* have been derived only once in the section, and therefore may represent an evolutionary dead-end by way of the loss of putative fungal mimicking features that attract pollinators, and that leaf shape and variegation pattern are convergent between *A. arifolium* and *A. speciosum*.

The recognition of segregate genera has persisted in some floristic treatments, especially those with a focus on the southeastern United States (Blomquist, 1957; Whittemore and Gaddy, 1997; Whittemore et al., 1997; Weakley, 2012). This in particular is the case for the North American members of *Asarum* subgenus *Heterotropa*, commonly recognized at the generic level as *Hexastylis* (Blomquist, 1957; Gaddy, 1987; Weakley, 2012). Barringer (1993) published combinations in *Asarum* for some taxa that had only been validly published in *Hexastylis*, although he did not create new combinations for all names in the group. Despite this attempt to update the nomenclature to recognize *Asarum* containing *Hexastylis*, the latter was still treated at the generic level by Whittemore and Gaddy (1997) in their treatment in the *Flora of North America* as
well as *The Flora of the Southern and Mid-Atlantic States* (Weakley, 2012). Section *Hexastylis* is the only section of *Asarum* whose taxa remain commonly recognized at the generic level. This lack of taxonomic conformity is most likely due to the paucity of highly supported and inclusive phylogenetic evidence of lower level relationships within Asaroideae, as well as the regional focus of authors of floristic works who often make taxonomic decisions rooted in differing appearance rather than in evolutionary relationships as evidenced through synapomorphy.

We recognize species that have evergreen leaves, lack stamen movement, bear flowers with abaxial sculpturing of high relief that does not extend from the proximal end of the sepal to the flower orifice, and a chromosome number of 2n = 26 as belonging to section *Hexastylis*.

*Subgenus Heterotropa section Longistylis – Asarum maximum, A. delavayi, and A. splendens* were recovered as a monophyletic group in all analyses and are united by molecular and morphological characters, though the identity of its sister clade shifted dramatically depending on the genomic compartment analyzed. These species have not been included in earlier studies of *Asarum*.

We recognize this newly identified group as a new section, *Longistylis*, the defining features of which are overhanging ovoid stigmas placed sub-laterally below highly developed style extensions that are bifid to varying degrees, and yellow pollen. *Asarum maximum* and *A. delavayi* both have broad, elongate style extensions that are
divergent and wing-like at anthesis; those of *A. splendens* are divergent, though not to the same degree as the former species.

**Classification of Asarum L. All subgenera and sections of Asarum s.l. are listed.**

- Flowers non-herkogamous and autonomously self-pollinating by way of stamen movement, ovaries inferior, styles connate, abaxial calyx surface pubescent, abaxial surface of the calyx not sculptured, sepals not connate for more than half their length, leaves paired (except in *A. himalaicum*).

  - Calyx actinomorphic, sepals conspicuously free and not forming an elongate tube; chromosome number $2n = (24)26$.

*Asarum L. subgenus Asarum*

  - *A. Asarum L. section Asarum*

  - Calyx zygomorphic, the sepals falsely connate and extending into a pseudotube; chromosome number $2n = 12$.


- **B. Asarum** L. section *Geotaenium* (F.Maek.) L.Kelly

  - Flowers herkogamous and not autonomously self-pollinating due to lack of stamen movement (with the exception of section *Asiasarum*), ovaries 3/4 inferior to fully superior, styles at least partly free, abaxial calyx surface glabrous or with unicellular glandular trichomes, adaxial surface of the calyx with raised sculpturing, sepals connate and forming a tube which greatly surpasses the gynoecium, leaves solitary (with the exception of the flowering nodes of section *Asiasarum* species).

*Asarum* L. subgenus *Heterotropa* (Morr. & Decne.)

**O.C.Schmidt**

- Leaves deciduous, those of flowering nodes paired, those of non-flowering nodes solitary, adaxial calyx sculpturing of only longitudinal lines, flowers nodding and held above ground level, stamens horizontal at anthesis, chromosome number $2n = 26$. 

35
○ *C. Asarum* L. section *Asiasarum* (F.Maek.) Araki

○ Leaves evergreen, solitary, often glossy and variegated, flowers produced at ground level, adaxial calyx sculpturing often of longitudinal and horizontal lines and forming deep pits (rarely solely longitudinal lines), stamens erect at flower opening.

- Floral orifice often constricted with a membranous ring, abaxial calyx surface sculptured proximally and distally.

  - Pollen white, stigmas dorsal or terminal, style extensions hardly or not surpassing stigmas and weakly bifurcate, chromosome number 2n=24.

○ *D. Asarum* L. section *Heterotropa* (Morr. & Decne.) A.Braun

  - Pollen yellow, stigmas dorsal, bifurcate style extensions long surpassing stigmas, chromosome number 2n=24, 26, or 39.

- Floral orifice without ring, abaxial calyx surface not sculptured distally, chromosome number 2n=26.

• *F. Asarum* L. section *Hexastylis* (Raf.) Araki


---

*Loss of autonomous self-pollination and derivation of putative floral mimicry of fungi*

The relationships of subgenus *Geotaenium* and section *Asiasarum* to the remainder of the genus hold important implications for the interpretation of floral evolution in *Asarum*. Self-pollination in angiosperms, due to the breakdown of barriers to
self-pollination, commonly represents a derived condition (Schoen et al., 1997; Takebayashi and Morrell, 2001). Unfortunately, our topologies do not unambiguously support a single story of floral evolution in the genus. An additional challenge to our understanding of floral evolution in Asarum is the lack of morphological comparability between the species that comprise it and Saruma henryi – this species is the closest relative of the genus, but little is comparable between the two in regards to floral morphology. Although very little is known of the pollination of Asarum species, we now know from the work of Kelly (1997; 1998) and results presented here, that a lack of mechanisms responsible for autonomous self-pollination and the presence of those which have been implicated in floral mimicry of fungi are associated with the majority of derived monophyletic groups within the genus, and serve as synapomorphies for subgenus Heterotropa, although none serve as synapomorphies for any section. The traits that are thought to facilitate fungal mimicry together characterize flowers that are not capable of autonomous self-pollination, and have superior ovaries and connate sepals forming elongate calyx tubes which are sculptured interiorly.

The lack of autonomous self-pollination in sections Asiasarum, Heterotropa, Hexastylis, and Longistylis can be interpreted as the derived condition in the genus if subgenus Geotaenium is sister to the remainder of the genus; this scenario is exemplified in our PartitionFinder-optimized combined ML analysis (Figure 5A). Alternatively, if Asarum is composed of two major clades -- that is if subgenera Asarum and Geotaenium were robustly recovered by future work to form a monophyletic group as recovered by our non- and genome-partitioned ML analyses and all BI analyses – then the optimization
of pollination mode would be ambiguous due to the non-comparable floral morphology of *Saruma henryi* (Figure 5B). Our MP and ML analyses of rDNA-only matrices support Kelly’s (1997; 1998) findings with the recovery of section *Asiasarum* as sister to the remainder of subgenus *Heterotropa*. The species of section *Asiasarum* exhibit stamen movement but are herkogamous; our rDNA-based topologies result in the interpretation of the loss of autonomous self-pollination mechanisms as derived (Figure 5C).
Figure 5. The presence of autonomous self-pollination (square symbol) and the absence of autonomous self-pollination (star symbol) plotted on collapsed representations of the PartitionFinder-optimized ML cladogram of combined plastid and nuclear data (A), non-genome-partitioned ML and BI trees (B), and the ML and MP cladograms inferred using only rDNA data (C). Note that Saruma henryi is marked with a question mark to underscore the lack of floral comparability between it and Asarum species.
Furthermore, rDNA suggests a single loss of delayed stamen movement and single origins of the constrictive floral orifice ring and an entirely sculptured calyx interior in sections Heterotropa and Longistylis, yet recovers section Hexastylis as paraphyletic. Contrastingly, the majority of our analyses suggest that either stamen movement was retained in section Asiasarum or that stamen movement has been lost independently in sections Hexastylis, Longistylis, and Heterotropa. In response to Kelly (1997), Takebyashi and Morrell (2001) remarked that the lack of robust estimations of outcrossing rates in herkogamous, non-autonomously selfing Asarum species precludes the assertion that a truly outcrossing condition has been derived from one of selfing. At minimum, the loss of ensured selfing via stamen movement, and the derivation of herkogamy in the most species-rich clades of Asarum, suggests that successful pollination by external forces has been pervasive enough during the diversification of Asarum to not impede cladogenesis in taxon-rich clades relative to sections Asarum, Asiasarum, and Geotaenium.

Influence of data partitioning

In our analyses, support values for clades where relationships were poorly resolved generally decreased with increased partitioning, and those of clades that were commonly recovered as highly supported were more often than not increased or unchanged (Table 2); this is especially evident in our combined plastid and nuclear analyses.
Our ML analyses were more sensitive to partitioning scheme choice than BI as implemented in BEAST; this is evidenced by our recovery of more likely, novel topologies identified only through partitioning (Figure 6; Table 4). The confident estimation of parameter probability distributions may be made more difficult as partition number increases and thus sequence length decreases (Fenn et al., 2008). Evidence of this may be the increased difficulty of reaching accepted sampling cutoffs, such as ESS, in highly partitioned analyses (Table 4).
Figure 6. The likelihood of ML (A) and posterior probability of BI (B) trees improved in our PartitionFinder and marker-partitioned analyses. Akaike Information Criterion values decreased with increased matrix partitioning under ML (A), but BIC values were only slightly impacted by marker partitioning, and then only negatively so. Partition number of each matrix analyzed is listed on the x axis.
Table 4. PartitionFinder and marker partitioning scheme matrices better fit models of evolution and produced trees of greater likelihood in ML analyses. PartitionFinder partitioned and marker partitioned matrices produced Bayesian trees of greater likelihood than those of less partitioned datasets. PartitionFinder (Likelihood, AIC, BIC) and RAxML (ML tree likelihood) scores for combined matrices partitioned singly, by genomic compartment, by PartitionFinder best scheme, and by marker are shown. The best scores for each criterion are denoted with an asterisk.
<table>
<thead>
<tr>
<th>Method</th>
<th>Matrix</th>
<th>PartitionFinder Likelihood</th>
<th>ESS</th>
<th>Posterior</th>
<th>ESS</th>
<th>BIC</th>
<th>AIC</th>
<th>RAXML Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI</td>
<td>1</td>
<td>-23300.401</td>
<td>2773.477</td>
<td>-23292.874</td>
<td>826.03</td>
<td>47938.21389*</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-22831.498</td>
<td>1413.526</td>
<td>-22866.534</td>
<td>450.794</td>
<td>48123.01567</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>BI</td>
<td>6</td>
<td>-21994.707</td>
<td>2696.334</td>
<td>-22045.046</td>
<td>915.582</td>
<td>48371.72424</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>BI</td>
<td>3</td>
<td>-21882.159*</td>
<td>1440.253</td>
<td>-21910.27*</td>
<td>52.265</td>
<td>52375.93314</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>ML</td>
<td>1</td>
<td>-23243.64345</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>46733.2869</td>
<td>-23209.36683</td>
</tr>
<tr>
<td>ML</td>
<td>8</td>
<td>-22776.32176</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>46044.64352</td>
<td>-22976.1245</td>
</tr>
<tr>
<td>ML</td>
<td>6</td>
<td>-21856.58057</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>44697.16114*</td>
<td>-22351.94749</td>
</tr>
<tr>
<td>ML</td>
<td>3</td>
<td>-21614.88262*</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>45197.76523</td>
<td>-22303.75914*</td>
</tr>
</tbody>
</table>
The partition scheme-induced changes in topology under a ML framework may be due to sounder estimations of model parameters by improving model fit (McGuire et al., 2007; Lanfear et al., 2014), since we found that increased matrix partitioning improved both tree likelihood (ML) and posterior probability (BI), as well as model fit values (AIC and BIC) for both ML and BI (Figure 6). However, we assume that an upper limit to the improvement of likelihood scores and AIC values exists, and that they would not improve from an ever-increasing number of partitions.

**Summary**

We argue that *Asarum* is best recognized as a single genus of three subgenera and six sections. The PartitionFinder-recommended partitioning scheme represents a biologically realistic way to compartmentalize our data, a strategy that Cameron et al. (2007) found to perform best, and we find it compelling that the most parsimonious explanation of our data and the most likely tree from the ML analysis of our PartitionFinder combined matrix are entirely congruent at the level of taxonomic section. We strongly encourage workers to compare the consequences of partitioning their matrices across multiple analysis types. The morphological synapomorphies put forth here represent consistent and defining differences in character states, and these traits will form the basis of future studies of character-associated diversification in *Asarum*. This may prove to be an especially fruitful line of investigation as the loss of stamen movement and the derivation of floral characters that have been implicated in fungal mimicry may have evolved multiple times in *Asarum*. It is our hope that the relationships
recovered by this work will demonstrate to those working on future floristic treatments that the point at which molecules and morphology meet, and that represents a substantive divergence from what can be thought of as the floral bauplan of *Asarum*, is the supporting node of subgenus *Heterotropa*, and that the seemingly unusual morphological features of section *Hexastylis* are not novelties but, based on taxonomic richness, more the rule in *Asarum*. 
Chapter 2. Putative floral brood-site mimicry, loss of autonomous selfing, and reduced vegetative growth are significantly correlated with increased diversification in *Asarum* (Aristolochiaceae).

**Introduction**

Approximately 90% of extant plant species are angiosperms (The Plant List, 2013). It is generally accepted that the cladogenic success of the flowering plants is due to myriad ecological and evolutionary opportunities that have been exploited by lineages that possess uniquely beneficial character states (see Crepet and Niklas, 2009), so-called key innovations (Simpson, 1969). However, it is clear that asymmetry exists in the ability of lineages to diversify, or that there is differential extinction, since species richness is unequally distributed throughout the angiosperm tree of life (Vamosi and Vamosi, 2010).

Recently, many authors have investigated the rates of speciation within their group of focus due to the ease of implementing Binary-State Speciation and Extinction Model (BiSSE) Net Diversification Likelihood (Maddison et al., 2007) and similar diversification analyses (see Pyron and Burbrink, 2013). Such analyses promise to do much to further our understanding of asymmetry in lineage richness by simultaneously investigating the effects of differential divergence (λ), extinction (μ), and character transition rates (q) in light of branch lengths on a given phylogeny, though they must be interpreted with caution (Rabosky, 2010; Davis et al., 2013). Character state changes
affecting the reproductive biology of a group can increase clade diversity in two ways, through reduced extinction rate resulting in a gradual accumulation of species or by causing an increase in speciation rate resulting in a rapid increase in the number of lineages in a clade (Armbruster, 2014). Asymmetrical character state transition rates (q) can be indicative of key innovation(s) when a pattern of convergent evolution is seen in a phylogeny; that is, a trait or traits has evolved many times, but has been infrequently lost (Johnson et al., 2011; Armbruster, 2014). Differential extinction rates (μ) can lead to increased diversity through the buildup of species in a clade due to the development of a character state that reduces the chance of extinction for its species relative to those of another clade. Diversification rate (λ) increases occur when a character state predisposes a lineage to undergo cladogenesis. Clades that harbor much morphological diversity and large numbers of species that are the result of an increased rate of diversification, rather than due to relaxed extinction rate, are commonly known as adaptive radiations (Simpson, 1969; Yoder et al., 2010).

Many have proposed that floral specialization (Nilsson, 1983; Hodges and Arnold, 1995; Van der Niet et al., 2014a; Armbruster, 2014) can play significant roles in influencing the diversification of lineages. Floral specialization in the form of mimicry can be found throughout the angiosperms, and has been documented in species-rich lineages such as Aristolochia (Oelschlägel et al., 2014), Amorphophallus (Kite and Hetterschieid, 1997), Arum (Urru et al., 2011), and the orchids (Ackerman, 1986; Tremblay et al., 2005). If we are to further refine hypotheses of angiosperm radiation,
targeted analyses of portions of the angiosperm tree that contrast greatly not only in extant species richness, but also phenotype, are needed.

One such case can be found in *Asarum* L. (Aristolochiaceae) of the Piperales, the sole order in the magnoliids that has experienced rapid diversification of herbaceous lineages (Isnard et al., 2012). A number of distinct growth strategies, floral ornamentation suites, biogeographical histories, and hypothesized pollination modes together comprise the life histories of species in *Asarum* (Cheng and Yang, 1983; Kelly, 1997, 1998, 2001; Lu and Wang, 2009; Shi et al., 2008; Shumei et al., 2003; Whittemore and Gaddy, 1997; Whittemore et al., 1997; Xiang and Soltis, 2001; Xiang et al., 2004). These herbaceous species variously produce solitary or paired evergreen or deciduous leaves along short or elongate rhizomes and bear solitary, apetalous flowers that differ in the presence of stamen-stigma separation (herkogamy), sepal connation, delayed stamen movement, ornamentation, and scent (Blomquist, 1957; Gaddy, 1987a; Kelly, 2001). The three major clades of *Asarum* (Euasarum, Geotaenium, and Heterotropa) differ in vegetative growth form, floral morphology, their ability to autonomously self-pollinate, and perhaps most greatly in species richness. Shifts in pollination biology have been shown to impact diversification (Johnson, 2006; Van der Niet et al., 2014a) and the flowers of these major *Asarum* clades differ in their possession of putative fungal mimic structures such as basidiomycete-shaped campaniform calyces with interior gill-like reticulations (Vogel, 1978; Lu 1982; Sugawara, 1988, Leins and Erbar, 2010). Furthermore, Kelly (1997, 1998) found that autonomous self-pollination is the pleisiomorphic condition in *Asarum*,
and since this reproductive mode is found in the minority of *Asarum* species, suggests that an increase in outcrossing may have played a role in the diversification of the genus.

The marked difference in species richness and floral and vegetative traits between *Asarum* subgenera suggests that directional selection incurred either by biased diversification, due to the development of key phenotypic innovations, or extinction rate has produced the observed differences in species richness between *Asarum* clades. Here, in the first BiSSE-informed character-associated diversification analysis in the magnoliids, we ask: 1) Is asymmetry in *Asarum* clade richness best explained by differential diversification, character-state transition, or extinction rates, singly or in combination? and 2) Do differences in diversification, character transition, and/or extinction rates significantly co-vary with putative floral mimic and/or vegetative characters that we can identify in *Asarum*?

**Materials and Methods**

**DNA extraction, sequence generation, and matrix construction**

Fresh leaf material was field collected and stored at -80°C. Total DNA extractions were conducted using the CTAB method (Doyle and Doyle, 1987) modified to use 67 mg of leaf tissue which was disrupted using a bead mill and stored at -20°C.

Eight regions were amplified: *rpoB-trnC* \(^{GCA}\), *rps16-trnK*, *trnL* \(^{UAA}\) exon – *trnF* \(^{GAA}\); *trnT* \(^{UGU}\) – *trnL* \(^{UAA}\); *trnL* \(^{UAA}\) – *trnL* \(^{UAA}\) exon, *trnS* \(^{UGA}\) - *trnFM* \(^{CAU}\), and *ycf1* 3850-5310 from the chloroplast and ITS1 – partial 26S from the nucleus. All taxa, regions amplified, and
resultant topologies are shown in Appendix A. Taxa from all segregate genera of *Asarum* s.l. were included in our analysis, with care taken to include the designated type species of each. PCR amplifications for all chloroplast markers were carried out using the *rpl16* program of Shaw et al. (2005). The PCR program for all rDNA sequence generated was as follows: initial template denaturation at 95°C for three minutes followed by 34 cycles of denaturation at 95°C for 30 seconds, primer annealing at 55°C for one minute, and primer extension at 72°C for one minute. The program was ended with an additional five minutes of primer elongation at 72°C and held at 4°C. ITS1 and 2 sequences were obtained from GenBank for *A. splendens* (58430).

Amplicons were sequenced by Beckman-Coulter Genomics and inspected and assembled in Geneious versions 5.3-6.1 (Drummond et al., 2011). Sequence assemblies were trimmed using an α of 0.05, and all were reference guided using the consensus read implied by an assembly constructed of the highest quality read from each locus. Sequences of poor quality and ambiguous base calls were edited by hand using context provided by the reference sequence. Alignments were generated using MAFFT (Katoh et al., 2002) in Geneious version 7.1 (Drummond et al., 2011) using automatic algorithm selection and the 200PAM/k=2 scoring matrix with a gap open penalty of 1.53 and an offset value of 0.123. The resulting alignment was then visually inspected and manually adjusted, and ambiguous regions of the alignment that could not be improved were removed.
Model selection and partitioning of matrices

PartitionFinder v1.1.0 (Lanfear et al., 2012) was used to assess evolutionary model fit and data partitioning schemes. All models that can be implemented in RAxML and PartitionFinder search types were used to evaluate 4140 possible schemes. The AIC criterion identified GTR + Γ + I as the model of best fit. We recognize that the chloroplast genome is a single non-recombinatory unit; however chloroplast markers were input into PartitionFinder as distinct units in order to fully explore their relative impacts on phylogenetic reconstruction.

Maximum likelihood analyses

The high performance computing Pthreads version of RAxML 7 (Stamatakis, 2006; Ott et al., 2007) implemented in RAxML GUI (Silvestro, 2012) was used for Maximum Likelihood analysis. We used the GTRGAMMAI model which includes the Γ rate heterogeneity model, estimates the proportion of invariant sites, and used 6 discrete rate categories. Search for the most likely ML tree was carried out using the ML + rapid bootstrap option with 1000 bootstrap replicates. BiSSE analyses were performed on the topology of highest likelihood.

Species tree inference

Bayesian coalescent inference (BCI) was conducted using *BEAST (Drummond et al., 2012) and the BEAGLE API was implemented for multicore processing (Ayres et al., 2011). This program estimates a phylogeny by averaging across multiple gene trees.
BEAST XML files were constructed using BEAUti (Drummond et al., 2012). The Ploidy Type setting was set to mitochondrial for all chloroplast regions to more accurately reflect inheritance and rate heterogeneity between markers; rDNA regions were set to autosomal nuclear. A coalescent process was used for species tree prior estimation (Kingman, 1982), and a piecewise linear and constant root population size model was used. Default priors were used for all other parameters. The MCMC chain was run for 100,000,000 generations and a 10 percent burn-in was used; parameters were logged every 1,000 generations. Tracer version 1.7 (Rambaut and Drummond, 2008) was used to monitor parameter estimations as well as ESS during analysis. The inferred species tree used for BiSSE analysis was constructed using Tree Annotator (Drummond et al., 2012), with a burn-in of 25% of the total trees generated and node annotation posterior probability cutoff set to > 0.95. Figtree version 1.4 (Rambaut, 2012) and TreeGraph2 (Stöver and Müller, 2010) were used for tree visualization and manipulation.

**BiSSE character-associated diversification analysis**

The BiSSE ln Likelihood Net Difference function of Mesquite (Maddison and Maddison, 2011) was used to test for character-associated diversification. We used twice the net difference of the proportional negative log likelihoods of our constrained and unconstrained models as a test statistic for comparison against critical values drawn from a chi-square distribution ($\alpha = 0.05; df = 1; critical value = 3.841$). Each constrained model represents a null hypothesis of symmetric speciation, extinction, or character state transition rate of a binary trait of interest distributed on our cladograms. Characters of
interest and their distribution on our trees are shown in Figure 7. The BiSSE method assumes exhaustive sampling. While we do not include all species of *Asarum s.l.* in our analysis, we sample proportionally throughout the genus, with approximately half of known species, as well as approximately half of each major subgeneric group.
Figure 7. ML (left) and BCI (right) trees used in this study; increased diversification is significantly correlated with decreased vegetative growth and characters of putative floral brood-site mimicry. Character states tested here are shown to the left of each taxon (legend in upper left). Change in taxon position in our BCI topology relative to position in the ML tree is denoted by lines. Clades are colored according to clade: Geotaenium = yellow, Euasarum = blue, Asiasarum = purple, Braun = red, Longistylis = black, and Hexastylis = green. Values above branches represent ML bootstrap and posterior probabilities greater than 50 and 0.95, respectively.
We chose not to collapse poorly supported nodes recovered by our analysis. Branch lengths that were too near zero for BiSSE analysis through Mesquite were assigned values of $1 \times 10^{-8}$. We argue that the inclusion of these nodes is more biologically justified than the use of a tree with arbitrarily resolved polytomies (Kuhn et al., 2011) or polytomies that result from collapsing nodes due to an arbitrary cutoff point for support values (FitzJohn et al., 2009).

The alternative hypothesis for each character is represented by our unconstrained models, with all parameters decoupled, and is therefore able to account for differential rates of speciation, extinction, and character state transitions. The unconstrained model was tested against each null in turn, constraining each of the three parameters iteratively. Maximum likelihood optimizations were minimized to four attempts, since sensitivity tests showed that optimization beyond this number failed to significantly affect recovered parameters, and thus p-values recovered by our likelihood ratio tests.

**Results**

*Sequences and alignment*

Our aligned rDNA sequences (TreeBASE submission 17127) totaled 868 base pairs in length, with missing data, including alignment gaps, making up 5.5% of the alignment. The alignment of our seven chloroplast markers yielded a matrix whose length was 7458 bp. Alignment of chloroplast sequences resulted in many more gaps than alignment of rDNA; missing data and alignment gaps accounted for 25.2% of the matrix.
**Phylogenetic inference results**

The topology recovered by ML analysis places the Geotaenium clade as sister to the remainder of *Asarum* (ML bootstrap < 50; Figure 7). In contrast to our BCI species tree, the Asiasarum clade is sister to the Heterotropa clade (ML bootstrap = 96). The relationship of the Longistyris clade to the Hexastylis clade is identical to that of our BCI species tree (ML bootstrap = 91).

Our BCI phylogeny (Figure 7) recovers the Geotaenium clade as sister to the Euasarum clade, and is mostly congruent with our ML analysis with the exception of the placement of the Asiasarum clade, which here is recovered within the Braun clade (Figure 7). The Longistyris clade is recovered as sister to the Hexastylis clade (posterior probability = 1.0). The resulting species tree serves as an additional hypothesis of diversification of *Asarum*.

**BiSSE character-associated diversification results**

Likelihood ratio tests identify statistically significant increases in model likelihoods for all five characters tested on our BCI species tree; in all cases, model likelihood improvement is driven by increased diversification rates in the major Heterotropa clade relative to the Euasarum clade. Increased rates of diversification were linked to shifts to partly or fully connate sepals (p < 0.001; all BCI values are shown in Table 5), at least a partially superior ovary (p = 0.01), one leaf per node (p < 0.002), the herkogamous condition (p < 0.001) and reduced internode length (p < 0.006).
Table 5. BiSSE model parameters, likelihoods, and likelihood ratio test results of constrained and unconstrained models of character state distributions tested on our *BEAST BCI species tree.
<table>
<thead>
<tr>
<th>Character</th>
<th>Model</th>
<th>Free Parameters</th>
<th>$\lambda_0$</th>
<th>$\lambda_1$</th>
<th>$\mu_0$</th>
<th>$\mu_1$</th>
<th>$q_{01}$</th>
<th>$q_{10}$</th>
<th>lnLik</th>
<th>$2\Delta$-lnL</th>
<th>Chi P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herkogamy</td>
<td>Unconstrained</td>
<td>6</td>
<td>E+02</td>
<td>E+02</td>
<td>E+05</td>
<td>E+02</td>
<td>E+01</td>
<td>E+01</td>
<td>E+02</td>
<td>HA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\lambda_0 = \lambda_1$</td>
<td>5</td>
<td>E+02</td>
<td>E+01</td>
<td>E+01</td>
<td>E+01</td>
<td>E+04</td>
<td>E+02</td>
<td>0.0004***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_0 = \mu_1$</td>
<td>5</td>
<td>E+02</td>
<td>E+05</td>
<td>E+01</td>
<td>E+01</td>
<td>E+01</td>
<td>E+02</td>
<td>0.0893</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$q_{01} = q_{10}$</td>
<td>5</td>
<td>E+02</td>
<td>E+04</td>
<td>E+02</td>
<td>E+01</td>
<td>E+02</td>
<td>E+02</td>
<td>0.4205</td>
<td></td>
</tr>
<tr>
<td>CalyxCONNATION</td>
<td>Unconstrained</td>
<td>6</td>
<td>E+02</td>
<td>E+02</td>
<td>E+05</td>
<td>E+02</td>
<td>E+04</td>
<td>E+01</td>
<td>E+02</td>
<td>HA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\lambda_0 = \lambda_1$</td>
<td>5</td>
<td>E+02</td>
<td>E+02</td>
<td>E+04</td>
<td>E+04</td>
<td>E+01</td>
<td>E+02</td>
<td>0.0043**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_0 = \mu_1$</td>
<td>5</td>
<td>E+02</td>
<td>E+02</td>
<td>E+06</td>
<td>E+01</td>
<td>E+02</td>
<td>E+00</td>
<td>0.0823</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$q_{01} = q_{10}$</td>
<td>5</td>
<td>E+02</td>
<td>E+02</td>
<td>E+04</td>
<td>E+01</td>
<td>E+02</td>
<td>E+00</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Ovary POSITION</td>
<td>Unconstrained</td>
<td>6</td>
<td>E+02</td>
<td>E+02</td>
<td>E+00</td>
<td>E+02</td>
<td>E+05</td>
<td>E+01</td>
<td>E+02</td>
<td>HA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\lambda_0 = \lambda_1$</td>
<td>5</td>
<td>E+02</td>
<td>E+02</td>
<td>E+05</td>
<td>E+01</td>
<td>E+02</td>
<td>E+02</td>
<td>0.0101*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_0 = \mu_1$</td>
<td>5</td>
<td>E+02</td>
<td>E+02</td>
<td>E+05</td>
<td>E+01</td>
<td>E+02</td>
<td>E+00</td>
<td>0.1826</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$q_{01} = q_{10}$</td>
<td>5</td>
<td>E+02</td>
<td>E+02</td>
<td>E+04</td>
<td>E+02</td>
<td>E+02</td>
<td>E+01</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Leaves per Node</td>
<td>Unconstrained</td>
<td>6</td>
<td>E+02</td>
<td>E+02</td>
<td>E+05</td>
<td>E+02</td>
<td>E+01</td>
<td>E+01</td>
<td>E+02</td>
<td>HA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\lambda_0 = \lambda_1$</td>
<td>5</td>
<td>E+02</td>
<td>E+02</td>
<td>E+05</td>
<td>E+02</td>
<td>E+02</td>
<td>E+02</td>
<td>0.0019**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_0 = \mu_1$</td>
<td>5</td>
<td>E+02</td>
<td>E+02</td>
<td>E+05</td>
<td>E+02</td>
<td>E+02</td>
<td>E+00</td>
<td>0.0791</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$q_{01} = q_{10}$</td>
<td>5</td>
<td>E+02</td>
<td>E+02</td>
<td>E+04</td>
<td>E+02</td>
<td>E+02</td>
<td>E+01</td>
<td>0.6278</td>
<td></td>
</tr>
<tr>
<td>Internode LENGTH</td>
<td>Unconstrained</td>
<td>6</td>
<td>E+02</td>
<td>E+02</td>
<td>E+03</td>
<td>E+02</td>
<td>E+04</td>
<td>E+02</td>
<td>E+02</td>
<td>HA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\lambda_0 = \lambda_1$</td>
<td>5</td>
<td>E+02</td>
<td>E+02</td>
<td>E+03</td>
<td>E+02</td>
<td>E+02</td>
<td>E+02</td>
<td>0.0051**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu_0 = \mu_1$</td>
<td>5</td>
<td>E+02</td>
<td>E+02</td>
<td>E+04</td>
<td>E+02</td>
<td>E+02</td>
<td>E+02</td>
<td>0.367</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$q_{01} = q_{10}$</td>
<td>5</td>
<td>E+02</td>
<td>E+02</td>
<td>E+03</td>
<td>E+02</td>
<td>E+02</td>
<td>E+01</td>
<td>0.3898</td>
<td></td>
</tr>
</tbody>
</table>
Enforcing symmetrical rates of both character state transition and extinction across lineages did not significantly affect the likelihood of any character distribution tested, although higher values for each parameter were generally found with the derived character states; this suggests that the rate of extinction and gains or losses of our tested characters, independent of their estimated rate of diversification or extinction, does not explain their observed distribution on our BCI tree.

All tested characters were also found to be significant when tested using our ML phylogram; the constraint of diversification rate significantly reduced model likelihood for herkogamy (p=0.0374; all ML values shown in Table 6), sepal connation (p=0.0374), ovary position (p=0.0227), number of leaves per node (p=0.0078), and internode length (p=0.0047). The constraint of character state transition rate or extinction rate did not significantly reduce model likelihood for any of the characters tested on our Maximum Likelihood topology (p-values spanning 0.1459 to 1.0), though higher values of each were generally found for the derived state of each character tested.
Table 6. BiSSE model parameters, likelihoods, and likelihood ratio test results of constrained and unconstrained models of character state distributions tested on our ML phylogram.
<table>
<thead>
<tr>
<th>Character</th>
<th>Model</th>
<th>Free Parameters</th>
<th>λ0</th>
<th>λ1</th>
<th>μ0</th>
<th>μ1</th>
<th>q01</th>
<th>q10</th>
<th>lnLik</th>
<th>2Δ-lnL</th>
<th>Chi P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herkogamy</td>
<td>Unconstrained</td>
<td>6 E=00</td>
<td>E=00</td>
<td>E=05</td>
<td>E=05</td>
<td>E=01</td>
<td>E=01</td>
<td>E=01</td>
<td>E=00</td>
<td>0.0374*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 λ0 = λ1</td>
<td>3.95</td>
<td>8.13</td>
<td>3.07</td>
<td>1.02</td>
<td>9.61</td>
<td>-1.01</td>
<td>4.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 μ0 = μ1</td>
<td>2.01</td>
<td>5.23</td>
<td>1.85</td>
<td>1.84</td>
<td>1.84</td>
<td>-1.23</td>
<td>-1.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 q01 = q10</td>
<td>2.01</td>
<td>5.23</td>
<td>1.04</td>
<td>5.84</td>
<td>1.85</td>
<td>-1.23</td>
<td>-6.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calyx</td>
<td>Unconstrained</td>
<td>6 E=00</td>
<td>E=00</td>
<td>E=05</td>
<td>E=05</td>
<td>E=04</td>
<td>E=04</td>
<td>E=01</td>
<td>E=00</td>
<td>0.0306*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 λ0 = λ1</td>
<td>4.19</td>
<td>1.69</td>
<td>2.07</td>
<td>1.25</td>
<td>2.92</td>
<td>-8.54</td>
<td>4.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 μ0 = μ1</td>
<td>2.07</td>
<td>4.84</td>
<td>9.46</td>
<td>1.13</td>
<td>2.21</td>
<td>-1.06</td>
<td>6.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 q01 = q10</td>
<td>2.03</td>
<td>4.77</td>
<td>2.12</td>
<td>2.65</td>
<td>-1.09</td>
<td>5.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ovary Position</td>
<td>Unconstrained</td>
<td>6 E=00</td>
<td>E=00</td>
<td>E=05</td>
<td>E=05</td>
<td>E=04</td>
<td>E=04</td>
<td>E=01</td>
<td>E=00</td>
<td>0.0227*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 λ0 = λ1</td>
<td>4.24</td>
<td>1.83</td>
<td>1.33</td>
<td>1.43</td>
<td>5.76</td>
<td>-2.72</td>
<td>5.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 μ0 = μ1</td>
<td>2.03</td>
<td>4.99</td>
<td>3.00</td>
<td>5.44</td>
<td>-5.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 q01 = q10</td>
<td>2.05</td>
<td>5.04</td>
<td>8.51</td>
<td>3.99</td>
<td>4.64</td>
<td>-5.17</td>
<td>2.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaves per Node</td>
<td>Unconstrained</td>
<td>6 E=00</td>
<td>E=00</td>
<td>E=06</td>
<td>E=04</td>
<td>E=04</td>
<td>E=01</td>
<td>E=00</td>
<td>E=00</td>
<td>0.5932</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 λ0 = λ1</td>
<td>1.89</td>
<td>5.23</td>
<td>1.15</td>
<td>3.99</td>
<td>3.54</td>
<td>4.64</td>
<td>-8.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 μ0 = μ1</td>
<td>2.05</td>
<td>4.13</td>
<td>1.12</td>
<td>9.39</td>
<td>3.20</td>
<td>4.45</td>
<td>-4.81</td>
<td>7.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 q01 = q10</td>
<td>1.89</td>
<td>5.23</td>
<td>3.00</td>
<td>5.44</td>
<td>-5.32</td>
<td>-1.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internode Length</td>
<td>Unconstrained</td>
<td>6 E=00</td>
<td>E=00</td>
<td>E=05</td>
<td>E=05</td>
<td>E=04</td>
<td>E=04</td>
<td>E=01</td>
<td>E=00</td>
<td>0.0078***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 λ0 = λ1</td>
<td>1.89</td>
<td>5.23</td>
<td>1.01</td>
<td>3.54</td>
<td>4.64</td>
<td>-8.35</td>
<td>-2.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 μ0 = μ1</td>
<td>1.92</td>
<td>2.37</td>
<td>4.58</td>
<td>1.01</td>
<td>4.64</td>
<td>-3.08</td>
<td>4.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 q01 = q10</td>
<td>1.78</td>
<td>5.37</td>
<td>5.90</td>
<td>9.00</td>
<td>1.43</td>
<td>8.07</td>
<td>-3.75</td>
<td>4.19</td>
<td></td>
</tr>
</tbody>
</table>

Total: 63
**Discussion**

As early as 1970, Stebbins recognized that pollinator shifts often involved few characters, and that these often occur in correlated sets. We find a tight coupling of vegetative and floral character transitions between the three major *Asarum* clades. The loss of stamen movement, derivation of herkogamy, and the connation of sepals are tightly correlated with the production of a single evergreen leaf per node, where internode length is more often short than not. Our BiSSE results indicate that increased diversification in the Heterotropa clade is significantly correlated with an increase in calyx complexity, loss of autonomous selfing, and a decrease in vegetative growth. This suggests that an increased investment in sexual reproduction and increased reliance on pollen vectors, even in light of potentially reducing seed set through the loss of delayed autonomous self-pollination, may positively influence lineage diversification relative to autonomously selfing lineages that invest more in vegetative growth.

**Floral morphology**

Flowers of *Asarum* vary from highly ornamented with ornately sculptured and curved calyx tubes and superior ovaries, to those with little ornamentation other than a high degree of pubescence, and free sepals and inferior ovaries. The flowers of the Euasarum (Figure 8 B – D) and Geotaenium clades have inferior ovaries with united styles, while the ovaries of the Heterotropa clade (Figure 8 E – M) are more often superior and the styles are not connate as in the other subgenera. In the Euasarum clade,
the lateral portions of the styles are united to form a central column. The styles are free and distinct in the Heterotropa clade, where an oftentimes highly contrasting depression is evident in the central portion of the gynoecium; the flowers of species engaged in mimicry are commonly reported with proximal portions that are lighter in color than the surrounding tissue (Urru et al., 2011; Jürgens et al., 2013).
Figure 8. Photographs of the flowers of *Asarum canadense* (Euasarum clade, B – D), *A. delavayi* (E – G), *A. arifolium* var. *arifolium* (H – J), and *A. shuttleworthii* var. *shuttleworthii* (K – M) of the Heterotropa clade showing the floral orifice (B, E, H, K), the floral interior (C, F, I, L), and the calyx tube (D, G, J, M).
It has been proposed that the predominant force in floral evolution is natural selection for reproductive isolation (Levin, 1971; Jones and Reithel, 2001), and recent research suggests that pollinator shifts can drive diversification (Forest et al., 2014; Van der Niet et al., 2014a; see section 1.4.2). The phylogenies presented here suggest that the Appalachian endemic Hexastylis clade is the result of diversification from a single introduction from central China. The lineages of this clade have diverged along two general evolutionary trajectories, one comprising ten species with highly sculptured calyces with broad floral orifices, and the other comprising three species with poorly sculptured calyces and highly restricted orifices. *Asarum arifolium*, of the less speciose lineage, is the most common and broadly distributed of the Hexastylis clade species, and can be found growing in sympatry with nearly all other species of the clade. We hypothesize that the marked differences in floral morphology, range sizes, and diversification (Figure 8 H – M) between the *A. arifolium* lineage and the remainder of the Hexastylis clade may be reflective of differences in pollination biology. BiSSE analyses indicate that lineages with complex, highly-sculptured calyces have diversified more rapidly than other lineages in *Asarum*; this is consistent with higher BiSSE-indicated diversification associated with increased floral investment and the putative attractive nature of the diverse calyx shapes in the genus.

*Pollination and putative floral mimicry of fungi*

The pollination biology of *Asarum* species is variable, since some are autonomously self-pollinated through delayed stamen movement while others are
herkogamous and have stationary stamens; species of the Asiasarum clade are herkogamous yet still exhibit stamen movement (Kelly, 1997, 2001). Stamens of the Euasarum and Geotaenium clades begin development at a 90° angle to the gynoecium, gradually becoming erect after the flower opens, and the dehisced anthers come into contact with the stigmatic surface. Herkogamous flowers rely on mechanisms, biotic or abiotic, to move pollen from the anther to the stigma. Additionally, all *Asarum* species are protogynous (Kelly, 1997), meaning that the stigmas are pollen receptive prior to anther dehiscence and also providing a brief window for cross pollination. The presence of herkogamy in some *Asarum* species is representative of a pollination mode that is predominantly outcrossing (Faegri and Van Der Pijl, 1979). In environments where pollinator service is limited, shifts toward increased reproductive performance (e.g. selfing) may be favored; in contrast, when reproductive isolation from congeners or release from interspecific interactions is favored, then reproductive divergence may arise (Johnson, 2006).

The relationships between herkogamous and non-herkogamous *Asarum* species have been rigorously investigated and interpreted in a phylogenetic framework by Kelly (1997) and Sinn et al. (2015a), though the pollination biology of *Asarum* species remains poorly understood. Wyatt (1954) determined that *A. virginicum* is self-fertile, yet self-pollination did not occur in any pollinator-excluded flowers; seed set was nearly total in *A. canadense*, a non-herkogamous species in the Euasarum clade that exhibits stamen movement. Vogel (1978) reported that *A. caudatum*, a western North American species, is outcrossing and deceives fungus gnats through floral mimicry of basidiomycete
sporocarps. He observed that male fungus gnats picked up pollen during copulation inside of the flower, and through subsequent copulations in other flowers, transferred pollen. Lu (1982) and Otte (1977) found a lack of support for frequent outcrossing in *Asarum*, with the former finding only 8% of female stage flowers with pollen on the stigma in representatives of the Euasarum clade, and the latter identifying a loose assemblage of leaf litter inhabitants as potential pollinators of the Hexastylis clade. It is important to note that Lu (1982) studied *A. caudatum* Lindl., a western North American member of the delayed-autonomously selfing *Asarum* clade; it is to be expected that little cross-pollination would be observed in non-herkogamous flowers with delayed autonomous self-pollination and without scent, though the great majority of pollination studies in *Asarum* have focused on such species. In the most conclusive investigation of pollination biology in a herkogamous species of *Asarum*, Sugawara (1988) found that fungus gnats (Mycetophilidae) not only visited and transported pollen between the flowers of *Asarum tamaense* (*Heterotropa tamaensis* (Makino) F.Maek., major *Heterotropa* clade) by way of hairs on their bodies, but also layed eggs in the reticulations at the base of the calyx in 23 - 46 % of flowers sampled. Even though empirical evidence of the potential for deceit pollination in only one species of the Heterotropa clade has been published, the flowers of this clade bear a remarkable resemblance to those of known brood-site deceptive species found throughout the angiosperms (see Jürgens et al., 2013 and Urru et al., 2011).

The flowers of the Heterotropa clade have raised longitudinal and transverse ridges that highly sculpture the interior of their calyces (Figure 8 F, I, and L). These
reticulations may be convergent on the gills of basidiomycete sporocarps (Vogel, 1978; Sugawara, 1988; Leins and Erbar, 2010). The outwardly smooth and inwardly sculptured calyces of the Heterotropa clade flowers and the gynostemium (the gynoecium with adnate stamens) may appear to fungus gnats or other dipterans to be the cap and stipe of the mimicked fungus, respectively (Figure 8C, F, I, and L). The flowers of the Euasarum and Geotaenium clades are of a similar bauplan, although their interiors are not highly sculptured (Figure 8C), the sepals do not greatly surpass the reproductive column as in the flowers of the Heterotropa clade (contrast Figure 8C with F, I and L), and their outside surface is nearly covered in long hairs (Figure 8D) that may be an obstacle to pollinators. The loss of exterior pubescence and delayed autonomous self-pollination coupled with the variation in calyx morphology between and within *Asarum* clades may be indicative of pollination shifts between lineages.

We speculate that the often limited geographic ranges of the species in the Heterotropa clade, and their small populations (see section 1.4.3), may predispose them to influence by spatial heterogeneity of pollinator service; Johnson and Bond (1992) and Kudo and Kasagi (2005) have shown that pollinator service can differ at the scale of tens to hundreds of meters, meaning that the pollinator mosaic can be of a finer scale than we often assume. Highly variable seed-set between populations comprising a single species has been observed during the course of our own fieldwork in the southeastern US, and highly variable seed-set has been noted between species (Otte, 1977; Jones et al., 2014); this suggests heterogeneity in pollinator service, possibly due to differences in floral morphology or nearby resources. Since the putative pollinators of these flowers are
visiting them due to their resemblance of fungi, and not due to floral rewards such as nectar, the presence of nearby resources (e.g. actual fungal brood sites) may influence the pollination of _Asarum_ species; such is the case of other non-rewarding flowers (Laverty, 1992; Johnson et al., 2003). However, Otte (1977) proposed that differences in floral morphology are responsible for asymmetrical seed set in the sympatric species _A._ arifolium and _A._ minus, where she found that seed set was 30% higher in the more trap-like flower of _A._ arifolium. Together these factors mean that _Asarum_ species that we regard as sympatric may overlap more in geography than in their interactions with pollinators, and it is conceivable that sympatric _Asarum_ species with highly divergent floral morphologies (see section 1.4.1) and different flowering periods, such as _A._ arifolium var. _arifolium_ and _A._ shuttleworthii var. _shuttleworthii_ (Figures 8H – M), may not share pollinators although they are commonly found in close proximity.

Investigation of the reproductive biology of the Hexastylis clade is currently underway; our aim here is not the demonstrated function of floral characters of presumed pollinator importance, but rather the inference of their importance from an evolutionary standpoint by studying their covariance with branch lengths, and by addressing their constancy and distribution throughout lineages. Evidence of tight-knit pollinator associations is lacking in _Asarum_ (Otte, 1977; Lu, 1982; Sugawara, 1988), but we find it difficult to surmise how else the association of increased lineage diversification rate and the characters tested here should be interpreted.
Vegetative morphology and dispersal

The vast majority of species in the *Asarum* clade have long internodes with pubescent, paired marcescent or deciduous leaves. These plants can produce large clonal mats, of which a single individual may occupy several square meters. In contrast, most species in the Heterotropa clade produce a single evergreen leaf per node and, due to shortened internodes, form clumps. Few species in the Heterotropa clade have long internodes and grow in large mats; only two species of the Hexastylis clade grow in this manner, and both are found at low elevation along rivers or swamps (Gaddy, 1987b).

We hypothesize that broadly distributed, mat-forming species may have a lower speciation rate due to clonal propagation through increased internode length and subsequent autonomous self-pollination, features that have been shown to limit the potential success of rare floral morphs (Rymer et al., 2010) and to limit the cladogenic potential of lineages (Igic and Busch, 2013; Wright et al., 2013). Interestingly, the species that occupy the largest geographic areas are members of the Euasarum clade; this might be due to their long internodes enabling a single individual to cover a large portion of the forest floor. Our data support the observation that lineages within the Heterotropa clade have produced more species relative to the Euasarum clade, while they occupy much smaller geographic areas, and are rarely mat-forming (Blomquist, 1957; Gaddy, 1987a; Shumei et al., 2003; Sugawara, 2006). *Asarum* seeds are ant dispersed (Gaddy, 1986; Lengyel et al., 2010; Turner and Frederickson, 2013), and we hypothesize that rare morphs in the Heterotropa clade may be more likely to persist as they can be more easily removed from direct spatial-, nutrient-, and pollinator-resource competition imposed by
nearby common floral morph relatives than the more common clonal species of clades Euasarum or Geotaenium. The strong dissimilarities in floral morphology and diversification rate between the Euasarum, Geotaenium, and Heterotropa clades may be indicative of one or more pollinator shifts; for example, the interior sculpturing of the calyces of flowers of the Heterotropa clade, which represents putative floral mimicry of basidiomycete sporocarp gills, is absent from the flowers of the Euasarum and Geotaenium clades.

Tremblay et al. (2005) posit that diversification need not be coupled with enhanced fruit set or increased vegetative growth. In fact, they state that pollen competition or reduced physiological constraint of pollinator shifts relative to the development of selfing has likely driven the dazzling diversity of floral displays, and the accompanying species richness, now seen in the orchids. Furthermore, Vamosi et al. (2006) suggest that small population size and pollen limitation can be drivers of biodiversity, both of which are exemplified by many species in Asarum.

Dispersal by ant species, such as Aphaenogaster rudis (Giladi, 2004; Ness et al., 2009; Canner et al., 2012;), whose movements are largely dictated by temperature and humidity changes (Smallwood, 1982; McGlynn et al., 2004; Warren et al., 2011), may have frequently resulted in movement of Asarum into novel areas with subsequent abandonment after climatic shifts that excluded the disperser but were not severe enough to cause extirpation of the dispersed plant taxa. Asarum species can thrive far removed from their native ranges, as evidenced by their horticultural use, and it has been shown experimentally that the realized niche of species in the Hexastylis clade is constrained by
ant dispersers (Warren et al., 2010). We hypothesize that the pseudo-random dispersal of seeds coupled with reduced ability to spread vegetatively and the derivation of a herkogamous condition may have contributed to a diversity of localized genotypes in the Heterotropa clade relative to the more clonal, delayed-autonomously selfing, less florally complex taxa in the Euasarum clade. The variation generated through this post-dispersal abandonment may have produced much fodder for myriad processes, such as founder-induced speciation, disruptive selection, competitive release, or any combination thereof (Rundle et al., 1998; Mittelbach et al., 2007; Rymer et al., 2010; Vamosi and Vamosi, 2010; Van der Niet et al., 2014b) that could account for the observed taxonomic richness in the Heterotropa clade.

**Methodological caveats**

The characters coded here represent a subset of many possibly relevant characters; the testing of all of these characters would be inappropriate as some characters have identical character plottings. Only characters that were unique in character state distribution were tested. Any morphological character distributed in like fashion, even one which has not been identified yet, might be responsible for the detected increase in diversification. The lack of knowledge of the totality of molecular and morphological characters is an issue with any ecological or phylogenetic study.

The BiSSE method assumes a completely sampled phylogeny (Maddison et al., 2007). The phylogeny reported here is not complete; however, we sample evenly throughout the genus with the exception of the Braun clade, the group which has
diversified at the highest rate. Due to the reduced sampling of this clade, our results and interpretation are most likely conservative. The inclusion of additional taxa from the Braun clade would undoubtedly reinforce and provide additional significance to our findings, since 46 species from this clade are currently recognized from Japan alone (Sugawara, 2006).

Davis et al. (2013) warn that BiSSE analyses of trees containing fewer than 300 terminals are prone to Type II error, though published studies have analyzed many fewer species (Johnson et al., 2011; Valente, et al., 2011; Blanco-Pastor and Vargas, 2013; Schneider et al., 2013). The majority of p-values reported here are highly significant; this suggests that the phylogenies analyzed did not unduly challenge the BiSSE method.

**Conclusions**

We show that character-associated diversification is strongly evident within *Asarum*. Diversification rate has increased in the Euasarum clade relative to the Geotaenium clade and in the Heterotropa clade relative to the Euasarum clade. Furthermore, we propose that the loss of delayed autonomous self-pollination, connation of sepals into an elongate floral tube, shift from an inferior to superior ovary, and reduction of vegetative growth represent innovations that have facilitated increased rates of diversification.

We have no *a priori* reason to believe that sepal connation, internode length, the number of leaves per node, ovary position, and the herkogamous condition are genetically linked traits, yet they may be mutually advantageous and concurrently
selected for. The correspondence of some character states such as reduced internode length or connate calyces is more readily explained than the shift from inferior to superior ovaries; the only explanation of ovary position that we can offer is possible convergence of the gynostemium on a fungal stipe. However, the strong co-occurrence of the features studied here aligns well with rapid speciation due to the development of an increasingly basidiomycete sporocarp-convergent floral brood-site deception syndrome and an accompanying reduction in vegetative growth. The linking of putative key-innovations, their biotic and/or abiotic interfaces, and determining whether they have been shaped by selective regimes represents the ultimate goal of our research; this work forms the foundation on which future questions of pollination biology and floral evolution in Asarum can be rooted.
Chapter 3. *Asarum chueyi* (Aristolochiaceae), a new species from the foothills of the Blue Ridge Mountains of Tennessee and Virginia, USA.

Introduction

The genus *Asarum* (Aristolochiaceae) comprises approximately 100 species distributed in both the Old and New Worlds (Cheng and Yang, 1983; Huang et al., 1995), 17 of which are found in North America (Gaddy, 1987; Whittemore and Gaddy, 1997, Whittemore et al., 1997; Kelly, 1997, 1998, 2001). Our own work (Sinn et al., 2015a) and that of others has shown (Kelly, 1997, 1998; Niedenberger, 2010) that the North American members of *Asarum* do not form a monophyletic group, but instead comprise two separate monophyletic groups. North American *Asarum* species with separate sepals, two deciduous leaves per node, and flowers that are autonomously self-pollinated via delayed stamen movement are found on both the east and west coasts, and form a clade that falls within *Asarum* subgenus *Asarum*. North American taxa that produce a single evergreen leaf per node and herkogamous flowers with connate sepals are restricted to the eastern portion of the continent and have been placed within *Asarum* subgenus *Heterotropa* (Kelly, 1997; 1998) section *Hexastylis* (Araki, 1937). Species of subgenus *Heterotropa* have showy, complex calyces that may mimic the sporocarps of basidiomycete fungi (Vogel, 1978; Lu, 1982; Sugawara, 1988; Leins and Erbar, 2010) and appear to have influenced the diversification of this subgenus (Sinn et al., 2015b).
The North American members of this subgenus remained under-collected and under-described, most obviously due to their often restricted ranges, low growth habit, early flowering period, and the difficulty of working with the deformed flowers of pressed material (Ashe, 1897; Weakley, 2012).

Several populations of difficult to identify herbarium specimens were located during fieldwork completed during the springs of 2012 and 2013; these populations represent a species new to science, *Asarum chueyi* B.T.Sinn. Additional *A. chueyi* localities in the foothills of the Blue Ridge Mountains of Virginia and Tennessee were discovered during the spring of 2014. This new species is described herein, and is distinguished from sympatric and morphologically similar species.

**Materials and Methods**

The holdings of BOON, CLEMS, FLAS, GA, GH, MO, MICH, NCSC, NY, PH, TENN, US, WCUH, WVA, and YUO were studied. Keys and descriptions published by Blomquist (1957), Gaddy (1987, 2011), Whittemore and Gaddy (1997), and Weakley (2012) were compared. Species discussed here are treated as hypotheses, and I use the Phylogenetic Species Concept per Nixon and Wheeler (1990) which defines species as, “the smallest aggregation of populations (sexual) or lineages (asexual) diagnosable by a unique combination of character states in comparable individuals (semaphoronts).”

Elliptical Fourier analysis (EFA; Giardina and Kuhl, 1977; Kuhl and Giardina, 1982) of silhouettes of calyx tube shape was conducted using SHAPE version 1.3 (Iwata and Ukai, 2002). Silhouettes were generated by digital photography of 42 field-preserved
flowers fixed in FAA solution (47% ethanol – 43% water – 5% acetic acid – 5% formaldehyde solution (37% formaldehyde – 10% methanol – 53% water) and stored in 70% ethanol. These accessions represent three described species, *Asarum heterophyllum* Ashe (1897: 33), *A. minus* Ashe (1897: 33), and *A. contractum* (Blomquist) Barringer (1993: 226), and *A. chueyi* described here, from a total of 20 populations throughout their ranges (Table 7).
### Table 7. Accessions analyzed by EFA, NMDS, and UPGMA.

<table>
<thead>
<tr>
<th>Species</th>
<th>Collector</th>
<th>Collector #</th>
<th>County</th>
<th>State</th>
<th>Stamen Filament Coding</th>
<th>Volume 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. chueyi</em></td>
<td>BTS</td>
<td>1190 B</td>
<td>Greene</td>
<td>Tennessee</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>A. chueyi</em></td>
<td>BTS</td>
<td>1111 A&amp;B</td>
<td>Albemarle</td>
<td>Virginia</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>A. chueyi</em></td>
<td>BTS</td>
<td>11112 B</td>
<td>Albemarle</td>
<td>Virginia</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>A. chueyi</em></td>
<td>BTS</td>
<td>1113 A</td>
<td>Albemarle</td>
<td>Virginia</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>A. chueyi</em></td>
<td>BTS</td>
<td>1195 B</td>
<td>Amherst</td>
<td>Virginia</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>A. contractum</em></td>
<td>BTS</td>
<td>1089 B-D</td>
<td>Caldwell</td>
<td>North Carolina</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>A. contractum</em></td>
<td>JVF</td>
<td>2722 A-C</td>
<td>Van Buren</td>
<td>Tennessee</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>A. heterophyllum</em></td>
<td>BTS</td>
<td>1169</td>
<td>Rabun</td>
<td>Georgia</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>A. heterophyllum</em></td>
<td>BTS</td>
<td>1075 A-F</td>
<td>Oconee</td>
<td>South Carolina</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>A. heterophyllum</em></td>
<td>BTS</td>
<td>1076 A-D</td>
<td>Oconee</td>
<td>South Carolina</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>A. heterophyllum</em></td>
<td>BTS</td>
<td>1077 A-F</td>
<td>Oconee</td>
<td>South Carolina</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>A. heterophyllum</em></td>
<td>BTS</td>
<td>1094 A&amp;B</td>
<td>Oconee</td>
<td>South Carolina</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>A. heterophyllum</em></td>
<td>BTS</td>
<td>1095 C</td>
<td>Oconee</td>
<td>South Carolina</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>A. heterophyllum</em></td>
<td>BTS</td>
<td>1145</td>
<td>Oconee</td>
<td>South Carolina</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>A. heterophyllum</em></td>
<td>BTS</td>
<td>1064 A-C</td>
<td>Unicoi</td>
<td>Tennessee</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>A. heterophyllum</em></td>
<td>BTS</td>
<td>1199</td>
<td>Nelson</td>
<td>Virginia</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>A. minus</em></td>
<td>BTS</td>
<td>1082 A&amp;C</td>
<td>Montgomery</td>
<td>North Carolina</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>A. minus</em></td>
<td>BTS</td>
<td>1060</td>
<td>Orange</td>
<td>North Carolina</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>A. minus</em></td>
<td>BTS</td>
<td>1061 A</td>
<td>Orange</td>
<td>North Carolina</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>A. minus</em></td>
<td>BTS</td>
<td>1112 A&amp;C</td>
<td>Albemarle</td>
<td>Virginia</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
The sepal lobes were cropped from the silhouettes such that only the calyx tube itself was analyzed by EFA (Figure 9). The EFA shape descriptors for 19 harmonics, corresponding to the 2nd through 20th harmonics, and two stamen characters (Table 7) were analyzed using Nonmetric Multidimensional Scaling (NMDS; Shepard, 1962) and the unweighted pair-group method using arithmetic averages (UPGMA, Sneath and Sokal, 1973). The 1st harmonic, which describes the fit of an ellipse to the shape of interest, was excluded from both analyses since it was used to standardize the orientation of all shapes and to cancel out the effects of floral size. PAST3 (Hammer, 2014) was used for all statistical analyses, using the mixed distance function for each analysis, where Chord distance was used for EFA descriptors and Jaccard distance was used to analyze binary character coding of stamen morphology.

Figure 9. Example of an *A. chueyi* (1195) calyx tube silhouette, with the sepal lobes cropped, used for Elliptical Fourier Analysis. The proximal and distal portions of the flower are to the left and right, respectively.
Results

NMDS of EFA shape descriptors resulted in the structuring of ordination space into four general groups (Figure 10) corresponding to *Asarum minus*, *A. heterophyllum*, *A. contractum*, and a new taxon, *A. chueyi* (Figure 11). Coordinates 1 and 2 had $R^2$ values of 0.84 and 0.08, respectively. A Shepard plot (Shepard, 1962) and a stress value of 0.16, a measure of data fit, suggest that the ordination represents a good explanation of the data.

![Figure 10. Ordination represented by coordinates 1 and 2 resulting from Nonmetric Multidimensional Scaling of Elliptical Fourier shape descriptors of calyx tube shape of *A. contractum* (+), *A. chueyi* (circles), *A. heterophyllum* (triangles) and *A. minus* (squares). The biplot depicts the impact of stamen filament morphology on data point spread.](image-url)
Figure 11. Illustrations of *A. chueyi* showing A) a whole plant (prepared from the holotype; scale represents 6.5 cm), and B) a flower with a single sepal removed to reveal the reproductive column, C) a dissected sepal, and D) an entire flower (each prepared from field-preserved flowers from Tennessee and Virginia; scale represents 2.3 cm).

Illustrations prepared by Ryan A. Folk.
UPGMA analysis recovered three major clusters with strong internal structuring which correspond to *A. contractum*, *A. chueyi*, and *A. minus* plus all but one individual sample of *A. heterophyllum* (Figure 12); the distance between these groups ranged from approximately 0.1 to 0.3. *Asarum chueyi* and *A. contractum* OTUs were more similar to their hypothesized conspecifics than they were to either *A. heterophyllum* or *A. minus*. The recovery of the *A. minus* OTUs 1082 A, 1112 A, and 1112 C with *A. heterophyllum* clusters is not supported and highly sensitive to matrix manipulation through bootstrapping of EFA descriptors alone. However, the bootstrapping of matrixes analyzed through mixed distance matrices is not supported in PAST3, and support of these nodes in our final combined EFA and stamen morphology analysis can only be inferred by the bootstrapping of EFA descriptors analyzed singly.
Figure 12. UPGMA phenogram of stamen filament characters and EFA shape descriptors of calyx tube shape of the flowers of *A. contractum* (red), *A. chueyi* (black), *A. heterophyllum* (green), and *A. minus* (pink).

**Discussion**

Elliptical Fourier analysis is a powerful tool for the mathematical description of shapes that are made up of curves, and is especially appropriate when determining the homology or placement of landmarks or semi-landmarks is non-trivial; such is the case for the calyx tube of *Asarum* species. EFA characterizes closed forms by the decomposition of a complex shape into component waves. Shapes of interest, in this
study the calyx tubes of *Asarum* flowers, are mathematically characterized using X and Y sine and cosine projections from the centroid of an ellipse fitted to the image. Elliptical shapes better fit the studied shape, and describe increasingly finer details of that shape, as the number of projections used increases; each set of projections is termed a harmonic (Macleod, 2012). This method can be viewed as a modification of Radial Fourier analysis (Schwarcz and Shane, 1969).

Elliptical Fourier Analysis of calyx tube shape, UPGMA analysis of EFA shape descriptors and stamen characters, and unique combinations of other morphological characters support the recognition of a new species, *A. chueyi* (Figures 10 & 12). The calyx tube of this new species occupies a position in morphospace between *A. contractum* and *A. minus*, and shows some similarity to *A. heterophyllum* individuals with flared calyx tubes; however, the calyx tube flare in *A. heterophyllum* is more distal than that of *A. chueyi*. Even though some overlap in morphospace between species is evident, calyx tube shape is but one of a combination of morphological characters that defines this taxon. It is critical to use only the flowers of robust, mature plants from *Asarum* for morphological study, as young plants of the genus may produce flowers with calyx tubes that vary greatly, even within a single plant. It is especially important to use only pollen-receptive *Asarum* flowers for morphometric analysis, as the calyx tube grows and becomes distorted after the flowers are successfully pollinated; this can influence the amount of “variation” that one finds in populations. It is also important to note that abortive or abnormally-formed floral organs can often be found within populations; this was especially the case in the Amherst Co., Virginia (B.T.Sinn 1195) locality where
intermediately round/strap-like stamen filaments and abnormal calyx lobe numbers and stamen extension morphology were observed in the field – sepal lobe number variation has also been observed in wild populations of what may be interpreted to be *A. memmingeri* in the Blue Ridge of North Carolina and *A. minus* in central portion of that same state (personal observation).

The calyx tube of *A. chueyi* is broadest at or just above the middle in contrast to that of *A. contractum*, which is broadest below the middle. The calyx of *A. chueyi* becomes much constricted distally, separating these flowers from those of *A. minus* and *A. sorriei* (new combination provided in Appendix C), which have broad openings (see Gaddy, 2011, Figure 4). *Asarum heterophyllum* sometimes has flared flowers, yet the protuberance is directly below the floral orifice, and the width of the floral orifice is equal to that of the middle of the floral tube. Even though calyx tube shape can be a useful and easy to observe character to delineate *Asarum* species, few section *Hexastylis* species can be reliably keyed based on this character alone, and the totality of reproductive features should be carefully studied prior to making a species diagnosis. Variation in calyx tube shape is seen between *A. chueyi* populations in Tennessee and Virginia; however, the tube is always inflated about the middle and the floral orifice is always highly constricted (Figure 13). In *A. chueyi*, the sepal lobes are as long as or longer than they are broad at the base, much like those of *A. heterophyllum*. This is in contrast to the sepals of *A. minus* where they are approximately the same length as their widest point.
Figure 13. Photographs of *A. chueyi* flowers from Tennessee (A-C) and Virginia (D and E) populations. Typical flower with one sepal removed (A, D), entire flower (B, E) and floral orifice (C). Note that the thecae have not dehisced in A. Scale is not maintained.
The adaxial, proximal portion of the calyx of *A. chueyi* is of high relief, with net-like reticulations that form irregularly shaped pits that are not as deep as those of *A. minus*, which sometimes is found in sympathy with *A. chueyi* in Virginia. In *A. chueyi*, this sculpturing ends abruptly, unlike the gradual smoothing of the sometimes, though weakly so, reticulated portion of the calyx of *A. contractum*. The transverse and longitudinal lines forming the reticulations within the calyx of *A. contractum* are thinner and of lower relief compared with *A. chueyi*. Furthermore, the transverse and longitudinal components are nearly equal in height, in contrast to the much more pronounced longitudinal ridges of *A. heterophyllum*; see Blomquist (1957) Figures 19 - 24 for detailed photographs of these features. The style extensions of *A. chueyi* are bifid and shorter than those of *A. minus* or *A. heterophyllum*; in the area of the type locality these extensions can be so little developed that the stigmas are nearly terminal.

Large-flowered *A. minus* has been found growing sympatrically with *A. chueyi* in areas surrounding the type locality. While the calyx of *A. minus* does flare, it does so past the middle, and it does not constrict thereafter to a degree that equals the diameter of the most proximal portion of the tube, as is the case in *A. chueyi*. Additionally, the style extensions of *A. minus* are approximately double the length of the extensions of *A. chueyi*. While the flowers of these northern *A. minus* individuals are larger than those of the same species from central North Carolina, it is evident from multivariate analysis of Elliptical Fourier shape descriptors that the shape of their calyces is no different, and
furthermore, that neither their flowers, nor those of *A. contractum* sampled elsewhere, are similar in overall calyx tube shape to the flowers of *A. chueyi* (Figure 10).

Some might postulate that the close proximity in which *A. chueyi* and *A. minus* occur points to a hybrid origin of *A. chueyi* between *A. contractum* and *A. heterophyllum* or *A. minus*. However, clearly delineated species can often be found at the same site, as it is often the case that *A. arifolium* var. *arifolium* or *A. arifolium* var. *ruthii* can be found growing in close proximity to many southeastern *Asarum* species. In northern Georgia, *A. heterophyllum*, *A. arifolium* var. *arifolium*, and *A. shuttleworthii* var. *shuttleworthii* can be found growing in an area of only 100 m² (personal observation). To my knowledge, populations near the type locality of *A. chueyi* that contain both *A. minus* and the former species, along with the reference above to sites in northern Georgia and South Carolina, represent the first published accounts of commonly misidentified section *Hexastylis* species found in sympatry.

Blomquist (1957) published two drawings of *A. contractum* (*Hexastylis contracta* Blomquist) in the only monographic work of section *Hexastylis*. This species is the only one for which two flower shapes were depicted. All available liquid preserved flowers that I have had access to, as well as populations visited in the field and all herbarium material identified as *A. contractum*, other than that identified herein as *A. chueyi*, are consistent with his Figure 7. Blomquist’s (1957) Figure 8 is more similar to *A. chueyi*, though no material that was examined by Blomquist that I have seen appears to be of this species. In his description of *A. contractum*, he does not discuss Figure 8, but instead points to Figures 6 and 7, the former of which is an illustration of *A. minus*. Blomquist
does not discuss Figure 8. Gaddy (1986) noted in his description of *H. rhombiformis*, and illustrated in his Figure 2, that *A. contractum* calyces are broadest below their midpoint. This species has been included in *A. contractum* by Barringer (1993), yet the features distinguishing these species are highly consistent.

The *A. chueyi* holotype designated here was annotated by Gaddy in 1981 as *H. contracta* Blomquist (1957: 279), upon which he provided the annotation: “First record from VA!” However, it is clear that this specimen was not representative of what he expected from that region. The range expansion suggested by such a find would have proven to be a significant contribution of his 1987 review. The discovery of this specimen was not published, but Gaddy did write at length of the unusual distribution of *A. contractum*, which is disjunct between the Blue Ridge Escarpment in Tennessee and North Carolina. It is possible that Gaddy was not sufficiently sure of his identification to publish an expansion in range for this species or that this annotation was simply lost since it is dated five years prior to the publication of his review. Weakley (2012) reports that *A. contractum* populations with a high degree of calyx tube reticulation can be found in the panhandle of Virginia; it would not be surprising if such populations were found to fall within *A. chueyi*, as some vegetative herbarium specimens that look remarkably similar to this species have also been seen from this same region.

The Tennessee population noted here is especially interesting, as the flowers of plants which came to be identified as *A. chueyi* from that locality were just opening when the stamens of other nearby plants that were keyed as *A. heterophyllum* had long dehisced. It would be unsurprising if future work investigating these and other
populations, and especially that which integrates molecular and morphological data, suggest that the variability often noted in populations of species in section *Hexastylis* is the result of mixed aggregations of the decedents of multiple lineages rather than species harboring extensive polymorphism in floral structure. I note again that many localities exist in which section *Hexastylis* species with remarkably different floral character suites exist in sympatry. Why should it be the case that section *Hexastylis* species with similar flowers cannot be found in sympatry? It is my hope that this description spurs interest in and discussion of species boundaries in this charismatic group of plants.

**Taxonomy**

*Asarum chueyi* B.T.Sinn, sp. nov. (Figure 11)

**Diagnosis** – Similar to both *A. contractum* (Blomquist) Barringer and *A. heterophyllum* Ashe, though notable for fusiform calyx tube broadest at its midpoint, highly developed reticulations forming irregular pits in the adaxial, proximal portion of the calyx tube, and poorly developed style extensions. Sepal lobes erect to only weakly spreading, as long as or longer than their width at their widest point. Style extensions shorter than the length of the stigmatic surfaces. Stamen filaments cylindrical and tapering distally causing the distal ends of the thecae to be nearer one another than the proximal ends; thecae latrorse or nearly so. Leaves less variegated and more oblong-cordate in shape than those of *A. minus*. 
**Type:**— UNITED STATES. Virginia: Albemarle County, limestone cliffs along Ravanna River at Route 29. 16 April, 1959. *Clyde F. Reed 42803.* (Holotype: MO!)

**Description:**— *Perennial herb,* clumped growth habit with short rhizome. *Petioles* glabrous, (8.5-) 9.5-13.5 cm long. *Leaf blades* minutely revolute, entire, and glabrous, (5.5-) 6.5-8.5 cm long from base of lobe to the apex of the acute leaf tip and (4-) 5.5-8 cm wide, roundly-hastate to oblong-cordate in shape, variegation originating along distal portion of primary veins and quickly fanning broadly along secondary veins; *leaf lobes* 1.5-2.5 cm long and 2-3 cm wide; *Calyx tube* 2-2.5 cm long, 0.5 cm wide at the orifice, fusiform and broadest (~1 cm) at its mid-point, deep maroon to purple with the exception of pale yellow base, red-villous within, trichomes ending abruptly at the start of the highly reticulate proximal portion of the tube; *calyx reticulations* composed of strong transverse and longitudinal lines of equal height, the strongest of which are not equal in height to the thickness of the calyx wall, and anastomosing to form deep irregular pits proximally; *sepal lobes* 1-1.5 cm long, erect to weakly spreading, white mottled, as long or longer than they are broad, with long-tomentose covering of uniformly red trichomes over abaxial surface with the exception of white areas; *stamens* latrorse to nearly so, thecae 1.5-2 mm long and 0.5 mm wide, filament shorter than the styles, with cylindrical filaments (2.5 – 3 mm long) tapering and surpassing the thecae by approximately 1 mm, the distal apex of the thecae nearer to one another than the proximal ends; *ovary* approximately 2/3 superior; *styles* white mottled with purple at base and gradually
becoming entirely purple at the poorly developed, deeply cleft extensions that surpass the ovoid, overhanging stigmas by 1-2 mm; stigmas approximately 1 mm long and 0.5 mm wide. Fruit and seeds not seen.

**Phenology:**— Flowering mid April – early May, fruiting not observed.

**Distribution:**— Foothills of the Blue Ridge Mountains in Tennessee and Virginia.

**Habitat:**— Shaded slopes with exposed rock near streams and rivers.

**Associates:**— *Quercus* sp., *Fagus grandifolia*, and *Kalmia angustifolia*.

**Conservation:**— Careful fieldwork is needed to document the occurrence of this taxon to the exclusion of two common species, *A. minus* and *A. heterophyllum*, which have been found growing in sympatry. Surveys for additional populations of this taxon should be conducted in order to better assess its distribution.

**Etymology:**— Named for Carl F. Chuey, a 43-year faculty member of Youngstown State University who botanized extensively throughout the United States, and worked tirelessly to establish a herbarium at his home institution. Perhaps most importantly, he served as a thoughtful and dedicated mentor to many first-generation college graduates, myself included.
Other specimens examined:— USA: Virginia: Albemarle Co., 16 April, 1959, C.F. Reed 42795 (MO); Albemarle Co., 27 April, 2013, B.T. Sinn 1148 (OS); Albemarle Co., 27 April, 2013, B.T. Sinn 1164 (OS); Albemarle Co., 27 April, 2013, B.T. Sinn 1166 (OS); Amherst Co., 21 April, 2014, B.T. Sinn 1195 (OS); Nelson Co., 19 April 1975, C.F. Reed 99775 (MO); Nelson Co., 19 April, 1975, C.F. Reed 99773 (MO). Tennessee: Greene Co., 20 April, 2014, B.T. Sinn 1190 (OS).

Artificial Key to species of Blomquist’s informal Virginica Group

1A. Trichomes of sepals and floral orifice pilose and exclusively white.

   *A. lewisii* Fernald. Leaf scars widely spaced and internodes long, plants forming mats; calyx tube broadly rhomboid, widest below the middle, tapering and becoming more constricted toward the orifice; leaves broadly hastate. VA, NC.

1B. Trichomes of sepals and floral orifice comprised of red or red and white cells … 2

2A. Calyx tube 3 cm or less in length, thick-walled and relatively robust, often odorless; ovary often somewhat inferior, the point of stamen insertion on the gynostemium no further from the base of the calyx tube than the height of the thecae … 3
3A. Stamen filaments broad and strap-like; thecae extrorse … 4

4A. Flower orifice not often highly constricted, or if constricted the orifice is not more than 5 mm from the widest portion of the calyx tube … 5

5A. Calyx tube greatly flared at the midpoint or higher, the floral orifice only slightly more constricted than the widest portion of the calyx tube; sepal lobes much broader than long; leaves usually variegated … 6

6A. Floral orifice nearly as wide as the interior of the calyx tube flare; calyx tube reticulations forming irregularly-shaped distally-facing pits, the longitudinal components only equaling the thickness of the calyx tube in height distally, if ever; leaves often heavily variegated.

_ A. minus _ Ashe. NC, SC, VA.

6B. Plants as above but with weakly variegated leaves and of frequently disturbed areas in the Sandhills region of NC and SC.

_ A. sorriei _ (Gaddy) B.T.Sinn.
5B. Calyx tube often flared above the middle, the floral orifice is of approximately the same width as the unflared portion of the calyx tube; sepal lobes as long or longer than their width at base; leaves often not variegated … 7

7A. Longitudinal and transverse components of the calyx tube sculpturing of approximately equal relief, the transverse components anastomosing and often not spanning the longitudinal portions forming irregularly shaped pits

*A. virginicum* L. Calyx lobes ± erect, much broader than long; leaves reniform to obovate-cordate, rarely variegated. NC, MD, WV, VA. **Note:** Plants of higher elevations in NC, WV, and VA with urceolate to cylindrical calyx tubes that are widest above the midpoint and narrowly urceolate have been interpreted as *A. memmingeri.*

7B. Longitudinal components of the calyx tube sculpturing of greater relief than the transverse components … 8

8A. Style extensions longer than the width of the stigmatic surfaces; stigmas dorsal and overhanging; calyx tube at least 1 cm in length.

*A. heterophyllum* Ashe. Calyx tube cylindrical, and sometimes broadly so, usually encircled by a flare distally (except in the SE portion of the range); sculpturing of the calyx tube highly developed, especially the
longitudinal components; some large-flowered southern populations with slight odor. AL, GA, KY, NC, SC, VA.

8B. Style extensions as long as or shorter than the width of the stigmatic surfaces; stigmas nearly terminal and in notched recess; flowers extremely small.

*Apocynum naniflorum* (H.L.Blomq.) B.T.Sinn. Calyx tube often approximately 1 cm in length; calyx tube cylindrical and slightly tapering or expanding toward orifice; sculpturing of the calyx tube poorly developed, often only the longitudinal components present. NC, SC.

4B. Floral orifice highly constricted, greater than 5 cm from widest point of calyx tube

*Apocynum rhombiforme* (Gaddy) B.T.Sinn. Interior sculpturing of calyx tube forming highly irregularly-shaped transversely-facing open pits, the longitudinal and lateral components consistently more than twice the thickness of the calyx in height. NC, SC.

3B. Stamen filaments cylindrical; thecae latrorse or nearly so … 9

9A. Stamen filaments not strongly tapering distally, proximal and distal portion of the thecae about equally spaced; calyx tube broadest below the middle.
*A. contractum* (H.L.Blomq.) Barringer. Sepal lobe length at most equal to the width of their widest point; abaxial surface of calyx often not sculptured or only weakly so, the broadly spaced longitudinal lines often not completely connected by the meandering transverse components which form distally-facing pits; leaves broadly reniform. TN, NC.

**9B.** Stamen filaments strongly tapering distally, thecae touching or nearly so distally; calyx fusiform and broadest at the midpoint.

* *A. chueyi* B.T.Sinn. Sepal lobes as long as or longer than the width of their widest point; abaxial proximal surface of calyx tube sculptured with longitudinal and transverse components of more or less equal relief forming transversely-facing pits; leaves broadly hastate. TN, VA.

**2B.** Calyx tube greater than 3 cm in length, thin-walled and relatively brittle, often with rancid odor; the point of stamen insertion on the gynostemium as far or further from the base of the calyx tube than the height of the thecae … 10

**10A.** Internodes short and leaf scars crowded on stem, plants not forming mats.

* *A. shuttleworthii* Britten & Baker var. *shuttleworthii*. GA, NC, SC, TN.
10B. Internodes long and leaf scars widely spaced on stem, plants forming mats.

*A. shuttleworthii* Britten & Baker *var. harperi* (Gaddy) Barringer. AL, GA.
Chapter 4: The first reported duplication of the Small Single Copy region in angiosperms: rearrangement, tRNA loss, IR boundary expansion, extreme regional AT-rich expansion, and episodic selection-regime shifts in *Asarum* plastomes.

**Introduction**

The majority of sequenced angiosperm plastid genomes (plastomes) are highly conserved (Guisinger et al., 2011; Wicke et al., 2011; Ruhlman and Jansen, 2014) with respect to their gene content, gene order (synteny), length, GC content, and are functionally tripartite due to the presence of a co-evolving, expansive inverted repeat (IR) regions (Knox, 2014) that separate the independently evolving Large Single Copy (LSC) and Small Single Copy (SSC) regions. A number of exceptional plastomes have been published in families of the Eudicots and Monocots (Chumley et al., 2006; Cai et al., 2008; Haberle et al., 2008; Magee et al., 2010; Guisinger et al., 2011; Sloan et al., 2014; Ruhlman and Jansen, 2014); however, rearranged or highly divergent plastomes have not been reported from the basal angiosperms (Cai et al., 2006; Li et al., 2013).

The magnoliids, a basal clade of angiosperms, comprise approximately 10,000 species (Stevens, 2001) in 19 families and four orders (Cai et al., 2006). In contrast to other basal angiosperm clades, the magnoliids comprise both herbaceous and woody
species (Isnard et al., 2012), as well as highly derived lineages characterized by suites of floral characters that facilitate tight-knit relationships with arthropod pollinators (Oelschlägel et al., 2015); these lineages have undergone radiation events (Isnard et al., 2011; Erkens et al., 2012; Sinn et al., 2015a). In spite of the unique taxonomic, vegetative, and floral richness of the magnoliids relative to other basal angiosperm lineages, they are commonly thought of as relictual holdouts rather than dynamic, derived entities (Friis et al., 2000; Albert et al., 2005). However, recent research reminds us that a basal phylogenetic position is not necessarily predictive of genomic conservatism (Rice et al., 2013).

Although the magnoliids represent the largest clade of basal angiosperms, only 20 magnoliid plastomes have been published (Cai et al., 2006; Kuang et al., 2011; Li et al., 2013). Relative to other angiosperm lineages, published magnoliid plastomes have generated little interest (Kuang et al., 2007), aside from their use in phylogenetic studies (Jansen et al., 2007; Moore et al., 2010; Soltis et al., 2011; Li et al., 2013; Ruhfel et al., 2014), as those sequenced to date are unrearranged and highly conserved in length, and gene and GC content (Cai et al., 2006; Li et al., 2013). Of the 625 complete land-plant plastomes databased by the National Center for Biotechnology Information, less than one percent are from the approximately 10,000 magnoliid species, and only five of these plastomes are from genera other than Magnolia, a woody, slowly evolving lineage (Kim et al., 2001). Further reinforcing a stereotype of the magnoliids, the only published mitochondrial genome from the clade, that of Liriodendron tulipifera, is the most conserved angiosperm mitochondrial genome known, and as a result has been termed a
“fossil genome” (Richardson et al., 2013). The recently published presence of character-associated diversification and the diversity of life history traits in the Piperales, and *Asarum* specifically, lead us to ask: Do 20 sequenced magnoliid plastomes provide a sound representation of the organellar genomes in the Piperales?

The diversification of the highly derived (Bliss et al., 2013) north-temperate magnoliid genus *Asarum* (Aristolochiaceae, subfamily Asaroideae (along with *Saruma*); ~110 species) has been driven by key vegetative and floral innovations (Sinn et al., 2015a), and the plastomes reported here provide additional evidence of these radiations. Early in our work, we found that magnoliid plastomes sequenced to date in fact do not adequately describe the plastome variability found within it and therefore answer the question posed above with a resounding “no.” In light of the surprising correspondence of the unexpected plastome rearrangement and asymmetrical lineage diversification within *Asarum* we ask: Do the plastomes of *Asarum* show evidence of release from purifying selection, and if so, is the topological distribution of differential selective regimes congruent with previously published differences in the diversification rate of these lineages (Sinn et al., 2015a)?

Divergent, yet functional, plastomes have been reported from eudicot and monocot families such as Campanulaceae (Haberle et al., 2008; Knox, 2014), Caryophyllaceae (Sloan et al., 2012), Ericaceae (Fajardo et al., 2013), Fabaceae (Magee et al., 2010; Gurdon and Maliga, 2014), Orchidaceae (Kim et al., 2014), and Geraniaceae (Chumley et al., 2006), suggesting either that divergence of the content and structure of angiosperm plastomes is common in derived lineages or is the result of acquired features
rather than physiological predisposition due to phylogenetic relationships. With only a single extant close relative, *Saruma henryi*, to serve as an outgroup, and a well-resolved phylogeny, the syntenic breaks, gene loss, rearrangement, and general disruption of *Asarum* plastomes promise to provide a backbone upon which to base hypotheses of generalizable mechanisms of plastome disruption. Lastly, we ask: Can generalizable precursors of plastome disruption and rearrangement be identified using *Asarum* and the polarity provided by a well-resolved phylogeny?

**Methods**

**DNA Extraction and Library Prep**

Total DNA extractions of one gram of leaf material were performed using the CTAB method (Doyle and Doyle, 1987). To reduce the concentration of residual proteins and RNase, phenol-chloroform cleanups were used, replacing phenol with phenol:chloroform:isoamyl alcohol (25:24:1). DNA concentration and purity were quantified using a Nanodrop 2000 spectrophotometer.

**DNA Sequencing and Data Handling**

Massively parallel sequencing (MPS) of all species was accomplished by multiplexing samples on multiple Illumina HiSeq and MiSeq lanes. MiSeq runs used version 3 chemistry to achieve paired-end reads up to 300 base pairs (bp) in length, and HiSeq sequencing produced 100 bp paired-end reads. Library preparation and DNA sequencing was conducted by the Molecular and Cellular Imaging Center of the Ohio
Agricultural Research and Development Center. Geneious version 7.1 (Drummond et al., 2011) was used for demultiplexing, mate-pair shuffling, and quality trimming of reads; reads were trimmed using an erroneous-nucleotide probability of $\alpha = 0.05$.

**Plastome Assembly and Annotation**

A combination of reference-based contig extension or “genome walking,” de novo assembly, and Sanger sequencing was used to assemble plastomes. Using the Geneious Read Mapper (Drummond et al., 2011), reads were mapped from each of the two species to the previously published *Piper cenocladum* plastome (Cai et al., 2006) using parameters that invoked only single bp gaps totaling up to 5% of a given mapped read. Mapped reads were then de novo assembled using the Geneious De Novo Assembler (Drummond et al., 2011); the resultant contigs were proofread and their consensus sequences were used as “seeds” for iterative reference-based assembly using the Geneious Read Mapper (Drummond et al., 2011) of all reads. Reads were mapped to all matching sites; this is particularly important in order for contigs to eventually overlap through iterative mapping. To avoid misassembly during contig extension, mapped reads were required to match perfectly for a minimum of 25 bp and the overall read divergence was capped at 1% of total read length. Extended contigs were inspected for accuracy every 25 iterations, and consensus sequences from each contig were used as input to the Geneious de novo assembler to combine contigs which overlapped from iterative mapping and extension. Extreme base pair compositional bias in several AT-rich regions prevented circularization of the plastome; primers (Appendix D) for Sanger sequencing
were developed in an attempt to bridge AT-rich regions that were either not represented in the sequencing libraries or were not recovered by Illumina sequencing, but this was largely unsuccessful.

The Live Annotation feature of Geneious was used for all plastome annotation, with like regions identified by a minimum of 70% similarity to published annotations of the *Piper cenocladum* plastome. The reading frame of each annotated region was visually inspected and hand-adjusted, preferring solutions that maintained open reading frames over those that resulted in pseudogenes.

**Repeat Detection**

Phobos version 3.3.12 (Mayer, 2006-2010) was used via the freely available Geneious plugin to discover, count, and annotate tandem repeats. Phobos uses a score which is reflective of the quality of a local alignment as an optimality criterion. The perfect search option was enabled, thereby identifying only perfect repeats, which were constrained to be a repeat unit length between 1 and 1000 base pairs and default score constraints; the “remove hidden repeats” setting was enabled so that repeats overlapping another repeat with a higher alignment score were not counted.

**Nuclear rDNA Array Assembly**

An artificial rDNA contig was constructed via genome walking in identical fashion to that described above for plastome construction. The initial seeds for this procedure were Sanger-sequenced contigs of ITS1-partial 28S stretches produced for
each of the species used in an earlier study (Sinn et al., 2015b). Genome walking continued until coverage dropped to a level in which ambiguous base calls needed to be made.

**Sequence Alignment, Matrix Construction and Partitioning Scheme Selection**

Sequence alignments were created and edited in Geneious using the MAFFT (Katoh et al., 2002) plugin with automatic algorithm selection and a 200PAM/k=2 scoring matrix with gap open penalty and offset values set to 1.53 and 0.123, respectively. The LSC of the *S. henryi* plastome was rearranged to match the syntenic configuration of *Asarum* plastomes, and the spacer regions forming the boundaries of this artificial inversion were manually trimmed and adjusted to maintain homology of alignment positions.

When aligning the nuclear rDNA contigs, it was apparent that genome walking had ventured into non-homologous regions of the genome when exiting 18S and 28S regions during the extension of contigs of some species. These non-homologous regions were found at the ends of our contigs which made them easy to identify and trim from the alignment.

Optimal partitioning of our matrix was defined as the BIC-optimal partitioning scheme as determined using PartitionFinder v1.1.1 (Lanfear et al., 2012). All genes, including exons and introns, and spacer regions were delimited and supplied to PartitionFinder; this allows the program to test the fit of models treating these genomic features singly or in aggregate. Due to the enormous size of the Bell number of the 259
hypothesized partitions, the heuristic, greedy algorithm was used to evaluate 67,341 of the possible partitioning schemes.

Maximum Likelihood Inference of Phylogeny

Maximum Likelihood (ML) inference of phylogeny was conducted using RAxML GUI version 1.3 (Silvestro, 2012) as a frontend to run the Pthreads, high-performance computing version of RAxML 7.4.2 (Stamatakis, 2006; Ott et al., 2007). Tree searches were conducted using ML + rapid bootstrapping (Stamatakis et al., 2008), while invoking the GTRGAMMA model with six discrete rate categories, and conducting 1000 bootstrap replicates.

Maximum Likelihood Inference of Selection

The CodeML module of PAML 4 (Yang, 2007) was run via PAMLX (Xu and Yang, 2013) to test for uniformity of neutrality or selection regime on plastome genes and gene families in the Asaroideae that either bordered regions of plastome disruption or were conspicuously divergent from orthologous regions in the plastome of *Saruma henryi*. The F61 (CodonFreq = 3) codon model (Goldman and Yang, 1994) was used, which parameterizes the codon rate matrix from the observed codon frequencies in the data matrix; this model is parameter-rich, but this allows for increased flexibility and biological realism by reducing *a priori* assumptions of codon frequency and nucleotide substitution rate distributions. The phylograms resulting from the analysis of rDNA singly and combined rDNA and plastome data were supplied to CodeML; branch lengths
were optimized one branch at a time. Models of selection regime linkage and decoupling are shown in Table 8.

Table 8. Key to models of \( \omega \) constraint tested in maximum likelihood analysis of selection regime.

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_1 )</td>
<td>All ( \omega ) equal</td>
</tr>
<tr>
<td>( M_{2A} )</td>
<td>( \omega_{Saruma} \neq \omega_{Asarum} )</td>
</tr>
<tr>
<td>( M_{2B} )</td>
<td>( \omega_{Saruma} = \omega_{A.epigynum+canadense} \neq \omega_{subgenus Heterotropa} )</td>
</tr>
<tr>
<td>( M_3 )</td>
<td>( \omega_{Saruma} \neq \omega_{A.epigynum+canadense} \neq \omega_{subgenus Heterotropa} )</td>
</tr>
<tr>
<td>( M_4 )</td>
<td>( \omega_{Saruma} \neq \omega_{A.epigynum} \neq \omega_{A.canadense} \neq \omega_{subgenus Heterotropa} )</td>
</tr>
<tr>
<td>( M_5 )</td>
<td>( \omega_{Saruma} \neq \omega_{A.epigynum} \neq \omega_{A.canadense} \neq \omega_{subgenus Heterotropa} \neq \omega_{A.asiasarum} )</td>
</tr>
<tr>
<td>( M_{6A} )</td>
<td>( \omega_{Saruma} \neq \omega_{A.epigynum} \neq \omega_{A.canadense} \neq \omega_{subgenus Heterotropa} \neq \omega_{A.asiasarum} \neq \omega_{A.minus} )</td>
</tr>
<tr>
<td>( M_{6B} )</td>
<td>( \omega_{Saruma} \neq \omega_{A.epigynum} \neq \omega_{A.canadense} \neq \omega_{subgenus Heterotropa} \neq \omega_{A.asiasarum} \neq \omega_{A.megacalyx} )</td>
</tr>
<tr>
<td>( M_7 )</td>
<td>( \omega_{Saruma} \neq \omega_{A.epigynum} \neq \omega_{A.canadense} \neq \omega_{A.delavayi} \neq \omega_{A.asiasarum} \neq \omega_{A.megacalyx} \neq \omega_{A.minus} )</td>
</tr>
</tbody>
</table>

**Results**

*Plastome Length*

The plastomes of *Asarum* and *Saruma* are highly variable in both their length and in the ease with which they can be circularized using standard MPS and assembly techniques. Due to the perceived diminishing returns of designing primers for Sanger sequencing and optimizing PCR protocols for extremely lengthy and troublesome AT-
repeat regions that we estimate to have very low melting points, we instead concatenated plastome contigs and present these mock plastomes. In contrast to any *Asarum* plastome, that of *Saruma henryi* forms a fully-circularized master circle 159,913 bp in length (Figure 14), whereas the plastomes of *Asarum* species could not be circularized and varied from 167,352 bp to 183,402 bp in length in *A. minus* (Figure 15) and *A. sieboldii* var. *sieboldii* (Figure 16), respectively. The remainder of sequenced plastomes totaled 172,882 bp in *A. epigynum* (Figure 17), 181,880 bp in *A. canadense* (Figure 18), 181,995 bp in *A. megacallyx* (Figure 19), and 182,680 bp in *A. delavayi* (Figure 20). Comparisons of plastome length, GC content, and spacer and gene regions can be found in Table 9. It is important to reiterate that, with the exception of *Saruma henryi*, the plastome lengths and GC contents of certain regions presented are estimates due to our inability to fully circularize these genomes.
Table 9. Comparison of length and percent GC, gene, and non-coding sequence for each plastome sequenced.

<table>
<thead>
<tr>
<th></th>
<th><em>S. henryi</em></th>
<th><em>A. epigynum</em></th>
<th><em>A. canadense</em></th>
<th><em>A. sieboldii</em></th>
<th><em>A. minus</em></th>
<th><em>A. megacalyx</em></th>
<th><em>A. delavayi</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>159,913</td>
<td>172,882</td>
<td>181,880</td>
<td>183,402</td>
<td>167,352</td>
<td>181,995</td>
<td>182,680</td>
</tr>
<tr>
<td>% GC</td>
<td>38.8</td>
<td>37.8</td>
<td>37.5</td>
<td>37.9</td>
<td>38</td>
<td>38.3</td>
<td>37.8</td>
</tr>
<tr>
<td>% Gene</td>
<td>68.3</td>
<td>70.3</td>
<td>67</td>
<td>69.1</td>
<td>68.4</td>
<td>70.1</td>
<td>68.4</td>
</tr>
<tr>
<td>% Non-coding</td>
<td>31.7</td>
<td>29.7</td>
<td>33</td>
<td>30.9</td>
<td>31.6</td>
<td>29.9</td>
<td>31.6</td>
</tr>
</tbody>
</table>
Figure 14. *Saruma henryi* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome.
Figure 15. *Asarum minus* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome.
Figure 16. *Asarum sieboldii* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome.
Figure 17. *Asarum epigynum* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome.
Figure 18. *Asarum canadense* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome.
Figure 19. *Asarum megacalycy* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome.
Figure 20. *Asarum delavayi* plastome physical map. Coding regions (CDS) are annotated in green, structural coding regions are annotated in red, tRNAs are annotated in pink. A 500 bp sliding window analysis of GC content is given as a percentage and illustrated by a blue line paralleling the genome.
Gene and tRNA Content

tRNA content, with the exception of the loss of trnH-GUG in Asarum, is identical across the sequenced taxa. The absence of trnH-GUG has been verified in each of the sampled lineages by mapping genome skimming data to the intact, but modified, copy found in Saruma henryi. Even though matches have been identified in the sequence pool of each taxon, the coverage of each copy is indicative of paralogous copies found in the nuclear and mitochondrial genomes; iterative genome walking confirmed that none of these suspected paralogs are housed in plastome sequence that was not identified during plastome assembly.

Aside from gene duplication events due to IR expansion and contraction and the loss of trnH-GUG described above, the gene content of Asaroidae plastomes is identical to that described from published magnoliid plastomes; however, several genes have been expanded or truncated in Asaroidae due to the repositioning of stop codons. In Piper cenocladium, the exons of rps16 sum to 294 bp, while these same exons in Saruma and Asarum sum to only 237 bp. The coding portion of rps16 totals 261 bp in Liriodendron tulipifera, 288 bp in Magnolia officinalis, and 270 bp in Drimys granadensis. The Asaroidae copy of rps16, which is not completely fixed between sampled lineages, contains five fixed and complete amino acid substitutions compared to that of P. cenocladium; each of these substitutions is due to the replacement of an entire codon rather than to non-synonymous SNPs.
Several plastid genes throughout the Asaroideae appear to have become pseudogenized. The *A. canadense* copy of *rpoC2* has been severely, and most likely recently, truncated to approximately half of its length due to a SNP in a poly-A repeat region. We propose that this is a relatively recent occurrence since the majority of the amino acid identities downstream of this mutation are intact, yet there are more indels, including two autapomorphic six bp insertions, in this previously coding region than in the same region of any other Asaroideae species. Similarly to *rpoC2*, one copy of *ycf2* in *A. canadense* and *accD* and *psaB* in *A. delavayi*, have all been drastically shortened due to frameshifts caused by SNPs resulting in premature stop codons. Sequence downstream of these premature terminators also shares near complete identity with homologous regions in all other species sampled in this study.

In contrast to the putative pseudogenes created by the creation of premature stop codons described above, the boundaries of the IRs fall within and appear to have disrupted the coding sequence of several duplicated genes throughout the Asaroideae, namely *ndhA, rpl2*, and *ycf1*. It appears that the IR\textsubscript{A} copy of *rpl2* in *A. canadense* is no longer functional due to the IR\textsubscript{A} boundary falling within the 3’ portion of the intron with subsequent loss of exon 2. While the IR boundary falls within *ndhA* in *A. epigynum* and *ycf1* in *A. minus*, both of which have been severely truncated, these genes do contain intact and reasonable reading frames suggesting that usable, yet altered product may still be produced.

*ycf1* is the longest gene found in Asaroideae plastomes, and its duplication in all *Asarum* plastomes sequenced to date means that it contributes the most additional coding
sequence of any single gene to the increased length of these genomes. In stark contrast to ycf2, we find that ycf1 has been expanded more than six-fold (5,601 bp in *A. delavayi*), compared to the copy found in *P. cenocladum* (927 bp). The expansion of this gene is not due to the insertion of novel sequence, but rather to mutations leading to a change in the location of the stop codon; this frameshift occurred through the expansion of an AT-rich region. With the exception of *A. minus*, both copies of ycf1 in *Asarum* are found within the IR boundaries. The asymmetry in ycf1 length and its co-occurrence with the IR boundary in *A. minus* suggest that the duplication, expansion and contraction of ycf1 have played a role in the duplication of the SSC in *Asarum* which is later discussed in detail.

ycf15 appears not to be functional in *S. henryi* due to a premature stop codon in Exon 1, while the reading frame is intact in all *Asarum* species except for *A. delavayi*; this loss of function seems to be independently derived in both *S. henryi* and *A. delavayi*, even though this exon contains two amino acid changes in *Asarum* relative to *P. cenocladum*, it is more similar to the latter than it is to that of *S. henryi*.

**The IR, SSC and Shifts of their Boundaries**

The boundaries of the IRs of *Asarum* are highly dynamic, and have experienced positional shifts at both the SSC- and LSC-boundaries (Figures 19 and 20). The greatest shift in the position of the LSC-IR boundary involved the incorporation of 3,492 bp downstream of the LSC-IRB boundary into IR_{A} of *A. epigynum*; this resulted in the duplication of the *rps19, rpl22, rps3, rpl16, and rpl14* genes. The remnants of this shift are present in all sampled *Asarum* species, with the exception of *A. canadense*; however,
a portion of this region in *A. sieboldii* var. *sieboldii* contains seemingly unique sequence that only otherwise closely matches plastid sequence reported from the distantly related *Illicium oligandrum*. A partial copy of *rps19* has been added to the LSC-IRₐ boundary of *Saruma henryi*, which is interpreted as the precursor of the expansion of IRa, the loss of *trnH*-GUG, and ultimately the rearrangement of the LSC in *Asarum*; we later discuss this in depth and propose a novel model of plastome disruption.
Figure 21. Gains (black bars) and losses (white bars) associated with Inverted Repeat-Large Single Copy (IR-LSC) and Inverted Repeat-Small Single Copy (IR-SSC) expansions plotted on a Maximum Likelihood cladogram inferred through the analysis of nuclear rDNA; bootstrap values are plotted along branches. Note the contrast between the rapid expansion and subsequent erosion of the IR-LSC boundary and the more gradual, and rarely lost, duplication of SSC through shifting of the IR-SSC boundary.
Figure 22. Gains (black bars) and losses (white bars) associated with Inverted Repeat-Large Single Copy (IR-LSC) and Inverted Repeat-Small Single Copy (IR-SSC) expansions plotted on a Maximum Likelihood cladogram inferred through the analysis of our plastome and combined datasets; bootstrap values for each are plotted along branches, respectively. Note the contrast between the rapid expansion and subsequent erosion of the IR-LSC boundary and the more gradual, and rarely lost, duplication of SSC through shifting of the IR-SSC boundary.
While changes in the position of the IR-LSC boundary appear to be highly plastic within Asaroideae, the repositioning of the SSC-IR bounds is more dramatic than what has been described to date in the angiosperms and has disproportionately influenced the lengthening of *Asarum* plastomes. We find that IR expansion has resulted in the total duplication, and therefore loss, of all typically single copy genes in the SSC regions of *A. canadense, A. delavayi, A. megacalyx,* and *A. sieboldii.* This means that the IR regions of the SSC-lacking species are continuous and palindromic; the longest of these contiguous IR regions is 96,542 bp long in *A. delavayi* – comprising nearly 53% of the total length of the genome. Providing critical, and polarizing, references are the plastomes of *Saruma henryi* and *A. epigynum,* the former of which harbors a completely intact SSC while that of the latter is duplicated from *ycf1* through the initial 3’ portion of the *ndhA* exon 2. Each of the completely duplicated SSC regions differs most strikingly in the length and divergence of two genes, *ycf1* and *ndhF;* these genes stand at the bounds of typical and pleisiomorphically arranged copies of the SSC which may suggest that they played a role in the duplication of this region in *Asarum.*

**GC Content**

When compared over their total length, GC content is not significantly different between any plastome sequenced here and those previously published from the magnoliids. However, increases, and especially decreases, in regional GC content are pronounced.
The GC content of several genes in the plastomes of Asaroideae species is lower than their homologs in *P. cenocladum*. The *ycf1* gene in *P. cenocladum* is comprised of 37.6% G or C, while the GC content of the same gene varies between 33.8 and 34% in Asaroideae. Additionally, *psbK* (41.1%), *psbT* (35.3%), *rpl22* (41.9%), *rps11* (48%), *rps16* (40.1%), and *rps19* (41.2%) in *P. cenocladum* contain between 2 and 4.1% higher GC content than their homologs in Asaroideae.

Contrastingly, the coding regions of many genes in Asaroideae have a higher GC content than do their homologs in *P. cenocladum*. For instance, *atpH* (46.3%), *matK* (36%), *petG* (36%), *psbJ* (42.3%), *psbM* (33.3%), *psbN* (47%), *rpl14* (41.7%), *rpl32* (35.8%), *rpl36* (39.5%), and *rps15* (35.9%) in *A. delavayi* range from 2.3 to 4% higher GC content compared to *P. cenocladum*.

The GC content of several spacer regions that, with the aid of Sanger sequencing, we have sequenced in their entirety shows strong phylogenetic signal corresponding to the nuclear rDNA phylogram with regard to their expansion and GC content (Figure 23). Expansion driven by AT repeats has been found in the *trnS*-GCU – *trnG*-GCC, *rpoA* – *petD*, and *ndhG* – *ndhE* intergenic spacer regions (Figure 23 A-C). Contrastingly, the intergenic spacers of *rpl14* – *rps8* and *ccsA* – *ndhD* appear to have shortened and undergone GC enrichment (Figure 23 D and E). Somewhat similarly, the *rpl32* – *trnL*-UAG and *ndhF* – *rpl32* intergenic regions have seen a rise in GC content in the Asaroideae compared to those of *P. cenocladum*, yet have only minimally lengthened compared to regions that have expanded due to an overwhelming addition of A or T residues (Figure 23 F-G).
Figure 23. Select genetic spacer regions characterized alongside Maximum Likelihood phylogenies which have been inferred from the analysis of nuclear rDNA. The magnitudes of AT-biased expansion (A-C), GC-biased compression (D-E), and GC-biased enrichment (F & G) appear to increase in concert with the derivation of lineages. We propose that AT-biased expansion is evidence of either the relaxation of genomic proofreading or a reduction in proofreading fidelity, while both GC-biased compression and enrichment are the result of overall genome GC content balancing in response to these AT-rich expansions.
<table>
<thead>
<tr>
<th></th>
<th>tmS-GCU -- tmG-GCC</th>
<th></th>
<th>rpoA -- petD</th>
<th></th>
<th>ndhG -- ndhE</th>
<th></th>
<th>rpl14 -- rps8</th>
<th></th>
<th>ccsA -- ndhD</th>
<th></th>
<th>rpl32 -- tmL-UAG</th>
<th></th>
<th>ndhF -- rpl32</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td>Length (bp)</td>
<td>GC Content (%)</td>
<td>AT-biased Expansion</td>
<td>Length (bp)</td>
<td>GC Content (%)</td>
<td>AT-biased Expansion</td>
<td>Length (bp)</td>
<td>GC Content (%)</td>
<td>AT-biased Expansion</td>
<td>Length (bp)</td>
<td>GC Content (%)</td>
<td>AT-biased Expansion</td>
<td>Length (bp)</td>
</tr>
<tr>
<td>Asarum minus</td>
<td>2,485</td>
<td>15.5</td>
<td></td>
<td>471</td>
<td>11.9</td>
<td></td>
<td>196</td>
<td>27.0</td>
<td></td>
<td>499</td>
<td>12.4</td>
<td></td>
<td>232</td>
</tr>
<tr>
<td>Asarum canadense</td>
<td>2,398</td>
<td>16.8</td>
<td></td>
<td>408</td>
<td>13.0</td>
<td></td>
<td>256</td>
<td>22.7</td>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Asarum epigynum</td>
<td>2,024</td>
<td>19.9</td>
<td></td>
<td>191</td>
<td>32.5</td>
<td></td>
<td>232</td>
<td>32.3</td>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Saruma henryi</td>
<td>1,944</td>
<td>30.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piper cenocladium</td>
<td>1,489</td>
<td>31.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asarum minus</td>
<td>832</td>
<td>7.8</td>
<td></td>
<td>471</td>
<td>11.9</td>
<td></td>
<td>196</td>
<td>27.0</td>
<td></td>
<td>499</td>
<td>12.4</td>
<td></td>
<td>232</td>
</tr>
<tr>
<td>Asarum canadense</td>
<td>471</td>
<td>11.9</td>
<td></td>
<td>408</td>
<td>13.0</td>
<td></td>
<td>256</td>
<td>22.7</td>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Asarum epigynum</td>
<td>408</td>
<td>13.0</td>
<td></td>
<td>191</td>
<td>32.5</td>
<td></td>
<td>232</td>
<td>32.3</td>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Saruma henryi</td>
<td>196</td>
<td>27.0</td>
<td></td>
<td>232</td>
<td>32.3</td>
<td></td>
<td>232</td>
<td>32.3</td>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Piper cenocladium</td>
<td>191</td>
<td>32.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asarum minus</td>
<td>499</td>
<td>12.4</td>
<td></td>
<td>404</td>
<td>15.6</td>
<td></td>
<td>256</td>
<td>22.7</td>
<td></td>
<td>499</td>
<td>12.4</td>
<td></td>
<td>232</td>
</tr>
<tr>
<td>Asarum canadense</td>
<td>404</td>
<td>15.6</td>
<td></td>
<td>232</td>
<td>32.3</td>
<td></td>
<td>232</td>
<td>32.3</td>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Piper cenocladium</td>
<td>?</td>
<td>?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>D</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asarum minus</td>
<td>176</td>
<td>27.8</td>
<td></td>
<td>187</td>
<td>28.3</td>
<td></td>
<td>187</td>
<td>28.3</td>
<td></td>
<td>176</td>
<td>27.8</td>
<td></td>
<td>226</td>
</tr>
<tr>
<td>Asarum canadense</td>
<td>187</td>
<td>28.3</td>
<td></td>
<td>238</td>
<td>24.8</td>
<td></td>
<td>238</td>
<td>24.8</td>
<td></td>
<td>187</td>
<td>28.3</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Asarum epigynum</td>
<td>238</td>
<td>24.8</td>
<td></td>
<td>226</td>
<td>25.7</td>
<td></td>
<td>226</td>
<td>25.7</td>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Saruma henryi</td>
<td>226</td>
<td>25.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piper cenocladium</td>
<td>226</td>
<td>25.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>E</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asarum minus</td>
<td>214</td>
<td>30.8</td>
<td></td>
<td>223</td>
<td>30.5</td>
<td></td>
<td>206</td>
<td>33.5</td>
<td></td>
<td>214</td>
<td>30.8</td>
<td></td>
<td>337</td>
</tr>
<tr>
<td>Asarum canadense</td>
<td>223</td>
<td>30.5</td>
<td></td>
<td>337</td>
<td>16.9</td>
<td></td>
<td>337</td>
<td>16.9</td>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Piper cenocladium</td>
<td>337</td>
<td>16.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>F</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asarum minus</td>
<td>1,171</td>
<td>26.1</td>
<td></td>
<td>1,169</td>
<td>26.3</td>
<td></td>
<td>852</td>
<td>27.1</td>
<td></td>
<td>1,171</td>
<td>26.1</td>
<td></td>
<td>1,034</td>
</tr>
<tr>
<td>Asarum canadense</td>
<td>1,169</td>
<td>26.3</td>
<td></td>
<td>852</td>
<td>27.1</td>
<td></td>
<td>852</td>
<td>27.1</td>
<td></td>
<td>1,169</td>
<td>26.3</td>
<td></td>
<td>852</td>
</tr>
<tr>
<td>Piper cenocladium</td>
<td>?</td>
<td>?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>G</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asarum minus</td>
<td>1,241</td>
<td>25.8</td>
<td></td>
<td>1,295</td>
<td>26.3</td>
<td></td>
<td>1,193</td>
<td>26.2</td>
<td></td>
<td>1,241</td>
<td>25.8</td>
<td></td>
<td>926</td>
</tr>
<tr>
<td>Asarum canadense</td>
<td>1,295</td>
<td>26.3</td>
<td></td>
<td>1,193</td>
<td>26.2</td>
<td></td>
<td>1,193</td>
<td>26.2</td>
<td></td>
<td>1,295</td>
<td>26.3</td>
<td></td>
<td>926</td>
</tr>
<tr>
<td>Piper cenocladium</td>
<td>926</td>
<td>22.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Introns**

The cumulative length of introns in Asaroideae plastomes is nearly 1,000 bp larger than that of *P. cenocladium* while the average GC content of these regions is nearly identical between all plastomes presented here and the former. The most apparent differences in the plastomes of the Asaroideae are the loss of the intron and exon 2 of *ycf15*, due to a premature stop codon in both *S. henryi* and *A. delavayi*, the loss of the 5’ portion of the intron and all of exon 2 of *rpl2* in *A. canadense*, and the hyper-mutability of intron 1 of *ycf3*.

The extension of the *ycf3* intron 1, and the resulting reduction of GC content in this region, is one of the most exaggerative differences between the plastomes of *Asarum* species and the remainder of the magnoliids. The insertion of approximately 420 bp in *ycf3* intron 1 of *A. minus* relative to *S. henryi*, the GC content of which is 6.3%, is remarkable, and is further discussed below.

**Intergenic Spacer Regions**

Nearly all expansions in Asaroideae plastomes are due to the insertion of AT-rich residues. Figure 23 A-G illustrates both the extreme nature of the expansion of some spacer regions and their surprising correspondence with the nuclear rDNA cistron topology. Spacer regions at the LSC-IR boundaries appear to be particularly prone to this type of expansion; 500 bp sliding window analyses of GC content plotted inside the circularized genomes in Figures 14 – 20 show dramatic decreases in GC content in these regions.
The expansion of *ycf3* intron 1 is similar to that of the *petD-rpoA, ndhG-ndhE*, and the *trnD-psbM* spacer regions; potentially of note, each of these regions lies approximately equidistant from one another when the plastome is shown in the stereotypical circular configuration – we present a hypothesis of the drivers of these regions in the discussion.

**Maximum Likelihood Inference of Phylogeny**

Topological inference using plastome or the nuclear rDNA cistron data singly, or in combination, results in the recovery of several supported but incongruent relationships across trees. Maximum likelihood analysis of nuclear rDNA sequence recovers *A. epigynum* (subgenus *Geotaenium*) as sister to the remainder of *Asarum* with bootstrap support of 73. However, ML inference using plastome data singly or in combination with the nuclear rDNA cistron data recovers an *A. canadense* (subgenus *Asarum*) + *A. epigynum* clade that is sister to the remainder of the genus, which is supported by bootstrap values of 59 and 69, respectively.

Relationships within subgenus *Heterotropa* were highly supported (bootstrap values between 93 and 100) when plastome or rDNA data were analyzed singly or in combination, yet the relationships evidenced in each genomic compartment are highly incongruent. Relationships inferred through the analysis of the nuclear rDNA cistron recover *A. sieboldii* var. *sieboldii* (section *Asiasarum*) as sister to *A. minus* (section *Hexastylis*; bootstrap = 93) + (*A. megacalyx*; section *Heterotropa* + *A. delavayi*; section *Longistylis*; bootstrap = 94). However, plastome data analyzed singly recovers two clades
within subgenus *Heterotropa*; one comprising sections *Hexastylis* and *Longistylis* and the other sections *Asiasarum* and *Longistylis* – these relationships are supported by bootstrap values of 100.

*Maximum Likelihood Inference of Selection*

Directionality of selective pressure or evidence of selective neutrality is significantly different across loci and/or Asaroideae lineages for multiple genes and one gene family; all PAML analyses that identified significant selection regime shifts within the Asaroideae are shown in Tables 10 and 11.
Table 10. Results of maximum likelihood tests of selection regime in PAML 4 using the branch selection and the F61 codon models on the nuclear rDNA cistron-based phylogram. Significance of parameter change on model improvement was determined via likelihood ratio tests. Constraint schemes are described in Table 8.
<table>
<thead>
<tr>
<th>Gene</th>
<th>ω Model</th>
<th>Free Parameters</th>
<th>Neg Ln Likelihood</th>
<th>LRT Pair</th>
<th>Chi Value</th>
<th>P Value</th>
<th>Sig. Post Bonferroni</th>
</tr>
</thead>
<tbody>
<tr>
<td>ycfI</td>
<td>M₁</td>
<td>73</td>
<td>-7813.763045</td>
<td>M₁ vs M₂A</td>
<td>10.534878</td>
<td>0.0012**</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>M₂A</td>
<td>74</td>
<td>-7808.495606</td>
<td>M₁ vs M₂B</td>
<td>6.433078</td>
<td>0.0112*</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>M₂B</td>
<td>74</td>
<td>-7810.546506</td>
<td>M₂A vs M₃</td>
<td>0.029014</td>
<td>0.8648</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₃</td>
<td>75</td>
<td>-7808.481099</td>
<td>M₂B vs M₃</td>
<td>4.130814</td>
<td>0.0421*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₄</td>
<td>76</td>
<td>-7807.935709</td>
<td>M₃ vs M₄</td>
<td>1.09078</td>
<td>0.2963</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₅</td>
<td>77</td>
<td>-7807.526175</td>
<td>M₄ vs M₅</td>
<td>0.819068</td>
<td>0.3654</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₆A</td>
<td>78</td>
<td>-7805.790017</td>
<td>M₅ vs M₆A</td>
<td>3.472316</td>
<td>0.0624</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₆B</td>
<td>78</td>
<td>-7807.495002</td>
<td>M₆A vs M₇</td>
<td>0.062346</td>
<td>0.8029</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₇</td>
<td>79</td>
<td>-7805.658929</td>
<td>M₆B vs M₇</td>
<td>3.672146</td>
<td>0.0553</td>
<td></td>
</tr>
<tr>
<td>atpA</td>
<td>M₁</td>
<td>73</td>
<td>-2078.548133</td>
<td>M₁ vs M₂A</td>
<td>0.218302</td>
<td>0.6403</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₂A</td>
<td>74</td>
<td>-2078.438982</td>
<td>M₁ vs M₂B</td>
<td>3.462436</td>
<td>0.0628</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₂B</td>
<td>74</td>
<td>-2076.816915</td>
<td>M₂A vs M₃</td>
<td>5.498686</td>
<td>0.019*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₃</td>
<td>75</td>
<td>-2075.689639</td>
<td>M₂B vs M₃</td>
<td>2.254552</td>
<td>0.1332</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₄</td>
<td>76</td>
<td>-2072.924391</td>
<td>M₃ vs M₄</td>
<td>5.530496</td>
<td>0.0187*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₅</td>
<td>77</td>
<td>-2072.925472</td>
<td>M₄ vs M₅</td>
<td>0.002162</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₆A</td>
<td>78</td>
<td>-2072.917164</td>
<td>M₅ vs M₆A</td>
<td>0.016616</td>
<td>0.8975</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₆B</td>
<td>78</td>
<td>-2072.916406</td>
<td>M₅ vs M₆B</td>
<td>0.018132</td>
<td>0.893</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₇</td>
<td>79</td>
<td>-2072.916406</td>
<td>M₆A vs M₇</td>
<td>0.001516</td>
<td>0.9691</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₈</td>
<td>79</td>
<td>-2072.916406</td>
<td>M₆B vs M₇</td>
<td>0.002656</td>
<td>0.893</td>
<td></td>
</tr>
<tr>
<td>ccsA</td>
<td>M₁</td>
<td>73</td>
<td>-1196.440361</td>
<td>M₁ vs M₂A</td>
<td>0.9454</td>
<td>0.3309</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₂A</td>
<td>74</td>
<td>-1195.967661</td>
<td>M₁ vs M₂B</td>
<td>0.39014</td>
<td>0.5322</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₂B</td>
<td>74</td>
<td>-1196.245291</td>
<td>M₂A vs M₃</td>
<td>0.03386</td>
<td>0.8539</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₃</td>
<td>75</td>
<td>-1195.950731</td>
<td>M₂B vs M₃</td>
<td>0.58912</td>
<td>0.4428</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₄</td>
<td>76</td>
<td>-1195.923401</td>
<td>M₃ vs M₄</td>
<td>0.05466</td>
<td>0.8154</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₅</td>
<td>77</td>
<td>-1195.469767</td>
<td>M₄ vs M₅</td>
<td>0.907268</td>
<td>0.3408</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₆A</td>
<td>78</td>
<td>-1194.35392</td>
<td>M₅ vs M₆A</td>
<td>2.231694</td>
<td>0.1352</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₆B</td>
<td>78</td>
<td>-1192.108783</td>
<td>M₅ vs M₆B</td>
<td>6.721968</td>
<td>0.0095**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₇</td>
<td>79</td>
<td>-1192.107455</td>
<td>M₆A vs M₇</td>
<td>4.49293</td>
<td>0.034*</td>
<td></td>
</tr>
</tbody>
</table>

Table 10 cont.
Table 10 cont.

<table>
<thead>
<tr>
<th>Genes</th>
<th>M1</th>
<th>M2A</th>
<th>M2B</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6A</th>
<th>M6B</th>
<th>M7</th>
<th>M6A vs M7</th>
<th>M6B vs M7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>psaA</strong></td>
<td>-2855.146596</td>
<td>-2852.822328</td>
<td>-2853.742098</td>
<td>-2852.641618</td>
<td>-2851.778491</td>
<td>-2850.817457</td>
<td>-2850.363674</td>
<td>-2850.817457</td>
<td>-2850.294482</td>
<td>0.138384</td>
<td>1.04595</td>
</tr>
<tr>
<td></td>
<td>M1 vs M2A</td>
<td>2.808996</td>
<td>0.36142</td>
<td>2.20096</td>
<td>1.726254</td>
<td>1.922068</td>
<td>0.907566</td>
<td>0.080956</td>
<td>0.138384</td>
<td>1.04595</td>
<td>0.3064</td>
</tr>
<tr>
<td></td>
<td>M1 vs M2B</td>
<td>0.0345*</td>
<td>0.5477</td>
<td>0.1379</td>
<td>0.1889</td>
<td>0.1656</td>
<td>0.3408</td>
<td>0.7761</td>
<td>0.7095</td>
<td>0.3064</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M1 vs M3</td>
<td>0.0937</td>
<td>0.1748</td>
<td>0.1889</td>
<td>0.2087</td>
<td>0.2033</td>
<td>0.3081</td>
<td>0.2033</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M1 vs M4</td>
<td>0.0937</td>
<td>0.5477</td>
<td>0.1889</td>
<td>0.3081</td>
<td>0.2033</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M1 vs M5</td>
<td>0.0937</td>
<td>0.5477</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rpl20</td>
<td>-501.067395</td>
<td>-499.473767</td>
<td>-500.146731</td>
<td>-499.337491</td>
<td>-498.823167</td>
<td>-498.032864</td>
<td>-497.992386</td>
<td>-495.617323</td>
<td>-495.617429</td>
<td>4.749914</td>
<td>0.0293*</td>
</tr>
<tr>
<td></td>
<td>M1 vs M2A</td>
<td>3.187256</td>
<td>0.6016</td>
<td>0.2033</td>
<td>1.028648</td>
<td>1.580606</td>
<td>0.080956</td>
<td>4.831082</td>
<td>4.749914</td>
<td>0.0293*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M1 vs M2B</td>
<td>0.0742</td>
<td>0.6016</td>
<td>0.2033</td>
<td>0.3081</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M1 vs M3</td>
<td>0.6016</td>
<td>0.2033</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M1 vs M4</td>
<td>0.6016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M1 vs M5</td>
<td>0.6016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All pet</td>
<td>-3333.755833</td>
<td>-3333.389129</td>
<td>-3332.174516</td>
<td>-3332.152015</td>
<td>-3331.530827</td>
<td>-3330.910452</td>
<td>-3327.414588</td>
<td>-3329.690279</td>
<td>-3327.414588</td>
<td>4.749914</td>
<td>0.0293*</td>
</tr>
<tr>
<td>genes</td>
<td>M1 vs M2A</td>
<td>0.733408</td>
<td>0.0753</td>
<td>0.1157</td>
<td>0.045002</td>
<td>1.24075</td>
<td>6.991728</td>
<td>2.440346</td>
<td>0.000212</td>
<td>1.04595</td>
<td>0.3064</td>
</tr>
<tr>
<td></td>
<td>M1 vs M2B</td>
<td>0.3918</td>
<td>0.0753</td>
<td>0.1157</td>
<td>0.5023</td>
<td>0.265</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M1 vs M3</td>
<td>0.3918</td>
<td>0.0753</td>
<td>0.1157</td>
<td>0.265</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M1 vs M4</td>
<td>0.3918</td>
<td>0.0753</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M1 vs M5</td>
<td>0.3918</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M2A vs M2B</td>
<td>0.1157</td>
<td>0.5023</td>
<td>0.265</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M2A vs M3</td>
<td>0.1157</td>
<td>0.265</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M2A vs M4</td>
<td>0.1157</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M2A vs M5</td>
<td>0.1157</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M2B vs M3</td>
<td>0.5023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M2B vs M4</td>
<td>0.5023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M2B vs M5</td>
<td>0.5023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M3 vs M4</td>
<td>0.265</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M3 vs M5</td>
<td>0.265</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M4 vs M5</td>
<td>0.265</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M6A vs M6B</td>
<td>0.1157</td>
<td>0.5023</td>
<td>0.265</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M6A vs M7</td>
<td>0.1157</td>
<td>0.265</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M6B vs M7</td>
<td>0.1157</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M6B vs M5</td>
<td>0.5023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M6B vs M4</td>
<td>0.5023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M6B vs M3</td>
<td>0.5023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.05, ** p < 0.01
Table 11. Results of maximum likelihood tests of selection regime in PAML 4 using the branch selection and the F61 codon models on the combined data phylogram. Significance of parameter change on model improvement was determined via likelihood ratio tests. Constraint schemes are described in Table 8.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Model</th>
<th>Free Parameters</th>
<th>Neg Ln Likelihood</th>
<th>LRT Pair</th>
<th>Chi Value</th>
<th>P Value</th>
<th>Sig. Post Bonferonni</th>
</tr>
</thead>
<tbody>
<tr>
<td>All pet genes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₁</td>
<td>73</td>
<td>-3319.38</td>
<td>M₁ vs M₂A</td>
<td>0.015686</td>
<td>0.9003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₂A</td>
<td>74</td>
<td>-3319.37</td>
<td>M₁ vs M₂B</td>
<td>2.159326</td>
<td>0.1417</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₂B</td>
<td>74</td>
<td>-3318.3</td>
<td>M₂A vs M₃</td>
<td>1.03207</td>
<td>0.3097</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₃</td>
<td>75</td>
<td>-3318.86</td>
<td>M₂B vs M₃</td>
<td>-1.11157</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₄</td>
<td>76</td>
<td>-3318.24</td>
<td>M₃ vs M₄</td>
<td>1.242324</td>
<td>0.265</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₅</td>
<td>77</td>
<td>-3318</td>
<td>M₄ vs M₅</td>
<td>0.469898</td>
<td>0.493</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₆A</td>
<td>78</td>
<td>-3315.14</td>
<td>M₅ vs M₆A</td>
<td>5.716146</td>
<td>0.0168*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₆B</td>
<td>78</td>
<td>-3317.42</td>
<td>M₅ vs M₆B</td>
<td>1.164786</td>
<td>0.2805</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₇</td>
<td>79</td>
<td>-3315.14</td>
<td>M₆A vs M₇</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M₆B vs M₇</td>
<td>4.55136</td>
<td>0.0329*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ycfI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₁</td>
<td>73</td>
<td>-7769.84</td>
<td>M₁ vs M₂A</td>
<td>3.590918</td>
<td>0.0581</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₂A</td>
<td>74</td>
<td>-7768.05</td>
<td>M₁ vs M₂B</td>
<td>3.501904</td>
<td>0.0613</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₂B</td>
<td>74</td>
<td>-7768.09</td>
<td>M₂A vs M₃</td>
<td>0.706946</td>
<td>0.4005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₃</td>
<td>75</td>
<td>-7767.69</td>
<td>M₂B vs M₃</td>
<td>0.79596</td>
<td>0.3723</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₄</td>
<td>76</td>
<td>-7767.15</td>
<td>M₃ vs M₄</td>
<td>1.090722</td>
<td>0.2963</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₅</td>
<td>77</td>
<td>-7767.15</td>
<td>M₄ vs M₅</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₆A</td>
<td>78</td>
<td>-7764.91</td>
<td>M₅ vs M₆A</td>
<td>4.471352</td>
<td>0.0345*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₆B</td>
<td>78</td>
<td>-7767.04</td>
<td>M₅ vs M₆B</td>
<td>0.217436</td>
<td>0.641</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₇</td>
<td>79</td>
<td>-7764.81</td>
<td>M₆A vs M₇</td>
<td>0.19679</td>
<td>0.6573</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M₆B vs M₇</td>
<td>4.450706</td>
<td>0.0349*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ycf1 and the pet gene family are inferred to be under positive selection when analyzing either the rDNA or combined phylograms. For ycf1 selection inference using the rDNA phylogram, significant improvement in model likelihood is achieved by decoupling Saruma from Asarum (M1 vs. M2A, p=0.001; M1 vs. M2B, p=0.01) and further improvement is gained when Saruma, Asarum subgenus Asarum, subgenus Geotaenium, and Asarum subgenus Heterotropa are decoupled (M2B vs. M3, p=0.04); the two former model comparisons are significant following Bonferroni correction. ycf1 selection regime is also inferred to be positive-shifted in Asarum lineages relative to Saruma when analyzing the combined phylogram; however, they are not significant following Bonferroni correction. The pet gene family is inferred to be under positive selection in both Asarum subgenus Geotaenium and section Hexastylis; in similar fashion to ycf1, this result is inferred from analysis of both phylograms.

In contrast to the robustness of selection regime inference of ycf1 and the pet family across phylograms, analysis of codon substitution upon the rDNA topology identifies selection regime shifts in multiple genes that are not identified when analyzing the combined phylogram. atpA is identified as being positive-shifted in Asarum subgenus Asarum (M2A vs. M3, p=0.019; M3 vs. M4, p=0.019) while the remainder of Asarum lineages are selectively neutral. Similarly, ccsA in Asarum sections Asiasarum and Heterotropa is positive-shifted relative to the remainder of Asarum (M5 vs. M6B, p=0.009; M6A vs. M7, p=0.034). Contrastingly, rpl20 is shifted toward purifying selection in Asarum sections Asarum, Hexastylis, and Longistylis (M5 vs. M6B, p=0.028; M6A vs. M7, p=0.029). P values resulting from likelihood ratio tests of model likelihood
improvement discussed above are not significant post-Bonferroni correction; however, these results do suggest that selective regime is not uniform across *Asarum* lineages and plastid loci.

**Discussion**

*Phylogenetic Incongruence between Genomic Compartments*

As in Sinn et al. (2015b) we find that the nuclear rDNA cistron and the plastome evidence strikingly different evolutionary histories of *Asarum*; yet, with the exceptions of the relationships of sections *Asarum* and *Geotaenium* to the remainder of the genus, the analysis of each result in bootstrap values that typically instill confidence in the resultant phylograms.

The most parsimonious plotting of morphological (Sinn et al., 2015b) and genomic changes is implied by the phylogram resulting from the analysis of nuclear rDNA data. Combined and plastome-only analyses imply multiple restorations of at least a portion of the SSC in *A. epigynum* and *A. minus*, in contrast with a single restoration of this region in *A. minus* when considering the rDNA topology.

*AT-driven Expansion of Intron and Intergenic Regions and Balancing of Overall GC Content*

It has long been recognized that angiosperm plastomes are AT rich, commonly with overall GC content less than 40% (Ruhlman and Jansen, 2014). The conservation of
plastome-wide GC content throughout the magnoliids especially in light of AT-biased expansion and the total duplication of the SSC in several *Asarum* lineages, a region long known to harbor the lowest GC content of each of the three functional portions of the plastome (Cai et al., 2006), suggests that an overall balance of GC content is maintained by some selective force. Furthermore, regional shifts in GC content appear to be driven by hypermutable regions which likely result from slipped-strand errors and subsequent error-prone correction (Magee et al., 2010). GC content shifts of similar length and magnitude have only otherwise been described from *Cypripedium* (Orchidaceae; Kim et al., 2014), a derived monocot genus.

The GC content of previously characterized magnoliid plastomes ranges from 38 to 39% (Cai et al., 2006), and despite the lengthy AT repeats found in *Asarum* plastomes, their GC content ranges from 37.5% and 38.3%. The similarity between overall GC content of these plastomes is remarkable, especially since the majority of plastome lengthening in *Asarum*, aside from that caused by the expansion of the IR, is almost entirely due to AT-rich residues. These AT-repeat regions result in an extreme number of tandem repeats, the total number of which are otherwise only rivaled by *Pelargonium* (Table 12). Highly conserved overall GC content, in light of many regions of *Asarum* plastomes where AT composition and length has only recently been similarly described (Kim et al., 2014), suggests that strong pressure can be exerted on the plastome to reconcile extreme regional substitution bias with genome-wide trends, potentially due to plastome stability tradeoffs.
Table 12. Total number of perfect repeats identified using Phobos in plastomes from across the angiosperm tree of life compared with those newly described here.

<table>
<thead>
<tr>
<th>Plastome</th>
<th>Total Number of Repeats</th>
<th>% increase over L. tulipifera</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liriodendron tulipifera</td>
<td>728</td>
<td></td>
</tr>
<tr>
<td>Magnolia officinalis</td>
<td>745</td>
<td>2.3</td>
</tr>
<tr>
<td>Drimys granadensis</td>
<td>839</td>
<td>15.2</td>
</tr>
<tr>
<td>Piper cenocladum</td>
<td>840</td>
<td>15.2</td>
</tr>
<tr>
<td>Saruma henryi</td>
<td>863</td>
<td>18.5</td>
</tr>
<tr>
<td>Vitis vinifera</td>
<td>929</td>
<td>27.6</td>
</tr>
<tr>
<td>Drimys granadensis</td>
<td>839</td>
<td>15.2</td>
</tr>
<tr>
<td>Magnolia officinalis</td>
<td>745</td>
<td>2.3</td>
</tr>
<tr>
<td>Piper cenocladum</td>
<td>840</td>
<td>15.2</td>
</tr>
<tr>
<td>Saruma henryi</td>
<td>863</td>
<td>18.5</td>
</tr>
<tr>
<td>Vitis vinifera</td>
<td>929</td>
<td>27.6</td>
</tr>
<tr>
<td>Pelargonium alternans</td>
<td>935</td>
<td>28.4</td>
</tr>
<tr>
<td>Amborella trichopoda</td>
<td>949</td>
<td>30.4</td>
</tr>
<tr>
<td>Vaccinium macrocarpon</td>
<td>980</td>
<td>34.6</td>
</tr>
<tr>
<td>Asarum megacalyx</td>
<td>1060</td>
<td>45.6</td>
</tr>
<tr>
<td>Asarum minus</td>
<td>1110</td>
<td>52.5</td>
</tr>
<tr>
<td>Pelargonium x hortorum</td>
<td>1122</td>
<td>54.1</td>
</tr>
<tr>
<td>Asarum epigynum</td>
<td>1132</td>
<td>55.5</td>
</tr>
<tr>
<td>Asarum sieboldii</td>
<td>1152</td>
<td>58.2</td>
</tr>
<tr>
<td>Asarum delavayi</td>
<td>1157</td>
<td>58.9</td>
</tr>
<tr>
<td>Asarum canadense</td>
<td>1247</td>
<td>71.3</td>
</tr>
</tbody>
</table>

Some intron and intergenic regions were found to be highly divergent from those of Piper and Saruma, especially due to the presence of highly repetitive AT-rich expansions, yet many such regions were not able to be sequenced using either Illumina or Sanger methods. For example, the trnS-GCU – trnG-GCC spacer (Figure 23) which in Piper cenocladum is 1,489 bp in length and has a GC content of 31.7% (Cai et al., 2006), varies from 2,024 bp in length with a GC content of 19.9% in A. epigynum to 2,485 bp in length and a GC content of 15.5% in A. minus. When plotted on our rDNA cistron-inferred topology, this and similar AT-rich expansions show absolute correspondence. Critically, the trnS-GCU – trnG-GCC spacer in Saruma henryi is extremely similar to
that of *Piper cenocladum* in GC content (30.6% and 31.7%, respectively), despite being 455 bp longer; this same region is only 80 bp longer in *A. epigynum*, which is sister to the remainder of *Asarum*, but GC content is 10.7% lower. Strong phylogenetic signal in these genomic changes suggests that with denser sampling we may be able to identify the precursors of genomic change and generate mechanistic hypotheses.

In similar fashion to some intergenic spacer regions, *ycf3* intron 1 has increased in length and AT content due to putative slipped-strand replication and subsequent faulty error correction. The physical location of this expansion, and the presence of four similarly evolving, nearly equidistant points of potential steric stress found in *Asarum* plastomes, leads us to posit that the hypermutability of this and similar features represent regions of reduced replication efficacy due to decreased molecular stability.

For instance, it has long been established that genes contained within the IRs of the plastome evolve more slowly than those outside of this region (Ruhlman and Jansen, 2014). Conversely, sequence found at either the boundary of the IR or at the midpoint of one of the single copy regions may experience a heightened rate of divergence due to physical (steric) stress on that region. Since a single intron of *ycf3*, which is found at the center of the LCS, has behaved in much the same way as spacer regions near the boundaries of both IRs and the central portion of the SSC, this may suggest that the location of this feature in the center of the LSC, along with lengthy AT-rich expansions in the LSC and an extraordinary lengthening of the plastome as a whole, is a consequence of increased steric stress on this portion of the molecule which has resulted in an increased rate of poorly corrected sequence. It is important to note that while
repetitive sequence is often thought of as “junk DNA,” some studies have identified regulatory and even adaptive advantages that such stretches of DNA confer to their home genome (see review by Zhou et al., 2014).

A Mechanistic Explanation of the Expansion of the IR-LSC Boundary, Loss of trnH-GUG, and the LSC Inversion

The possibility of bidirectional recombination between IRs in magnoliid plastomes is necessary for the model of Asarum plastome rearrangement that we outline below; we envision IR boundary expansion in Asarum to proceed in a stepwise, reciprocal fashion where a boundary disruption of one repeat is corrected against the intact remaining copy via crossing over (Maréchal and Brisson, 2010). We know that trnH-GUG can be impacted by recombination between IR boundaries since Cai et al. (2006) reported its duplication in IR_B of Drimys, another magnoliid genus – this means that expansion of IR_A propagated to IR_B via intramolecular recombination. If both boundaries are unstable, however, then the chance of intramolecular recombination and correction errors being made may increase – we propose that the state of the IR-LSC boundaries of the Saruma plastome predisposed those of Asarum to this very situation.

In stark contrast to the duplication of trnH-GUG in the plastome of Drimys, trnH-GUG has been lost in all sampled Asarum lineages; this represents the first reported loss of a plastid tRNA in the magnoliids. Coincidently, the Saruma plastid copy of trnH-GUG contains two SNPs, one of which results in a codon substitution, relative to the copy in Piper cenocladum; this suggests a change or loss of function in addition to being flanked
by regions of low complexity and high similarity. Direct repeats, which the flanking regions of *trnH-GUG* very closely resemble, are known to have the potential to undergo recombination and yield genomic isomers of the mitochondrial genome that contain less than a single genomic equivalent, so called sublimons (Maréchal and Brisson, 2010). The ability of plastome regions to be excised by intramolecular recombination between approximately 600 bp direct repeats has been demonstrated by Kode et al. (2006) as a way in which to conduct gene knockout experiments. However, the length of the direct repeats in the *trnH-GUG* flanking regions of *S. henryi* that we describe above may not be unduly short; Odahara et al. (2015) demonstrates that the loss of RECG, a recombination suppressing gene in the nucleus, leads to atypical recombination between repeats of only 12 bp in length plastomes of knockout lines – this suggests that if recombination suppression was affected in proto-*Asarum*, that only a few SNP mutations in a highly mutable region would have been required for a heightened chance of *trnH-GUG* excision.

We propose that *trnH-GUG* was lost from the plastome prior to the diversification of *Asarum* due to development of direct repeats in close proximity which developed from highly similar, hypermutable low-complexity flanking regions; these repeats then led to the creation of viable plastome isomers that eventually became more numerous than those containing the tRNA.

In *Saruma*, the expansion of the LSC-proximal portion of IR<sub>A</sub> includes a highly degraded copy of *rps19* (Figure 24). We propose that this subtle expansion of the IR<sub>A</sub> represents the precursor of the IR<sub>A</sub> boundary shifts, *trnH-GUG* loss, and the LSC inversion in *Asarum*. 

143
Figure 24. Alignment of the IR$_B$-LSC boundary of magnoliid plastomes; note the marked expansion of IR$_B$ post loss of trnH-GUG, and subsequent contraction that coincides with the duplication of the SSC. Black portions of sequence are indicative of a lack of comparability between aligned portions.
The development of a low-complexity region due to the propagation of a TAA repeat likely instantiated by an inversion mutation (ATA to TAT) within the duplicated IR_A copy of rps19 evolved in concert with a region of similar complexity upstream of trnH-GUG; the 5’ portions of these regions are extremely similar, and we are unable to completely rule out that this region was duplicated downstream from trnH-GUG.

Alignment of the 3’ end of the rps19 pseudogene and the region downstream of trnH-GUG reveals that these two sequences are highly similar (overall similarity of 78%) and include a 17 bp inverted repeat and no fewer than four direct repeats ranging in length from five to ten bp in length. The sequence separating these direct repeats is comprised solely of A or T, and due to the stochastic nature of the propagation of such low complexity sequence, it is plausible that this intervening sequence could have evolved to further extend these direct repeats between the common ancestor of Asarum and Saruma and extant representatives of the former. Of the explanations for the origin of these two highly-similar regions, we hypothesize that they were initiated independently and converged on one-another due to slipped-strand mispairing.

The ability for intramolecular recombination to operate between IRs suggests that the void created by the excision of the rps19-trnH-GUG region would be corrected against IR_B, yet due to the low GC content of the resulting splice-site this would also result in a novel region of reduced complexity. The remaining necessary steps to produce what we observe today in the plastomes of Asarum could have proceeded in multiple ways; here we will discuss two possibilities. Firstly, the plastome could have existed for some time with the IR boundary within the rps19 reading frame, especially if the
majority of the time it exists as a series of linear chromosomes (Bendich, 2004). Alternatively, and we argue more likely, is that the disruption of the IRA boundary caused by the excision of $trnH$-GUG initiated a sequence correction response that corrected IRA against IRB but resulted in the extension of this region to incorporate $rps19$ through at least a partial copy of $rps14$ in the ancestor of *Asarum* (Figure 24).

The remaining inversion that must be invoked to account for the contemporary arrangement of *Asarum* plastomes is the most difficult syntenic break to explain, but it is hard to imagine that this disruption is not also due to intramolecular recombination due to regions of low complexity. It has long been established that tRNAs are often flanked by GC-poor regions (Chumley et al., 2006; Guisinger et al., 2011), and the $trnE$-$trnD$ syntenic block is no exception (Figure 25; Downie and Jansen, 2014). For example, Kim and Kim (2014) found the $trnY$-$trnE$ intergenic spacer of *Equisetum arvense* to be hypervariable even between geographically isolated variants of the same species; changes in this tRNA-rich syntenic block account for the majority of differences between the compared genomes, and the authors propose that the propagation of a repetitive motif is responsible. While evidence of repeat-driven expansion and degradation of $trnE$-$trnD$ block intergenic spacers is also evident in *Asarum*, the IR$_A$ copy of $rpl14$ in *A. epigynum* is separated from $trnE$ by only 29 bp of GC-rich sequence; this suggests that the inversion of the LSC, such that the $trnE$-$trnD$ block is adjacent to IR$_A$, was either not caused by repeat-mediated recombination or that the repetitive sequence was later replaced via error correction against IR$_B$. The latter is supported by the spasmodic history of the IR$_A$ boundary; if expansions and contractions can occur prior to a rearrangement, it
seems plausible that they can occur afterward as well. Since the longest intact expansion of is found in *A. epigynum*, which is recovered as sister to the remainder of *Asarum* by our previous analyses (Sinn et al., 2015b) and by our rDNA phylogram reported here, this leads us to propose that a single expansion of IR\textsubscript{A} occurred in proto-*Asarum*. One certainty is that the original expansion of IR\textsubscript{A} has degraded and retracted, and has done so differentially throughout the evolution of *Asarum*; it is not clear whether the precursor of this instability was the cause or is a symptom of the LSC inversion.
Figure 25. Alignment of the 5’ flanking region of \textit{trnE}-UUC through the 5’ flank of \textit{psbM} in 11 magnoliid species representing four orders. The 5’ flank of \textit{trnE}-UUC in \textit{Asarum} species is preceded upstream by \textit{rpl16} while this same gene is preceded upstream by \textit{trnT}-GGU in all other published magnoliid plastomes. Note the contrast between the level of conservation of the tRNA genes and their spacer regions with the spacer regions outside of the \textit{trnE}-UUC – \textit{trnD}-GUC syntenic block.
Duplication of the Small Single Copy Region

The SSC region has been hypothesized to confer structural stability to a circularized plastid genome (Knox, 2014; Ruhlman and Jansen, 2014), although it has become clear that the plastome is infrequently observed in circular form when condensed (Bendich, 2004). One of the most interesting implications for a duplication of the SSC region is the novel structure suggested by the absence of single-copy sequence separating the large IR regions found between the LSC and the SSC (Figure 26); the lack of the SSC region means that the plastomes of many Asarum species are largely comprised of IR sequence and are therefore functionally bipartite, a novel condition for angiosperm plastomes. Bipartite plastid genomes have been reported from the Eudicots (see Ruhman and Jansen, 2014 for a thorough review), but those described here are the first to owe their two-parted organization to expansion of the IRs rather than the loss of one.
Figure 26. Possible structures based on stereotypical plastome circularization suggested by assemblies of Asaroideae plastomes.

The duplication of the entirety of the SSC region has, to our knowledge, not been reported elsewhere in the angiosperms. Furthermore, and from a parsimony perspective, depending on whether the presence of this region is plotted on the topology resulting from the analysis of the rDNA cistron, plastid-only or combined matrix the SSC has been
lost gradually then regained in *A. minus* or regained twice in *A. epigynum* and *A. minus*, respectively. In light of the apparent rarity of a duplication event of this magnitude, we find it especially interesting that the plastid phylogeny suggests independent restorations of the SSC in both *Asarum* subgenera *Geotaenium* and *Heterotropa*.

**Episodic Selective Regime Shifts**

In addition to the non-uniformity of plastome structure and synteny across the Asaroideae, we present evidence of the operation of genic- and lineage-specific selective regimes. We notice a general relaxation of purifying selection in various *Asarum* lineages relative to *Saruma*, and find that this pattern is coincident with lineages that are suggested to have undergone increased rates of diversification (see Sinn et al., 2015b). The inconsistency of plastome structure, synteny, and genic content among *Asarum* plastomes coupled with our earlier work identifying asymmetry in diversification rate within the genus suggests that it is likely that no single set of selective pressures is driving the development of these peculiar genomes. We propose that their continued evolution has proceeded semi-independently, and more divergently than is typical for plastomes of closely related species, post establishment of the precursors that we hypothesize to be responsible for their disruption.

In addition to the overall relaxation of purifying selection in *Asarum* inferred by PAML, the presence of several pseudogenes that represent autapomorphies in our phylogenies further suggests a strong decoupling between the evolutionary trajectories across *Asarum* lineages and their plastids. Similarly strong divergences in pseudogene
presence and selective regime across closely-related lineages have been identified in the Lentibulariaceae (Wicke et al., 2013), Campanulaceae (Guisinger et al., 2011), and the Caryophyllaceae (Sloan et al., 2014).

Surprisingly, we found that differences in selective regime do not colocalize with syntenic breaks and hypermutable intragenic spacer regions in Asarum. The locations of selectively divergent genes and pseudogenes present no easily identifiable pattern across the Asaroideae, and we suggest that this represents further evidence of strong decoupling of the evolutionary trajectories amongst these plastomes.

The species whose plastomes are presented here comprise only six percent of the described diversity of the Asaroideae. Increased sampling would not only improve the accuracy of the phylogenetic reconstructions that serve as the framework upon which to test selective regime, but given the incremental changes already inferred from this coarse level of sampling it could also provide for high resolution inference of selection regime shifts and their precursors and effects.

**Summary: the roles of recombination, slipped-strand replication and error repair in the evolution of the plastome**

Many studies and reviews have reported that repetitive and low complexity motifs are commonly found in the adjacent flanking regions of rearrangements, losses and duplications in the plastomes of angiosperms (Aii et al., 1997; Chumley et al., 2006; Cai et al., 2008; Haberle et al., 2008; Wang et al., 2008; Ruhlman and Jansen, 2014). During this same period of time, our understanding of the molecular mechanisms which generate
these potential disruption-associated sequences has greatly increased (Levinson and Gutman, 1987; Maréchal et al., 2009; Vaughn and Bennetzen, 2014; Iyer et al., 2015; Odahara et al., 2015). We have put forth hypotheses through the implication of particular regions and repeats in the syntenic disruption of *Asarum* plastomes, and we agree with Guisinger et al. (2011) and Sloan et al. (2014) that reduction in the efficacy of nuclear-encoded proteins that promote faithful replication and maintenance of organellar genomes coupled with selection on the plastome are likely at the root of the observed rearrangements and divergence of *Asarum* plastomes relative to others currently described from the magnoliids.

While tandem repeats are commonly found in plastomes (Ruhlman and Jansen, 2014), and duplications within close proximity to one another have been reported in many systems (Thomas et al., 2004), the mechanisms responsible for these genomic aberrations are not fully agreed upon (Vaughn and Bennetzen, 2014). Regardless of how repetitive regions are created, we present evidence and suggest models of disruption that are consistent with the work of other researchers (Guisinger et al., 2010; Guisinger et al., 2011) who have hypothesized that recombination between paralogous copies of identical sequence can lead to plastome disruption on an evolutionary timescale, in addition to generating the expected genomic variability that can be found between the genomes of geographic or otherwise isolated populations of the same species (Gurdon and Maliga, 2014; Kim and Kim, 2014).

Recent investigations involving the plastomes of multiple taxa interpreted within a phylogenetic framework, the present study included, have revealed that signal of
positive or relaxation of purifying selection can be found in the plastomes of disparate lineages (Wicke et al., 2013; Barrett et al., 2014; Sloan et al., 2014; Hu et al., 2015); this suggests that relaxation of error correction in the plastome may not only correspond to increased disruption, but that disruption itself might lead to novel function that can become fodder for selective processes. The discovery here, and elsewhere (Barrett et al., 2014), that wildly different selective regimes can be simultaneously influencing the plastomes of closely related species suggests that relatedness is not sufficient to predict the state of the plastome; furthermore, this suggests that the chloroplast plays a more prominent role in the realized niche of its host plant than we are commonly led to believe (Bock et al., 2014; Hu et al., 2015). Future investigation of the state of proteins and pathways that are implicated in recombinatory repair of the organellar genomes may be particularly fruitful avenues of research from which to further our understanding of the constraints and predispositions of plastome plasticity and the role of this genome in, or as a recorder of, speciation (Maréchal and Brisson, 2010, Bock et al., 2014).

Although the association of hypermutable (Magee et al., 2010) and repetitive sequence has been directly implicated in the syntenic disruption of plastomes by many studies (Chumley et al., 2006; Guisinger et al., 2010; Guisinger et al., 2011; Knox, 2014), the density of taxon sampling has often been insufficient to provide clear polarity of change, that is, to establish the pleisiomorphic state of the affected flanking regions just prior to rearrangement. Isolating and describing the role of individual regions of low complexity or repetitive sequence in many lineages is crucial to better developing our understanding of plastome disruption at an evolutionary timescale; this can most
effectively be accomplished in non-model organisms through dense taxon sampling and in a phylogenetic framework. Only when we have sequenced plastomes from the majority of taxa in multiple, disparate lineages will we be able to understand whether atypical plastome features are precursors or symptoms of plastome disruption. We have found that the evolutionary history of *Asarum* and *Saruma* has provided one such case, and future sequencing efforts in the group promise to further develop our understanding of the precursors of plastome change.
Literature Cited


Armbruster, W. S. 2014. Floral specialization and angiosperm diversity: phenotypic
divergence, fitness trade-offs and realized pollination accuracy. AoB Plants 6: plu003.


Canner, J. E., R. R Dunn, I. Giladi, and K. Gross, 2012. Redispersal of seeds by a


Drummond, A. J., B. Ashton, S. Buxton, M. Cheung, A. Cooper, C. Duran, M. Field, J.
Heled, M. Kearse, S. Markowitz, R. Moir, S. Stones-Havas, S. Sturrock, T.
Auckland, NZ.

phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and
Evolution*.  http://dx.doi.org/10.1093/molbev/mss075.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high

innovations in an early branching angiosperm lineage (Annonaceae;

Faegri, K. and L. Van Der Pijl. 1979. The principles of pollination biology, third ed.

sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements

Farris, J. S., V. A. Albert, M. Källersjö, D. Lipscomb, and A. G. Kluge. 1996. Parsimony

Felsenstein, J. 1981. Evolutionary trees from DNA Sequences: a maximum likelihood
approach. *Journal of Molecular Evolution* 17: 368-376.


Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. *Journal of Molecular Evolution* 70: 9317.


Knox, E. B. 2014. The dynamic history of plastid genomes in the Campanulaceae sensu


http://dx.doi.org/10.1186/1471-2148-14-82.


Linnaeus, C. 1753. Species Plantarum. Vol. 1


Rafinesque, C. S. 1825. *Neogenyton*, or indication of sixty-six new genera of plants of North America. pg. 3. Published by the author.


Schneider, H., L.-J. He, J. Marquardt, L. Wang, J. Heinrichs, S. Hennequin, and X.-C.


Silvestro, M. 2012. raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity and Evolution* 12, 335-337.


Sinn, B. T., L. M. Kelly, and J. V. Freudenstein. 2015a. Phylogenetic relationships in


Tavaré, S. 1986. *Some probabilistic and statistical problems in the analysis of DNA*
sequences. In: Miura, R. M. [Ed.], Some mathematical questions in biology – DNA sequence analysis, 57-86. Providence, USA.


Weakley, A. S. 2012. Flora of the Southern and Mid-Atlantic States. Published by the author, Chapel Hill.


Weakley, A. S. 2012. Flora of the Southern and Mid-Atlantic States (Draft – 30
November 2012). 1225 pp. Available from:


Williamson, E. M., A. Lorenc, A. Booker, and N. Robinson. 2013. The rise of traditional


Yamaji, H., J. Yokoyama, H. Ohashi, and M. Maki. 2006. Concordant clines and


Zhou, K., A. Aertsen, and C. W. Michiels. 2014. The role of variable DNA tandem
Appendix A. Herbarium and GenBank voucher information for material used in Chapter 1.

**Taxon**: Collector initials and number; GenBank accession number: *rpoB-trnC^GCA*, *rps16-trnK*, 

*trnL^UAA* exon – *trnF^GAA*, *trnT^UGU* – *trnL^UAA*, *trnL^UAA* – *trnL^UAA* exon, *trnS^UGA* – 

*trnFM^CAU*, *ycf1* 3850-5310, rDNA; *Voucher specimen*, Collection locale; 

Herbarium. Regions that failed to amplify are represented by ---.

*Asarum arifolium* Michx. var. *arifolium*; BTS 1062; KJ671647, KJ888434, KJ888542, KJ888599, KJ888655, KJ888758, KJ888708, KJ888487; United States, North Carolina; 

KJ888450, KJ888559, KJ888617, KJ888672, KJ888772, KJ888725, KJ888505; Japan;
KEW. *A. hartwegii* S.Watson; JVF 2928; KJ671661, KJ888451, KJ888561, KJ888619,
KJ888674, ---, KJ888727, KJ888507; OS. *A. hatsushimae* F.Maek. ex Hatus. &
Yamahata; LK 1040; KJ671639, KJ888452, KJ888562, KJ888620, KJ888675,
KJ888773, KJ888728, KJ888508; Japan, Kagoshima Pref.; KEW. *A. heterophyllum*
Ashe; BTS 1058; KJ671658, KJ888454, KJ888564, KJ888621, KJ888676, KJ888774,
KJ888730, KJ888510; United States, South Carolina; OS. BTS 1076A; KJ671649,
KJ888455, KJ888565, KJ888622, KJ888677, KJ888775, KJ888731, KJ888455; United
States, South Carolina; OS. *A. hexalobum* (F.Maek.) F.Maek. var. perfectum F.Maek.;
USDA 47938*J; KJ671676, KJ888466, KJ888577, KJ888634, KJ888689, KJ888786, ---,
KJ888522; collection locale unknown; OS. *A. himalaicum* Hook.f. & Thomson ex
Klotzsch; LK 942; KJ671637, KJ888456, KJ888566, KJ888623, KJ888678, KJ888776,
KJ888732, KJ888511; China, Sichuan Prov.; BH. *A. kumageanum* Masam.; PJZ;
KJ671682, KJ888457, KJ888567, KJ888624, KJ888679, KJ888777, KJ888733,
KJ888512; horticulture, collection locale unknown. *A. lemmonii* S.Watson; UC Bot
Garden 88-0824; KJ671629, KJ888458, KJ888568, KJ888625, KJ888680, KJ888778,
KJ888734, KJ888513; United States, California; OS. *A. lewisii* Fernald; BTS 1138;
KJ671660, ---, KJ888569, KJ888626, KJ888681, ---, KJ888735, KJ888514; United
States, North Carolina; OS. *A. maximum* Hemsl.; USDA 70824*H; KJ671673,
KJ888459, KJ888570, KJ888627, KJ888682, KJ888779, KJ888736, KJ888515;
collection locale unknown; OS. *A. megacalyx* (F.Maek.) T.Sugaw.; USDA 58437*L;
KJ671672, KJ888460, KJ888571, KJ888628, KJ888683, KJ888780, KJ888737,
- KJ888584, KJ888640, KJ888695, ---, KJ888747, KJ888529; collection locale unknown; OS. *A. shuttleworthii* Britten & Baker var. *harperi* (Gaddy) Barringer;
FC91; KJ671680, ---, KJ888560, KJ888618, KJ888673, ---, KJ888726, KJ888506;
United States, Alabama, OS. *A. shuttleworthii* Britten & Baker var. *shuttleworthii*;
BTS 1103A; KJ671656, KJ888472, KJ888583, KJ888639, KJ888694, KJ888792, ---,
KJ888528; United States, North Carolina; OS. BTS 1121A; KJ671657, KJ888431,
KJ888539, KJ888596, KJ888652, KJ888755, KJ888705, KJ888484; United States, North
Carolina; OS. BTS 1125A; KJ671658, KJ888432, KJ888540, KJ888597, KJ888653,
KJ888756, KJ888706, KJ888485; United States, North Carolina; OS. *A. sp.;* USDA
39518*H; KJ671663, KJ888473, KJ888585, KJ888641, KJ888696, KJ888793,
KJ888748, KJ888473; collection locale unknown; OS. *A. sp. nov. 2;* BTS 1137;
KJ671685, KJ888468, KJ888579, KJ888636, KJ888690, KJ888788, KJ888743,
KJ888524; United States, South Carolina; OS. *A. speciosum* (R.M.Harper) Barringer;
FC 105; KJ671677, KJ888474, KJ888586, KJ888642, KJ888697, KJ888794, KJ888749,
KJ888530; United States, Alabama, OS. *A. splendens* (F.Maek.) C.Y.Cheng &
C.S.Yang; USDA 74149*H; KJ671674, KJ888476, KJ888588, KJ888644, KJ888698, ---,
KJ888750, KJ888531; collection locale unknown; OS. USDA 74150*H; KJ671675,
KJ888477, KJ888589, KJ888645, KJ888699, ---, KJ888751, KJ888532; collection locale
unknown; OS; PJZ; KJ671686, KJ888478, KJ888590, KJ888646, KJ888700, ---,
KJ888796, KJ888533; collection locale unknown. USDA 58430; KJ671670, KJ888477,
KJ888587, KJ888643, ---, KJ888795, ---, KJ888475; collection locale unknown; OS. *A.
takaoi* F.Maek.; LK 1041; KJ671641, KJ888479, KJ888591, KJ888647, ---, ---,
KJ888752, KJ888534; Japan, Kyoto Pref.; KEW. *A. virginicum* L.; BTS 1128B;
KJ671659, KJ888433, KJ888541, KJ888598, KJ888654, KJ888757, KJ888707,
KJ888486; United States, Virginia; OS. *A. yakusimense* Masam.; USDA 40201*J;
KJ671665, KJ888483, KJ888595, KJ888651, KJ888704, KJ888800, KJ888754,
KJ888538; collection locale unknown; OS. *Saruma henryi* Oliv.; USDA 49482*J;
KJ671667, KJ888453, KJ888563, ---, ---, ---, KJ888729, KJ888509; collection locale
unknown; OS.
Asarum shuttleworthii var. shuttleworthii 1125A

Asarum naniflorum PZ
Asarum rhombiformis
Asarum contractum PZ
Asarum shuttleworthii var. shuttleworthii 1121A
Asarum minus 1060
Asarum sp. nov. 1
Asarum heterophyllum 1058
Asarum heterophyllum 1076A
Asarum minus 1082B
Asarum virginicum
Asarum minus 53052
Asarum shuttleworthii var. shuttleworthii 1103B
Asarum minus 1083
Asarum contractum 1089A
Asarum memmingeri 1093A
Asarum memmingeri 1092A
Asarum sp. nov. 2
Asarum shuttleworthii var. harperi
Asarum arifolium var. ruthii
Asarum arifolium var. arifolium
Asarum lewisii
Asarum speciosum
Asarum delavayi
Asarum splendens 74149
Asarum splendens PZ
Asarum splendens 74150
Asarum splendens 59430
Asarum maximum
Asarum takaoi
Asarum minamitanianum
Asarum savatieri subsp. savatieri
Asarum species 39518
Asarum asperum
Asarum hexalobum var. perfectum
Asarum megacalyx
Asarum blumei
Asarum muramatsui
Asarum asarooides
Asarum hatsushimae
Asarum gelatinum
Asarum crassum
Asarum kumageanum
Asarum fusinoi
Asarum yakusimense
Asarum forbesii
Asarum sieboldii var. sieboldii
Asarum lemonii
Asarum hartwegii
Asarum caudatum
Asarum canadense
Asarum debile
Asarum caudigereellum
Asarum caulescens
Asarum himalaicum
Asarum pulchellum
Asarum epigynum
Saruma henryi
Asarum memmingeri 1092A
Asarum memmingeri 1093A
Asarum virginicum
Asarum minus 1082B
Asarum shuttleworthii var. shuttleworthii 1121A
Asarum shuttleworthii var. shuttleworthii 1125A
Asarum contractum 1089A
Asarum minus 53052
Asarum contractum PZ
Asarum sp. nov. 1
Asarum shuttleworthii var. harperi
Asarum heterophyllum 1058
Asarum minus 1060
Asarum naniflorum
Asarum sp. nov. 2
Asarum rhombiformis
Asarum shuttleworthii var. shuttleworthii 1103B
Asarum minus 1083
Asarum lewisii
Asarum asaroides
Asarum muramatsui
Asarum fusinoi
Asarum gelatinum
Asarum hatsushimae
Asarum kumageanum
Asarum yakusinense
Asarum forbesii
Asarum savatieri subsp. savatieri
Asarum takaoi
Asarum asperum
Asarum minimilianum
Asarum megacalyx
Asarum hexalobum var. perfectum
Asarum crassum
Asarum blumei
Asarum delavayi
Asarum splendens 74149
Asarum splendens 74150
Asarum splendens PZ
Asarum splendens 58430
Asarum maximum
Asarum arifolium var. arifolium
Asarum arifolium var. ruthii
Asarum speciosum
Asarum sieboldii var. sieboldii
Asarum caudatum
Asarum lemontii
Asarum hartwegii
Asarum canadense
Asarum caudigerellum
Asarum debile
Asarum pulchellum
Asarum himalayicum
Asarum caulescens
Asarum epigynum
Saruma henryi
S14
Appendix B. Herbarium and GenBank voucher information for
material used in Chapter 2. Accessions denoted with an asterisk were
used in our RAxML analysis; all accessions were used for BCI.

Taxon; Collector initials and number; GenBank accession number: rpoB-trnC\(^{GCA}\), rps16-
trnK,

\(trnL\(^{UAA}\)\) exon – \(trnF\(^{GAA}\)\), \(trnT\(^{UGU}\) – \(trnL\(^{UAA}\)\), \(trnL\(^{UAA}\) – \(trnL\(^{UAA}\)\) exon, \(trnS\(^{UGA}\) -
\(trnFM\(^{CAU}\)\), \(ycf1\) 3850-5310, rDNA; Voucher specimen, Collection locale;
Herbarium. Missing data is represented by ---. Unknown values at time of
submission are represented by ???.

*Asarum arifolium Michx. var. arifolium*; BTS 1062; KJ671647, KJ888434, KJ888542,

*Asarum arifolium Michx. var. ruthii (Ashe) Barringer*; BTS 1067A; KJ671648,
KJ888470, KJ888581, KJ888638, KJ888692, KJ888790, KJ888745, KJ888526; United
States, North Carolina.

Asarum asperum F.Maek var. asperum*; LK 1037; KJ671664, KJ888436, KJ888544, KJ888601, KJ888657, KJ888760, ---, KJ888489; 89-4015, Japan, Osaka Pref.

Asarum blumei Duch.*; LK 1042; KJ671642, KJ888437, KJ888545, KJ888602, KJ888658, KJ888761, KJ888710, KJ888490; Japan, Shizuoka Pref.

Asarum canadense L.*; JVF 2392; KJ671662, KJ888438, KJ888546, KJ888603, KJ888659, KJ888762, KJ888711, KJ888491; United States, Missouri.

Asarum caudatum Lindl*.; UC Bot Garden 64-1174; KJ671627, KJ888439, KJ888547, KJ888604, KJ888660, KJ888763, KJ888712, KJ888492; United States, California.


Asarum caulescens Maxim. *; LK 937; KJ671636, KJ888441, KJ888549, KJ888606, KJ888662, KJ888765, KJ888714, KJ888494; China, Sichuan Prov.


Asarum crassum F.Maek. *; LK 1046; KJ671643, KJ888445, KJ888554, KJ888611, KJ888667, KJ888768, KJ888719, KJ888499; Japan, Kagoshima Pref.

Asarum debile Franch. *; LK 936; KJ671635, KJ888446, KJ888555, KJ888612, ---, KJ888769, KJ888720, KJ888500; China, Sichuan Prov.

Asarum delavayi Franch. *; USDA 58431*H; KJ671671, ---, ---, KJ888613, KJ888668, ---, KJ888721, KJ888501; collection locale unknown.

Asarum epigynum Hayata*; LK 1049 ; KJ671644, KJ888447, KJ888556, KJ888614, KJ888669, KJ888770, KJ888722, KJ888502; Taiwan.
Asarum forbesii Maxim. *; LK 681; KJ671631, KJ888448, KJ888557, KJ888615, KJ888670, KJ888771, KJ888723, KJ888503; China, Zhejiang Prov.

Asarum fudsinoi T.Itô*; LK 684; KJ671632, KJ888449, KJ888558, KJ888616, KJ888671, ---, KJ888724, KJ888504; Japan, Kagoshima Pref.

Asarum gelasinum (F.Maek.) Hatus. *; LK 1038; KJ671639, KJ888450, KJ888559, KJ888617, KJ888672, KJ88872, KJ888725, KJ888505; Japan.

Asarum hartwegii S.Watson*; JVF 2928; KJ671661, KJ888451, KJ888561, KJ888619, KJ888674, ---, KJ888727, KJ888507; ???.

Asarum hatsushimae F.Maek. ex Hatus. & Yamahata*; LK 1040; KJ671639, KJ888452, KJ888562, KJ888620, KJ888675, KJ888773, KJ888728, KJ888508; Japan, Kagoshima Pref.

Asarum hexalobum (F.Maek.) F.Maek. var. perfectum F.Maek. *; USDA 47938*J; KJ671676, KJ888466, KJ888577, KJ888634, KJ888689, KJ888786, ---, KJ888522; collection locale unknown.


Asarum kumageanum Masam. *; PJZ; KJ671682, KJ888457, KJ888567, KJ888624, KJ888679, KJ888777, KJ888733, KJ888512; collection locale unknown.

Asarum lemmonii S.Watson*; UC Bot Garden 88-0824; KJ671629, KJ888458, KJ888568, KJ888625, KJ888680, KJ888778, KJ888734, KJ888513; United States, California.

Asarum lewisii Fernald*; BTS 1138; KJ671660, ---, KJ888569, KJ888626, KJ888681, ---, KJ888735, KJ888514; United States, North Carolina.

Asarum maximum Hemsl. *; USDA 70824*H; KJ671673, KJ888459, KJ888570, KJ888627, KJ888682, KJ888779, KJ888736, KJ888515; collection locale unknown.


*Asarum muramatsui* Makino; USDA 45326*P; KJ671666, KJ888464, KJ888575, KJ888632, KJ888687, KJ888784, KJ888740, KJ888520; collection locale unknown.
Asarum naniflorum B.T. Sinn*; PJZ; KJ671684, KJ888465, KJ888576, KJ888633, KJ888688, KJ888785, KJ888741, KJ888521; collection locale unknown.

Asarum pulchellum Hemsl. *; LK 890; KJ671634, KJ888467, KJ888578, KJ888635, ---, KJ888787, KJ888742, KJ888523; China, Sichuan Prov.


Asarum savatieri Franch. subsp. savatieri*; LK 1041; KJ671640, KJ888471, KJ888582, ---, KJ888692, KJ888791, KJ888746, KJ888527; Japan, Shizuoka Pref.

Asarum sieboldii Miq. var. sieboldii*; USDA 56474*DH; KJ671669, ---, KJ888584, KJ888640, KJ888695, ---, KJ888747, KJ888529; collection locale unknown.

Asarum shuttleworthii Britten & Baker var. harperi (Gaddy) Barringer*; FC; KJ671680, ---, KJ888560, KJ888618, KJ888673, ---, KJ888726, KJ888506; United States, Alabama.

Asarum shuttleworthii Britten & Baker var. shuttleworthii; BTS 1103A*; KJ671656, KJ888472, KJ888583, KJ888639, KJ888694, KJ888792, ---, KJ888528; United States,
Asarum sp. *; USDA 39518*H; KJ671663, KJ888473, KJ888585, KJ888641, KJ888696, KJ888793, KJ888748, KJ888473; collection locale unknown.

Asarum sp. nov. 2*; BTS 1137; KJ671685, KJ888468, KJ888579, KJ888636, KJ888690, KJ888788, KJ888743, KJ888524; United States, South Carolina.


Asarum takaoi F.Maek. *; LK 1041; KJ671641, KJ888479, KJ888591, KJ888647, ----, -- -, KJ888752, KJ888534; Japan, Kyoto Pref.


Asarum yakusimense Masam. *; USDA 40201*J; KJ671665, KJ888483, KJ888595, KJ888651, KJ888704, KJ888800, KJ888754, KJ888538; collection locale unknown.

Saruma henryi Oliv. *; USDA 49482*J; KJ671667, KJ888453, KJ888563, ----, ----, ----, KJ888729, KJ888509; collection locale unknown.
Appendix C. A new combination for *Asarum sorriei*.

Appendix D. Primers designed for scaffolding of *Asarum* plastome contigs.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>33327F</td>
<td>AAACAGGATGGCCGGATCTT</td>
</tr>
<tr>
<td>33942R</td>
<td>CTCTATGGATTCTCTCTAATTGACTCCT</td>
</tr>
<tr>
<td>83371F</td>
<td>ACGGATCGTGGCTAAAGATTCA</td>
</tr>
<tr>
<td>83758R</td>
<td>TTGGACATTCTACAGAAGGATTTTCAC</td>
</tr>
<tr>
<td>15344F</td>
<td>CGCTCTCCTCTTTATAGATAGGACT</td>
</tr>
<tr>
<td>15904R</td>
<td>GTCGTGGCATTACGGCTTTTA</td>
</tr>
<tr>
<td>126696F</td>
<td>TGCAATGTGCTATGGCTTC</td>
</tr>
<tr>
<td>127167R</td>
<td>GCCACCTAAAAACAAAGTACATGCT</td>
</tr>
<tr>
<td>185732F</td>
<td>GACTCCGACACGCACAGAGTAC</td>
</tr>
<tr>
<td>185732R</td>
<td>TGGAATGTCTGGATTTCCAGTATCC</td>
</tr>
<tr>
<td>45989F</td>
<td>CCCGAATCATCAGTCTCTTTTCT</td>
</tr>
<tr>
<td>46532R</td>
<td>CGGGAGAAAAGGAGGCATT</td>
</tr>
<tr>
<td>10181F</td>
<td>TAGCAATCCGCCGCTTTAGT</td>
</tr>
<tr>
<td>11128R</td>
<td>CCGTGGGAGGACAGAATAGG</td>
</tr>
</tbody>
</table>