A First Complete Phylogenomic Hypothesis for Diploid Blueberries (*Vaccinium* section *Cyanococcus*)

Andrew A. Crowl[1,2,*]; Peter W. Fritsch[3]; George P. Tiley[2,4]; Nathan P. Lynch[1,5]; Thomas G. Ranney[1,5]; Hamid Ashrafi[1]; Paul S. Manos[2]

[1]Department of Horticultural Science, North Carolina State University, Raleigh, NC 27607, USA
[2]Department of Biology, Duke University, Durham, NC 27708, USA
[3]Botanical Research Institute of Texas, Fort Worth, TX 76107, USA
[4]Royal Botanic Gardens Kew, Richmond TW9 3AE, UK
[5]Department of Horticultural Science, North Carolina State University, Mountain Horticultural Crops Research & Extension Center, Mills River, NC 28759, USA
[*]Author for correspondence: andy.crowl@gmail.com

Manuscript received _____; revision accepted _____.

Running head: Phylogeny of diploid blueberries

ABSTRACT

The premise of the study:
The true blueberries, (*Vaccinium* sect. *Cyanococcus*; Ericaceae), endemic to North America, have been intensively studied for over a century. However, with species estimates ranging from 9 to 24 and much confusion regarding species boundaries, this ecologically and economically valuable group remains inadequately understood at a basic evolutionary and taxonomic level. As a first step toward understanding the evolutionary history and taxonomy of this species complex, we present the first phylogenomic hypothesis of the known diploid blueberries.
Methods:
We used flow cytometry to verify the ploidy of putative diploid taxa and a target-enrichment approach to obtain a genomic dataset for phylogenetic analyses.
Results:
Despite evidence of gene flow, we found that a primary phylogenetic signal is present. Monophyly for all morphospecies was recovered, with two notable exceptions: one sample of *V. boreale* was consistently nested in the *V. myrtilloides* clade and *V. caesariense* was nested in the *V. fuscatum* clade. One diploid taxon, *Vaccinium pallidum*, is implicated as having a homoploid hybrid origin.
Conclusions:
This foundational study represents the first attempt to elucidate evolutionary relationships of the true blueberries of North America with a phylogenomic approach and sets the stage for multiple avenues of future study such as a taxonomic revision of the group, the verification of a homoploid hybrid taxon, and the study of polyploid lineages within the context of a diploid phylogeny.

Keywords
Alleles; Ericaceae; Homoploid hybridization; HybSeq; Phasing; Phylogenetics; Target enrichment; *Vaccinium*

INTRODUCTION

A ubiquitous component of heathlands and other acidophilic plant communities, as well as a food source for wildlife and humans, the true blueberries (*Vaccinium* section *Cyanococcus* A. Gray; henceforth "*Cyanococcus*") are of immense ecological and economic value. Commercially cultivated blueberries originate from this group—representing one of only a handful of widely cultivated plants originating in North America. Despite its economic importance, *Cyanococcus* has suffered from conflicting taxonomies with poorly defined species boundaries and little investigation into the evolutionary history of wild populations.

Cyanococcus is a reticulate species complex of ca. 9–24 species comprising diploids ($2n = 2x = 24$), tetraploids, and hexaploids distributed across much of temperate North America (Fig. 1). The section is easily distinguished from other sections of *Vaccinium* L. by several unique or otherwise diagnostic characters, e.g., verrucose branchlets, articulated pedicels, awnless anthers, and pseudo-10-locular berries (Camp, 1945; Vander Kloet, 1983). In addition to morphological characters, the available molecular data suggest that the group forms a clade (Kron et al., 2002; Crowl et al., unpublished), although sufficient sampling has yet to be undertaken to satisfactorily test monophyly.

Cyanococcus served as a model system during the Modern Synthesis (Huxley, 1942), playing a pivotal role in furthering our understanding of polyploidy and expanding the scope of the movement to include plants. Toward the goal of crop improvement, W.H. Camp and colleagues (Camp, 1942, 1945; Camp and Gilly, 1943; Darrow and Camp, 1945) used data from morphology, crossing studies, genetics, and cytology to

propose a complex series of ancestor-descendant polyploid species relationships in *Cyanococcus*, some through autopolyploidy, others through allopolyploidy. In some cases, Camp (1945) documented size differences correlated with ploidy, such as larger stature and flowers, which has recently been confirmed in one mixed diploid and tetraploid population (Poster et al., 2017). Finally, by equating artificially produced hybrid progeny with morphologically similar plants in the wild, Camp concluded that natural hybrids are rampant among blueberry species, although a strong triploid block, now well known among plant breeders (e.g., Lyrene et al., 2003), was seen to inhibit the viability of progeny with odd-numbered sets of chromosomes.

Subsequently, S.P. Vander Kloet revised Camp's taxonomy in the context of morphological phenetics. The most consequential of Vander Kloet's conclusions from this work was the supposition that all *Cyanococcus* species greater than 1 m tall ("highbush") have been derived from a genetic amalgamation of mostly diploid species less than 1 m tall ("lowbush"), thus forming a "compilospecies" (Harlan and de Wet, 1963) of multiple origins and of variable ploidy (Vander Kloet, 1980, 1983, 1988). In this context, Vander Kloet aggregated 12 of Camp's species into a single highly variable highbush blueberry, *V. corymbosum* L. Although many authors have questioned this extremely broad concept based on habit, leaf, flower, and stem morphology, phenology, and ecology (e.g., Uttal, 1987; Weakley, 2020; Fritsch et al., in press), this taxonomic view of *Cyanococcus* is currently considered the standard, having been adopted by the USDA, plant breeders, and many local and regional floras, including the *Flora of North America* (Vander Kloet, 2009).

93 Much prior research on *Cyanococcus* has highlighted the challenges involved in
94 disentangling this group, but more recent research suggests that the prospects are hopeful
95 for resolving longstanding questions regarding its species composition, patterns of
96 speciation, and evolutionary history (Fritsch et al., in press). In this respect, the rapid
97 maturation of genomic approaches to the study of complex groups of organisms affords a
98 timely opportunity to revisit the evolution of the true blueberries. The multiple ploidy
99 levels inherent in *Cyanococcus*, the group's ecological and economic importance, and the
100 genomic resources now available make *Cyanococcus* an ideal system for understanding
101 polyploidy and cryptic speciation in flowering plants. Surprisingly, however, the
102 evolution of the group as a whole has yet to be studied with such approaches. This has
103 left *Cyanococcus* in an unsatisfactory state to both evolutionary biologists and plant
104 breeders alike.
105 Here we provide a first glimpse into the evolutionary history of *Cyanococcus* with
106 genomic data by reconstructing a diploid phylogeny with genomic data from hundreds of
107 nuclear loci, with flow cytometry analyses conducted to verify ploidy of all currently
108 recognized putative diploid taxa. Our results will be useful for future study of polyploid
109 *Cyanococcus* lineages and updating the taxonomy of this important group of plants.
110
111 **MATERIALS AND METHODS**
112 ***Flow cytometry***
113 Ploidy was estimated with flow cytometry at the Mountain Horticultural Crops
114 Research and Extension Center (North Carolina, USA). Leaf samples were quickly dried
115 in the field with silica gel. This dried tissue (approximately 1.5 cm^2) was finely chopped

with a razor blade in a Petri dish with 400 mL of nuclei extraction buffer (CyStain UV Precise P Nuclei Extraction Buffer, Sysmex Partec, Görlitz, Germany). The solution was incubated for 1 to 2 min at approximately 24°C and then filtered through Partec CellTrics disposable filters with a pore size of 50 µm to remove tissue debris. Nuclei were stained with 1.6 mL of 4',6-Diamidino-2-phenylindole (DAPI) staining buffer (CyStain UV Precise P Staining Buffer, Sysmex Partec). Stained nuclei were analyzed with a flow cytometer (Partec PA-II, Partec) to determine relative genome size. Counts exceeded a minimum of 3000 cells per sample and two subsamples were run for each sample. Genome sizes were determined by comparing mean relative fluorescence of each sample with an internal standard, *Pisum sativum* L. 'Ctirad,' with a known genome size of 8.76 pg (Doležel et al., 2007) and calculated as: 2C genome size of sample = 8.76 pg × (mean fluorescence value of sample/mean fluorescence value of standard). The validity of this method for estimating ploidy levels in *Vaccinium* has been previously demonstrated (with fresh leaf material) by Hummer et al. (2015) and Costich et al. (1993), the latter showing that an observed increase in nuclear DNA content is concurrent with an equivalent increase in ploidy.

Sampling and sequencing

We sampled 36 *Cyanococcus* individuals, each from different natural populations, representing eight putative diploid species (Appendix S1; see Supplementary Data with this article). Species determination followed the morphospecies concepts summarized in Weakley (2020) in addition to the *V. boreale* I.V. Hall & Aalders concept of Vander Kloet (1988). Three additional taxa, *V. arboreum* Marshall (*Vaccinium section*

139 *Batodendron*), *V. macrocarpon* Aiton (*Vaccinium sect. Oxycoccus*), and *V. stamineum* L.
140 (*Vaccinium sect. Polycodium*) comprised the outgroup.
141 DNA extractions were carried out with a modified CTAB approach for all
142 samples (Doyle and Doyle, 1987). The concentration of DNA from extractions was
143 quantified with a Qubit 2.0 (Invitrogen, Carlsbad, California, USA) and the Qubit dsDNA
144 Broad Range Assay Kit as per the manufacturer's recommendations. Samples ranging
145 from 115 to 3000 ng of DNA were sent to Arbor Biosciences (Ann Arbor, Michigan,
146 USA) for library preparation and DNA sequencing on a NovaSeq S4 sequencer (Illumina,
147 San Diego, California, USA) with 2x150 bp chemistry. The Angiosperms353 v1 target
148 capture kit (Johnson et al., 2019) was used for targeted enrichment of each sample.
149
150 ***Sequence data processing***
151 Raw sequences were filtered and processed with the Trim Galore wrapper script
152 (v.0.6.5), which uses Cutadapt (v.2.6; Martin, 2011) and FastQC (v.0.11.9; Andrews,
153 2010) to trim adapters and low-quality reads based on a given Phred quality score cutoff
154 (-q 20). Consensus read assembly for target loci was performed with the default settings
155 in HybPiper v1.3.1 (Johnson et al., 2016). Following the recommendations of McLay et
156 al. (2021), we included available Ericales sequences in the target reference file in
157 addition to the standard Angiosperms353 targets to improve the recovery of targeted loci.
158 Supercontig sequences were then assembled with the *intronerate.py* script available as a
159 part of HybPiper. To screen for potential paralogs, we identified loci/samples in which
160 multiple contigs were generated during the assembly step with the
161 *paralog_investigator.py* script. All loci in which a paralog was suspected were removed

from the dataset. The remaining consensus reads were used as the reference to generate both IUPAC and allele datasets (see below).

Allele phasing

HybSeq data is typically processed in a way that results in single consensus sequences for loci, thus ignoring allelic variation (Andermann et al., 2018; Tiley et al., 2021). However, allelic data may be important in the estimation of species networks when gene flow among taxa is present (Tiley et al., 2021). To include this variation, we employed the recently developed bioinformatics pipeline PATÉ (Tiley et al., 2021) to phase alleles. The pipeline uses consensus loci (in this case, supercontig sequences) created with HybPiper as reference sequences and Illumina reads are mapped back to these loci using the BWA-MEM algorithm from BWA (Li and Durbin, 2009). Variant calling is carried out at the ploidy level determined by flow cytometry for each individual using the HaplotypeCaller program from GATK (McKenna et al., 2010). Potentially erroneous variant calls are filtered out based on the following parameters outlined in DePristo et al. (2011) with the VariantFiltration program in GATK: (1) QD < 2.0, (2) FS > 60.0, (3) MQ < 40.0, (4) ReadPosRankSum < 8.0. We also remove variants present on less than 5% or more than 95% of reads (AF < 0.05 || AF > 0.95) and variants with a depth less than 10 reads (DP < 10). The resulting vcf file for each individual is passed to H-PoPG (Xie et al., 2016) for allele phasing, which solves for the specified number of haplotypes that minimizes the number of switch errors among the reads present in the BAM file using a dynamic programming solution. PATÉ then takes variants from the largest phase block, combines them with sequences from regions of the locus that could

not be phased because of insufficient read overlap, and replaces them with ambiguity codes so that the resulting alleles are the same length as the original consensus loci, similar to previous phasing strategies exclusive to diploids (Kates et al. 2018). PATÉ additionally provides full IUPAC sequences in which all heterozygous sites are replaced by ambiguity codes, which were analyzed alongside individual allele sequences.

Maximum likelihood analyses

Alignments were carried out with FSA (Bradley et al., 2009). To reduce potential issues with missing data and poorly aligned ends, we removed alignment columns containing more than 50% missing data. Individual IUPAC gene trees and allele trees were constructed with IQ-TREE v.1.6.9 (Nguyen et al., 2015). ModelFinder Plus was used to first select the best model for each locus. To assess topological support, we implemented the ultrafast bootstrap approximation UFBoot2 (Hoang et al., 2018) with 1000 replicates in which sites within partitions (loci) were resampled, an approach that is similar to the standard nonparametric bootstrap.

A concatenated alignment was produced for the IUPAC dataset with the *pxcat* command in Phyx (Brown et al., 2017). A partitioned phylogenetic analysis, where partitions were individual loci, was performed with IQ-TREE. The best-fit partitioning scheme was chosen with the PartitionFinder algorithm (*-m TESTMERGE*; Lanfear et al., 2012) implemented in IQ-TREE. A relaxed clustering algorithm (*-rcluster* 10; Lanfear et al., 2014) was implemented to only consider the top 10% of partitioning schemes. As above, 1000 ultrafast bootstrap replicates were performed to assess support.

208 ***Species tree analyses***

209 Multiple species-tree methods were used to estimate a diploid species tree for
210 *Cyanococcus*. Singular value decomposition quartet species-tree estimation
211 (SVDquartets; Chifman and Kubatko, 2014) implemented in Paup* (v.4a142; Swofford,
212 2002) was run on the concatenated IUPAC data matrix, all possible quartets were
213 evaluated, and support was assessed with 100 bootstrap replicates. We also used
214 ASTRAL-III (v.5.5.6; Zhang et al., 2018) on the individual IUPAC gene trees and allele
215 trees. Alleles were assigned to individuals or species with the allele mapping (-a) option.
216 We additionally used STACEY (Jones, 2017), available as part of the BEAST2 package
217 (Bouckaert et al., 2014), to estimate a species tree from the IUPAC and allele datasets in
218 a Bayesian framework. Substitution models, clock models, and gene trees were unlinked
219 for all loci. The birth-death-collapse model was used as a species-tree prior. To enable
220 ambiguous site processing of the IUPAC dataset, we manually added *useAmbiguities*
221 *="true"* to the gene tree likelihood priors in the XML file. All analyses were run for
222 10,000,000 generations, retaining one sample every 10,000 generations, or until
223 convergence of all parameters (ESS values > 200), as assessed with Tracer v1.7.2
224 (Rambaut et al., 2018).
225
226 ***Network analyses***

227 Hybridization is thought to be common in *Cyanococcus* (Camp, 1945; Vander
228 Kloet, 1988). To investigate potential reticulation between diploid taxa, we used a
229 pseudolikelihood approach as implemented in SNaQ (Solís-Lemus and Ané, 2016). For
230 each dataset (IUPAC and alleles) we tested models in which we allowed a maximum of

231 zero to three hybridization events (*hmax* = 0–3) and used the log pseudolikelihood profile
232 of these runs to estimate the best fitting model. Gene trees inferred from IQ-TREE were
233 used as input. Twenty independent runs were used for each *hmax* value. The
234 computational constraints of this method precluded the estimation of a network with
235 every sample represented as a tip in the tree. Instead, alleles from individual allele trees
236 were assigned to species, resulting in a network in which tips represented species. The
237 IUPAC dataset was subsampled such that each species was represented by one to three
238 samples. To more precisely estimate the placement of the hybrid event suggested by these
239 analyses (i.e., was a single *V. pallidum* population involved or did the hybrid event pre-
240 date all sampled *V. pallidum* populations), we constructed an additional IUPAC dataset
241 including all eight sampled individuals of *V. pallidum*.
242
243 ***Concordance-discordance analyses***
244 Because high bootstrap support can be recovered from phylogenetic analyses
245 despite a low number of genes supporting the topology (e.g., Minh et al., 2020), we
246 additionally assessed conflict within our dataset using gene concordance factors (gCF;
247 percentage of genes supporting a given clade) and site concordance factors (sCF;
248 percentage of informative sites) as implemented in IQ-TREE. Individual IUPAC gene
249 trees were used to calculate both gCF and sCF with 1000 random quartets in the sCF
250 analysis (–scf 1000) for each of the topologies inferred from concatenated and species
251 tree analyses (see above).
252 Discordance was additionally assessed with PhyParts v0.0.1 (Smith et al., 2015).
253 The best individual IUPAC gene trees inferred from IQ-TREE were rooted and outgroup

taxa were removed with Phyx. Results from these analyses were visualized with the *PhyPartsPieCharts* script. As in the gCF/sCF analyses, we tested each of the topologies inferred from concatenated and species tree analyses.

RESULTS

Flow cytometry

Flow cytometry analysis of silica-dried leaf material provided clear genome size estimation for 33 of 36 *Cyanococcus* samples (Appendix S1). Average 2C values ranged from 1.08–1.65 pg, within the range for diploid *Vaccinium* individuals previously determined by Hummer et al. (2015) and Redpath et al. (2022). Although we are in the process of reassessing the morphological characters traditionally used to define species in *Cyanococcus*, ploidy estimates mostly conformed to expectations based on morphological identification and observations of the size and density of stomata on second-year branchlets (Fritsch et al., in press). The one conspicuous exception is *V. boreale*, which was nearly indistinguishable on the basis of morphology from its tetraploid counterpart, *V. angustifolium*, although more detailed analysis of stomatal size and density may facilitate identification (Aalders and Hall, 1962).

Sequence data

Of the 353 loci targeted with the Angiosperms353 probe set, we successfully captured and sequenced 348. Of these, 25 were flagged as potentially containing paralogs. After removing these loci and all columns containing more than 50% missing data, the final concatenated IUPAC alignment consisted of 323 loci of alignment length

672,737 bp (= characters); 22,421 of the characters were parsimony-informative. Individual supercontig gene (and allele) alignments ranged in length from 272 bp to 7064 bp.

Maximum likelihood analyses

Concatenated maximum-likelihood (ML) analyses of the IUPAC dataset with IQ-TREE resulted in an overall well-supported topology and maximally supported *Cyanococcus* clade (Fig. 2A). A northern lineage of *V. boreale* and *V. myrtilloides* was placed as sister to a large clade composed of the remaining taxa with distributions extending into the southeastern United States. Within this clade, we found three sister-species relationships: *V. elliottii*-*V. pallidum*, *V. darrowii*-*V. tenellum*, and *V. fuscatum*-*V. caesariense*. This diploid analysis distinguished six maximally supported terminal groups. One sample of *V. boreale* was found to be nested within *V. myrtilloides* and our only sample of *V. caesariense* nested within *V. fuscatum*.

Species tree analyses

The SVDquartets analysis (IUPAC dataset) recovered *V. elliottii* as non-monophyletic, with one sample sister to the *V. fuscatum*-*V. caesariense* clade and the other two in a much deeper position in the tree, albeit with low support (Fig. 2B). The remaining relationships were consistent with the results from IQ-TREE and ASTRAL-III, including the non-monophyly of *V. boreale* and the nested position of *V. caesariense* within the *V. fuscatum* clade (Fig. 2). ASTRAL-III analyses recovered a topology (Fig. 2C and 2D) largely consistent with the concatenated ML results. However, the placement

of *V*. *elliottii* differed between IUPAC (Fig. 2C) and allele analyses (Fig. 2D). This taxon was recovered as sister to *V*. *pallidum* with the IUPAC dataset, whereas it was recovered as sister to other diploid highbush taxa, *V*. *fuscatum* and *V*. *caesariense*, with the allele dataset, again with low support. This conflicting placement was observed regardless of whether alleles were assigned to individuals (Fig. 2) or species (Fig. 3). Species tree analyses with STACEY placed *V*. *elliottii* sister to the *V*. *fuscatum*-*V*. *caesariense* clade and *V*. *pallidum* as a stand-alone lineage. This topology was recovered with both the IUPAC and allele datasets and is consistent with the topology inferred in our ASTRAL analysis of allele data. A unique topology in which *V*. *pallidum* is sister to the *V*. *boreale*-*V*. *myrtilloides* clade was observed when scrutinizing the posterior distribution of trees (Fig. 4). This signal, however, is only present in the lowest 5% of the posterior distribution from the IUPAC analysis.

Network analyses

Network analyses of both the IUPAC and allele data with SNaQ suggested a single hybridization event in our sampling of diploid taxa (Fig. 4; Appendix S2). Analysis of the allele data in which alleles were assigned to species recover *V*. *pallidum* as a hybrid taxon with parental lineages identified as *V*. *elliottii* and the clade comprising *V*. *boreale* and *V*. *myrtilloides* (Fig. 4A). Our estimates suggest a nearly equal parental contribution from these two lineages (gamma = 0.57 from *V*. *elliottii* and gamma = 0.43 from *V*. boreale-*V*. *myrtilloides*). Subsequent analysis of the IUPAC data (in which sequences were assigned to samples rather than species) including eight *V*. *pallidum* individuals confirmed that the hybrid event predates the divergence of all sampled *V*.

323 *pallidum* populations and a nearly equal genomic contribution from *V*. *elliottii* (gamma =
324 0.56) and an ancestor of *V*. *boreale*-*V*. *myrtilloides* (gamma = 0.44; Fig. 4C).
325
326 ***Concordance-discordance analyses***
327 High levels of discordance were found within the IUPAC dataset. Despite high
328 bootstrap and posterior probability values, we found relatively low gene (gCF) and site
329 (sCF) concordance factors for the major clades recovered in concatenated and species
330 tree analyses (Fig. 2). Regarding the inconsistent placement of *V*. *elliottii*, 1.9% of genes
331 (41% of sites) place it sister to *V*. *pallidum* whereas 0.6% of loci (36% of sites) support *V*.
332 *elliottii* as sister to *V*. *fuscatum*. These results are consistent with those obtained with
333 PhyParts (Appendix S3).
334
335 **DISCUSSION**
336 Despite the reputation of *Cyanococcus* as taxonomically intractable, results from
337 this study in addition to recent field experience has led us to agree with Ward (1974) that
338 *Cyanococcus* "*...is difficult but not in any way an irresolvable tangle of intergrading*
339 *populations*" (p. 192). Although high levels of gene-tree discordance and topological
340 differences between concatenated ML and species tree methods were observed, the
341 overall topology, monophyly of major clades corresponding to various morphospecies
342 concepts, and placement of these clades were consistent across analyses and datasets. All
343 analyses resolve a northern lineage of *V*. *boreale* and *V*. *myrtilloides* sister to the
344 remaining primarily southeastern taxa. Moreover, the analyses consistently recover a
345 close association between *V*. *darrowii* and *V*. *tenellum* and between *V*. *fuscatum* and *V*.

346 *caesariense*. These results are consistent with an early allozyme study of diploid
347 *Cyanococcus* populations based on phenetic analysis (Bruederle and Vorsa, 1994).
348 Observed areas of discordance are primarily from inconsistencies in the
349 placement of *V. pallidum* and *V. elliottii*, suggesting hybridization involving these taxa.
350 Network estimation specifically implicated *V. pallidum* as a hybrid taxon. Further
351 analyses including numerous *V. pallidum* individuals sampled across a wide geographic
352 range yielded results showing that the hybrid event predates the divergence of all
353 sampled populations, suggesting that *V. pallidum* is a species of homoploid-hybrid origin.
354 Parental taxa are suggested to be *V. elliottii* and the lineage giving rise to *V. boreale* and
355 *V. myrtilloides*. A recent study of expressed sequence tag-polymerase chain reaction
356 markers (Rowland et al., 2021) inferred *V. pallidum* as a close relative of *V. boreale* and
357 *V. myrtilloides*, consistent with this supposition. Although several of our analyses
358 inferred a sister relationship of *V. pallidum* with *V. elliottii*, none found *V. pallidum* to be
359 sister to the *V. boreale-myrtilloides* clade. This signal does, however, appear to be present
360 in our dataset when examining the posterior distribution of trees from a Bayesian analysis
361 in STACEY. *Vaccinium pallidum* occupies a geographic range largely overlapping those
362 of its two putative parents (which do not overlap in range), extending further north than
363 *V. elliottii* and further south than either *V. boreale* or *V. myrtilloides* (Fig. 1).
364 Morphologically, there are not immediately clear characters consistent with the hybrid
365 origin of *V. pallidum*, though this would be expected if the hybrid event was ancient and
366 *V. pallidum* has had sufficient evolutionary time to accumulate morphological attributes
367 distinct from either parent. Moreover, the lack of intermediate morphological characters
368 does not preclude *V. pallidum* as a potential hybrid taxon as hybridization is not

369 necessarily expected to leave a consistent or predictable phenotypic signature (Anderson,
370 1948; Rieseberg et al., 1993).
371 Monophyly for all morphospecies was recovered, with two notable exceptions: *V*.
372 *boreale* and *V*. *fuscatum*. One sample of *V*. *boreale* consistently nested within *V*.
373 *myrtilloides* and our *V*. *casearíense* sample nested within *V*. *fuscatum* (see also Bruederle
374 and Vorsa, 1994). In the case of *V*. *boreale*, no evidence of gene flow was detected in our
375 dataset, although hybrids of *V*. *boreale* and *V*. *myrtilloides* have been reported (Hall and
376 Aalders, 1962). Gene flow was detected between *V*. *caesariense* and *V*. *fuscatum* in a
377 sub-optimal SNaQ network (not shown), potentially explaining the non-monophyly of *V*.
378 *fuscatum*. Alternatively, the longstanding decision to recognize *V*. *caesariense*
379 (essentially a glabrous version of *V*. *fuscatum* occurring on the coastal plain) as an
380 independent entity may be erroneous and the morphological attributes (i.e., the lack of
381 pubescence on stems and/or leaves) used to distinguish it from *V*. *fuscatum* may merely
382 be variation within a species. Regarding the *V*. *corymbosum* "highbush" concept, this
383 result and the apparent sister relationship of *V*. *elliottii* would appear to at least partially
384 corroborate Vander Kloet's decision to combine these taxa into a single species. The
385 morphologically distinct and phylogenetically cohesive *V*. *elliottii*, however, challenges
386 this broad concept. Unfortunately, without the inclusion of polyploid taxa we cannot yet
387 satisfactorily address this issue. Furthermore, we have sampled only two populations of
388 *V*. *boreale* and one population of *V*. *caesariense* in this study; meaningful conclusions
389 regarding these taxa must await further sampling and more in-depth analyses.
390 Although our study of the morphological characters defining species in
391 *Cyanococcus* is ongoing, our working morphospecies concepts for diploid *Cyanococcus*

taxa appear to be largely verified with molecular data, as is our hypothesis that the true species composition of this clade likely falls somewhere between the highly divided concept of Camp (1945) and the highly combined concept of Vander Kloet (1988).

Alleles -vs- IUPAC data

Recent studies have attempted to address questions as to the necessity of phasing alleles in phylogenetic reconstruction (e.g., Kamneva et al., 2017; Andermann et al., 2018; Kates et al., 2018; Tiley et al., 2021). We found that in the presence of hybridization, IUPAC and allele data resulted in different topologies. Analyses of IUPAC data consistently inferred a close phylogenetic association between *V*. *pallidum* and *V*. *elliottii*, often as sister lineages. Conversely, allele data inferred *V*. *pallidum* as a lone-lineage, phylogenetically intermediate between its two putative parental lineages. This pattern of phylogenetic intermediacy of hybrids relative to their parents has been previously observed across a wide range of time scales and data types, including morphological data from F_1s produced through controlled crosses (McDade, 1990), RADseq data from putative naturally formed F_1 hybrids (Hauser et al., 2017), and target-enrichment data from taxa involved in ancient introgression events (Crowl et al., 2020). Allele data resolved *V*. *elliottii* as sister to other "highbush" taxa, i.e., *V*. *fuscatum* and *V*. *caesariense*, consistent with our network analyses. This pattern is recovered regardless of whether alleles were assigned to individuals or species. These results suggest that phasing alleles is useful in datasets containing hybrid taxa.

On homoploid hybrids

Homoploid hybrid speciation is the process by which a new species is formed through hybridization of divergent parent lineages, but without an increase in ploidy (Grant, 1981; Rieseberg, 1997). Although several potential homoploid hybrid species are known in various plant groups, e.g., *Carex* (Hodel et al., 2022), *Senecio* (James and Abbott, 2005; Brennan et al., 2012), *Iris* (Arnold, 1993; Taylor et al., 2013; Zalmat et al., 2021), *Pinus* (Wang and Szmidt, 1994), *Penstemon* (Wolfe et al., 1998), *Paeonia* (Pan et al., 2007), they appear to be somewhat rare in nature (but see Nieto Feliner et al., 2017). Results from the present study suggest that *V. pallidum* is an additional example. Whereas hybridization is well known in *Vaccinium*, to our knowledge this is the first report of a naturally formed homoploid hybrid species in the group.

To further test this supposition, we additionally considered an F_1 homoploid (diploid) hybrid resulting from a controlled cross between *V. myrtilloides* x *elliottii*. When included in our dataset, network analyses correctly inferred the parents of this hybrid plant and an equal genomic contribution from each parent (Appendix S2). Although far from conclusive, this test case serves as a positive control of sorts and provides increased confidence that our genomic dataset and analytical approach can accurately identify a homoploid hybrid taxon. We caution, however, that much work is needed to verify these findings, including further sampling of putative parental taxa, tests of reproductive isolation, investigation of niche divergence, and a detailed morphological study.

What about polyploids?

Whereas our efforts have focused on the diploid species of *Cyanococcus*, the group contains numerous polyploid lineages. Polyploids, with more than two copies of each chromosome, remain difficult to analyze in a phylogenetic context. The central challenge of analyzing sequence data from polyploids, and especially allopolyploids, lies in identifying divergent homeolog copies from parental taxa. The majority of bioinformatic tools available for processing next-generation sequence data were developed for diploid organisms and therefore collapse variable homeolog sequences into a single consensus sequence for downstream analysis. For polyploids, this creates chimeric sequences that obscure signals of polyploidy and a polyploid mode of origin. Conversely, allelic data more accurately capture the complex genomic histories of polyploids and allow for the incorporation of divergent signals from polyploid loci into phylogenomic inference, thus distinguishing allopolyploidy from autopolyploidy and identifying parental taxa.

The diploid phylogenetic estimate presented here in combination with recent advances in phylogenetic network analysis and a recently developed bioinformatics approach to phasing alleles for arbitrary ploidy from target enrichment data (Tiley et al., 2021) provide an exciting opportunity to investigate polyploid *Cyanococcus* taxa and infer parentage and mode of polyploidization in this challenging group.

Acknowledgements

We thank Aaron Liston for many helpful comments that improved this manuscript. Funding for this research was provided by the Trinity College of Arts and Sciences at Duke University and the National Science Foundation (DEB-2038213 to Duke and North Carolina State University; DEB-2038217 to the Botanical Research Institute of Texas).

Author Contributions

AAC, PWF, HA, PSM designed the study; AAC, PWF, PSM carried out fieldwork; NPL conducted flow cytometry analyses; AAC ran phylogenomic analyses; all authors have contributed to the intellectual content and writing of this manuscript.

Data Availability Statement

Raw reads are deposited in the NCBI Sequence Reads Archive (BioProject: PRJNA854616). Final DNA alignment and gene-tree files are available from the Dryad repository (doi:10.5061/dryad.cc2fqz68x).

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Voucher table.

Appendix S2. Comparison of network analyses with different datasets.

Appendix S3. Results from concordance/discordance analyses.

Literature Cited

AALDERS, L.E., And I.V. HALL. 1962. New evidence on the cytotaxonomy of *Vaccinium* species as revealed by stomatal measurements from herbarium specimens. *Nature* 196: 694–694.

ANDERMANN, T., A.M. FERNANDES, U. OLSSON, M. TÖPEL, B. PFEIL, B. OXELMAN, A. ALEIXO, ET AL. 2018. Allele phasing greatly improves the phylogenetic utility of ultraconserved elements. *Systematic Biology* 68: 32–46.

ANDERSON, E. 1948. Hybridization of the habitat. *Evolution* 2: 1–9.

ANDREWS, S. 2010. FastQC - A quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

ARNOLD, M.L. 1993. *Iris nelsonii* (Iridaceae): Origin and genetic composition of a homoploid hybrid species. *American Journal of Botany* 80: 577–583.

BOUCKAERT, R., J. HELED, D. KÜHNERT, T. VAUGHAN, C.-H. WU, D. XIE, M.A. SUCHARD, ET AL. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10: e1003537.

BRADLEY, R.K., A. ROBERTS, M. SMOOT, S. JUVEKAR, J. DO, C. DEWEY, I. HOLMES, and L. PACHTER. 2009. Fast Statistical Alignment. *PLoS Computational Biology* 5: e1000392.

BRENNAN, A.C., D. BARKER, S.J. HISCOCK, and R.J. ABBOTT. 2012. Molecular genetic and quantitative trait divergence associated with recent homoploid hybrid speciation: a study of *Senecio squalidus* (Asteraceae). *Heredity* 108: 87–95.

BROWN, J.W., J.F. WALKER, and S.A. SMITH. 2017. Phyx: phylogenetic tools for unix. *Bioinformatics* 33: 1886–1888.

BRUEDERLE, L.P., and N. VORSA. 1994. Genetic differentiation of diploid blueberry, *Vaccinium* sect. *Cyanococcus* (Ericaceae). *Systematic Botany* 19: 337–349.

CAMP, W.H. 1942. On the Structure of populations in the genus *Vaccinium*. *Brittonia* 4: 189.

CAMP, W.H. 1945. The North American blueberries with notes on other groups of Vacciniaceae. *Brittonia* 5: 203–275.

CAMP, W.H., and C.L. GILLY. 1943. The structure and origin of species. *Brittonia* 4: 323–385.

CHIFMAN, J., and L. KUBATKO. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30: 3317–3324.

COSTICH, D.E., R. ORTIZ, T.R. MEAGHER, L.P. BRUEDERLE, and N. VORSA. 1993. Determination of ploidy level and nuclear DNA content in blueberry by flow cytometry. *Theoretical and Applied Genetics* 86: 1001–1006.

CROWL, A.A., P.S. MANOS, J.D. MCVAY, A.R. LEMMON, E.M. LEMMON, and A.L. HIPP. 2020. Uncovering the genomic signature of ancient introgression between white oak lineages (*Quercus*). *New Phytologist* 226: 1158–1170.

DARROW, G.M., and W.H. CAMP. 1945. *Vaccinium* hybrids and the development of new horticultural material. *Bulletin of the Torrey Botanical Club* 72: 1.

DEPRISTO, M.A., E. BANKS, R. POPLIN, K. V GARIMELLA, J.R. MAGUIRE, C. HARTL, A.A. PHILIPPAKIS, ET AL. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43: 491–8.

DOLEŽEL, J., J. GREILHUBER, and J. SUDA. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* 2: 2233–2244.

534 DOYLE, J.J., and J.L. DOYLE. 1987. A rapid DNA isolation procedure for small
535 quantities of fresh leaf tissue. *Phytochemical Bulletin*.

536 FRITSCH, P.W., A.A. CROWL, H. ASHRAFI, and P.S. MANOS. In press. Understanding
537 the systematics and evolution of *Vaccinium* sect. *Cyanococcus* (Ericaceae):
538 progress and prospects. *Rhodora*.

539 GRANT, V. 1981. Plant Speciation. Columbia University Press.

540 HALL, I.V., and L.E. AALDERS. 1961. Cytotaxonomy of lowbush blueberries in Eastern
541 Canada. *American Journal of Botany* 48: 199–201.

542 HARLAN, J.R., and J.M.J. DE WET. 1963. The compilospecies concept. *Evolution* 17:
543 497.

544 HAUSER, D.A., A. KEUTER, J.D. MCVAY, A.L. HIPP, and P.S. MANOS. 2017. The
545 evolution and diversification of the red oaks of the California Floristic Province
546 (*Quercus* section *Lobatae*, series *Agrifoliae*). *American Journal of Botany* 104:
547 1581–1595.

548 HODEL, R.G.J., R. MASSATTI, and L.L. KNOWLES. 2022. Hybrid enrichment of
549 adaptive variation revealed by genotype–environment associations in montane
550 sedges. *Molecular Ecology* 31: 3722–3737.

551 HOANG, D.T., O. CHERNOMOR, A. VON HAESELER, B.Q. MINH, and L.S. VINH. 2018.
552 UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology*
553 *and Evolution* 35: 518–522.

554 HUMMER, K.E., N.V. BASSIL, H.P. RODRÍQUEZ ARMENTA, and J.W. OLMSTEAD.
555 2015. *Vaccinium* species ploidy assessment. *Acta Horticulturae* 1101: 199–
556 204.

557 HUXLEY, J. 1942. Evolution. The Modern Synthesis. London: George Alien &
558 Unwin Ltd.

559 JAMES, J.K., and R.J. ABBOTT. 2005. Recent, allopatric, homoploid hybrid
560 speciation: The origin of *Senecio squalidus* (Asteraceae) in the British Isles
561 from a hybrid zone on Mount Etna, Sicily. *Evolution* 59: 2533–2547.

562 JOHNSON, M.G., E.M. GARDNER, Y. LIU, R. MEDINA, B. GOFFINET, A.J. SHAW,
563 N.J.C. ZEREGA, and N.J. WICKETT. 2016. HybPiper: Extracting coding
564 sequence and introns for phylogenetics from high-throughput sequencing reads
565 using target enrichment. *Applications in Plant Sciences* 4: 1600016.

566 JOHNSON, M.G., L. POKORNY, S. DODSWORTH, L.R. BOTIGUÉ, R.S. COWAN, A.
567 DEVAULT, W.L. EISERHARDT, ET AL. 2019. A universal probe set for targeted
568 sequencing of 353 nuclear genes from any flowering plant designed using k-
569 medoids clustering. *Systematic Biology* 68: 594–606.

JONES, G. 2017. Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *Journal of Mathematical Biology* 74: 447–467.

KATES, H.R., M.G. JOHNSON, E.M. GARDNER, N.J.C. ZEREGA, and N.J. WICKETT. 2018. Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of *Artocarpus*. *American Journal of Botany* 105: 404–416.

KRON, K.A., E.A. POWELL, and J.L. LUTEYN. 2002. Phylogenetic relationships within the blueberry tribe (Vaccinieae, Ericaceae) based on sequence data from *matK* and nuclear ribosomal ITS regions, with comments on the placement of *Satyria*. *American Journal of Botany* 89: 327–336.

LANFEAR, R., B. CALCOTT, S.Y.W. HO, and S. GUINDON. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29: 1695–1701.

LANFEAR, R., B. CALCOTT, D. KAINER, C. MAYER, and A. STAMATAKIS. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology* 14: 82.

LI, H., and R. DURBIN. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–60.

LYRENE, P.M., N. VORSA, and J.R. BALLINGTON. 2003. Polyploidy and sexual polyploidization in the genus *Vaccinium*. *Euphytica* 133: 27–36.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17: 10.

MCDADE, L. 1990. Hybrids and phylogenetic systematics I. Patterns of character expression in hybrids and their implications for cladistic analysis. *Evolution* 44: 1685–1700.

MCKENNA, A., M. HANNA, E. BANKS, A. SIVACHENKO, K. CIBULSKIS, A. KERNYTSKY, K. GARIMELLA, ET AL. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297–1303.

MCLAY, T.G.B., J.L. BIRCH, B.F. GUNN, W. NING, J.A. TATE, L. NAUHEIMER, E.M. JOYCE, ET AL. 2021. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences* 9: aps3.11420.

MINH, B.Q., M.W. HAHN, and R. LANFEAR. 2020. New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution* 37: 2727–2733.

606 NGUYEN, L.-T., H.A. SCHMIDT, A. VON HAESELER, and B.Q. MINH. 2015. IQ-TREE:
607 A fast and effective stochastic algorithm for estimating maximum-likelihood
608 phylogenies. *Molecular Biology and Evolution* 32: 268–274.

609 NIETO FELINER, G., I. ÁLVAREZ, J. FUERTES-AGUILAR, M. HEUERTZ, I. MARQUES, F.
610 MOHARREK, R. PIÑEIRO, ET AL. 2017. Is homoploid hybrid speciation that rare?
611 An empiricist's view. *Heredity* 118: 513–516.

612 PAN, J., D. ZHANG, and T. SANG. 2007. Molecular phylogenetic evidence for the
613 origin of a diploid hybrid of *Paeonia* (Paeoniaceae). *American Journal of*
614 *Botany* 94: 400–408.

615 POSTER, L.S., S.N. HANDEL, and P.E. SMOUSE. 2017. Corolla size and temporal
616 displacement of flowering times among sympatric diploid and tetraploid
617 highbush blueberry (*Vaccinium corymbosum*). *Botany* 95: 395–404.

618 RAMBAUT, A., A.J. DRUMMOND, D. XIE, G. BAELE, and M.A. SUCHARD. 2018.
619 Posterior summarization in Bayesian phylogenetics using Tracer 1.7.
620 *Systematic Biology* 67: 901–904.

621 REDPATH, L.E., R. ARYAL, N. LYNCH, J.A. SPENCER, A.M. HULSE-KEMP, J.R.
622 BALLINGTON, J. GREEN, ET AL. 2022. Nuclear DNA contents and ploidy levels
623 of North American *Vaccinium* species and interspecific hybrids. *Scientia*
624 *Horticulturae* 297: 110955.

625 RIESEBERG, L.H., N.C. ELLSTRAND, and M. ARNOLD. 1993. What can molecular and
626 morphological markers tell us about plant hybridization? *Critical Reviews in*
627 *Plant Sciences* 12: 213–241.

628 RIESEBERG, L.H. 1997. Hybrid origins of plant species. *Annual Review of Ecology*
629 *and Systematics* 28: 359–389.

630 ROWLAND, L.J., D.J. BELL, N. ALKHAROUF, N. V. BASSIL, F.A. DRUMMOND, L.
631 BEERS, E.J. BUCK, ET AL. 2012. Generating genomic tools for blueberry
632 improvement. *International Journal of Fruit Science* 12: 276–287.

633 SMITH, S.A., M.J. MOORE, J.W. BROWN, and Y. YANG. 2015. Analysis of
634 phylogenomic datasets reveals conflict, concordance, and gene duplications
635 with examples from animals and plants. *BMC Evolutionary Biology* 15: 150.

636 SOLÍS-LEMUS, C., and C. ANÉ. 2016. Inferring phylogenetic networks with
637 maximum pseudolikelihood under incomplete lineage sorting. *PLOS Genetics*
638 12: e1005896.

639 SWOFFORD, D.L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other
640 methods). Version 4. *Sinauer Associates, Sunderland, Massachusetts*.

641 TAYLOR, S.J., L.D. ROJAS, S.W. HO, and N.H. MARTIN. 2013. Genomic collinearity
642 and the genetic architecture of floral differences between the homoploid hybrid

643 species Iris nelsonii and one of its progenitors, Iris hexagona. *Heredity* 110:
644 63–70.

645 TILEY, G.P., A.A. CROWL, P.S. MANOS, E.B. SESSA, C. SOLÍS-LEMUS, A.D. YODER,
646 and J.G. BURLEIGH. 2021. Phasing alleles improves network inference with
647 allopolyploids. *bioRxiv*2021.05.04.442457.

648 VANDER KLOET, S.P. 1980. The taxonomy of the highbush blueberry, *Vaccinium*
649 *corymbosum*. *Canadian Journal of Botany* 58: 1187–1201.

650 VANDER KLOET, S.P. 1983. The taxonomy of *Vaccinium* § Cyanococcus : a
651 summation. *Canadian Journal of Botany* 61: 256–266.

652 VANDER KLOET, S.P. 1988. The genus *Vaccinium* in North America. Research
653 Branch, Agriculture Canada, Ottawa, Canada.

654 VANDER KLOET, S.P. 2009. Vaccinium. *In* Flora of North America Editorial
655 Committee [ed.], Flora of North America North of Mexico, vol. 8.
656 Magnoliophyta: Paeoniaceae to Ericaceae, 515–530. Oxford University Press,
657 New York, NY.

658 WANG, X.-R., and A.E. SZMIDT. 1994. Hybridization and chloroplast DNA variation
659 in a *Pinus* species complex from Asia. *Evolution* 48: 1020–1031.

660 WARD, D.B. 1974. Contributions to the Flora of Florida: 6, *Vaccinium* (Ericaceae).
661 *Castanea* 39: 191–205.

662 WEAKLEY, A. 2020. Flora of the Southern and Mid–Atlantic States. University of
663 North Carolina Herbarium, North Carolina Botanical Garden, Chapel Hill, NC.

664 WOLFE, A.D., Q.-Y. XIANG, and S.R. KEPHART. 1998. Diploid hybrid speciation in
665 *Penstemon* (Scrophulariaceae). *Proceedings of the National Academy of*
666 *Sciences* 95: 5112–5115.

667 XIE, M., Q. WU, J. WANG, and T. JIANG. 2016. H-PoP and H-PoPG: heuristic
668 partitioning algorithms for single individual haplotyping of polyploids.
669 *Bioinformatics* 32: 3735–3744.

670 ZALMAT, A.S., V.A. SOTOLA, C.C. NICE, and N.H. MARTIN. 2021. Genetic structure
671 in Louisiana Iris species reveals patterns of recent and historical admixture.
672 *American Journal of Botany* 108: 2257–2268.

673 ZHANG, C., M. RABIEE, E. SAYYARI, and S. MIRARAB. 2018. ASTRAL-III:
674 polynomial time species tree reconstruction from partially resolved gene trees.
675 *BMC Bioinformatics* 19: 153.

676

677

678

679 **Figure Legends**
680
681 **Figure 1**. Geographic distribution maps for diploid *Cyanococcus* morphospecies. Black
682 symbols indicate populations included in our broad survey of ploidy and morphology.
683 Yellow symbols indicate a subset of those samples sequenced and included in
684 phylogenomic analyses.
685
686 **Figure 2**. Comparison of topologies recovered from concatenated and species-tree
687 analyses for the diploid *Cyanococcus* clade (highlighted in blue). Note the inconsistent
688 placement of *V. pallidum* and *V. elliottii* populations between analyses and datasets.
689 Sample numbers refer to the voucher table in Appendix S1. Values above branches
690 indicate support (bootstrap or posterior probability). Values below branches indicate gene
691 concordance factors (gCF) and site concordance factors (sCF). These are reported as
692 gCF/sCF. Intraspecific (population-level) support values are not shown. (A) Phylogenetic
693 estimate from IQ-TREE analysis of the concatenated IUPAC dataset. (B) Species tree
694 inferred from SVDquartets analysis of the concatenated IUPAC dataset. (C) Species tree
695 inferred from ASTRAL-III analysis of the IUPAC dataset. (D) Species tree inferred from
696 ASTRAL-III analysis of the allele dataset.
697
698 **Figure 3**. Comparison of species trees inferred with IUPAC and allele data. In both
699 instances, alleles and IUPAC sequences were assigned to species. Note the inconsistent
700 placement of *V. pallidum* and *V. elliottii* between datasets. (A) Species tree inferred from
701 ASTRAL-III analysis of the IUPAC dataset. (B) Species tree inferred from ASTRAL-III
702 analysis of the allele dataset. Values on branches indicate local posterior probability
703 support.
704
705 **Figure 4**. Evidence for the homoploid hybrid origin of *Vaccinium pallidum*. (A) Network
706 inferred from the allele dataset in which alleles were assigned to species. Values on
707 hybrid edges are the estimated genomic contributions from each parent (gamma). (B)
708 Posterior distribution of Bayesian species-tree analysis. The lowest 5% of trees from the
709 posterior distribution are depicted in yellow, showing alternative placement of *V.*
710 *pallidum* sister to *V. myrtilloides* and *V. boreale*. (C) Network inferred from IUPAC
711 dataset with increased population sampling. Note that the hybrid event predates
712 divergence of all sampled *V. pallidum* populations.

Figure 1

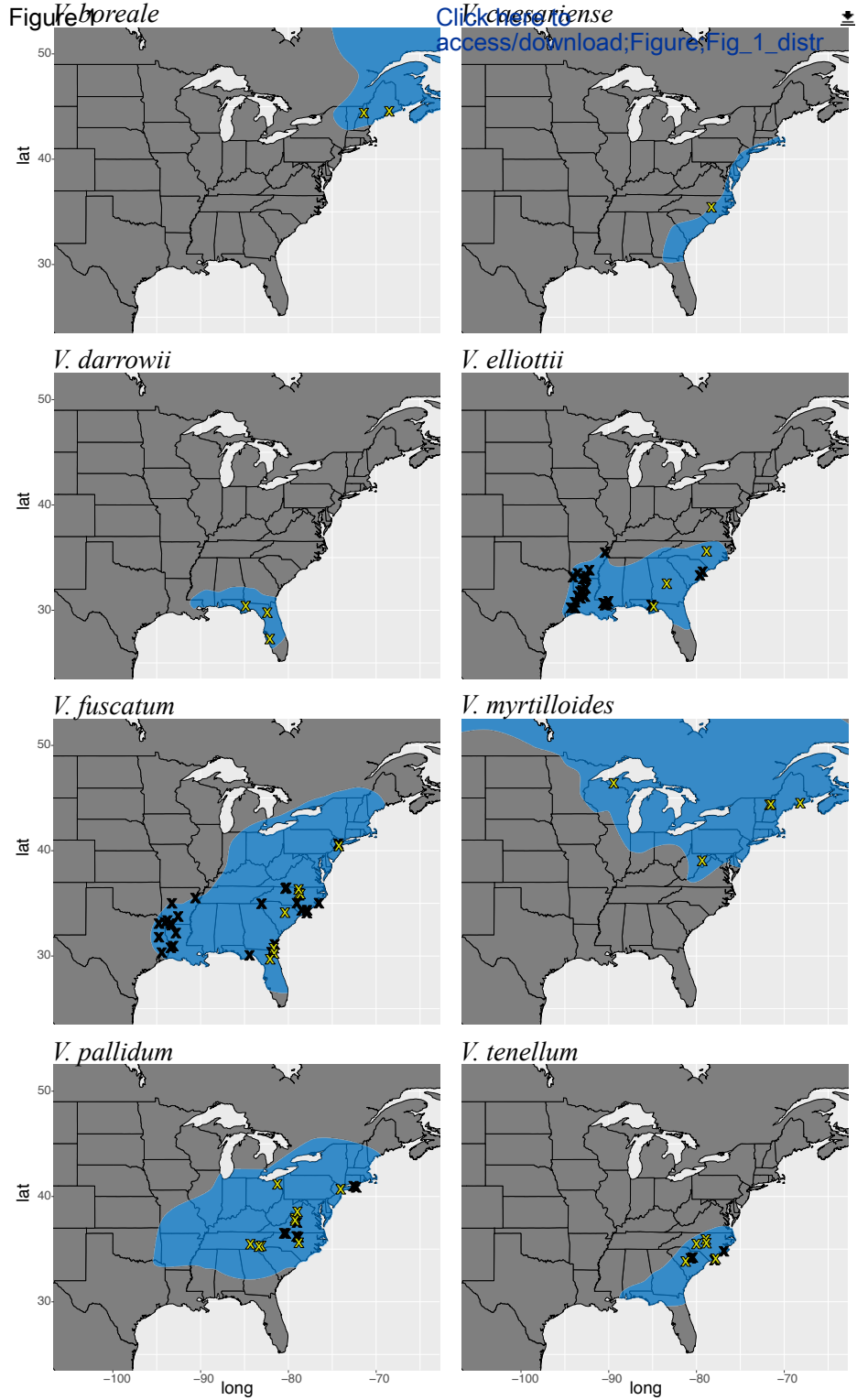


Figure 2

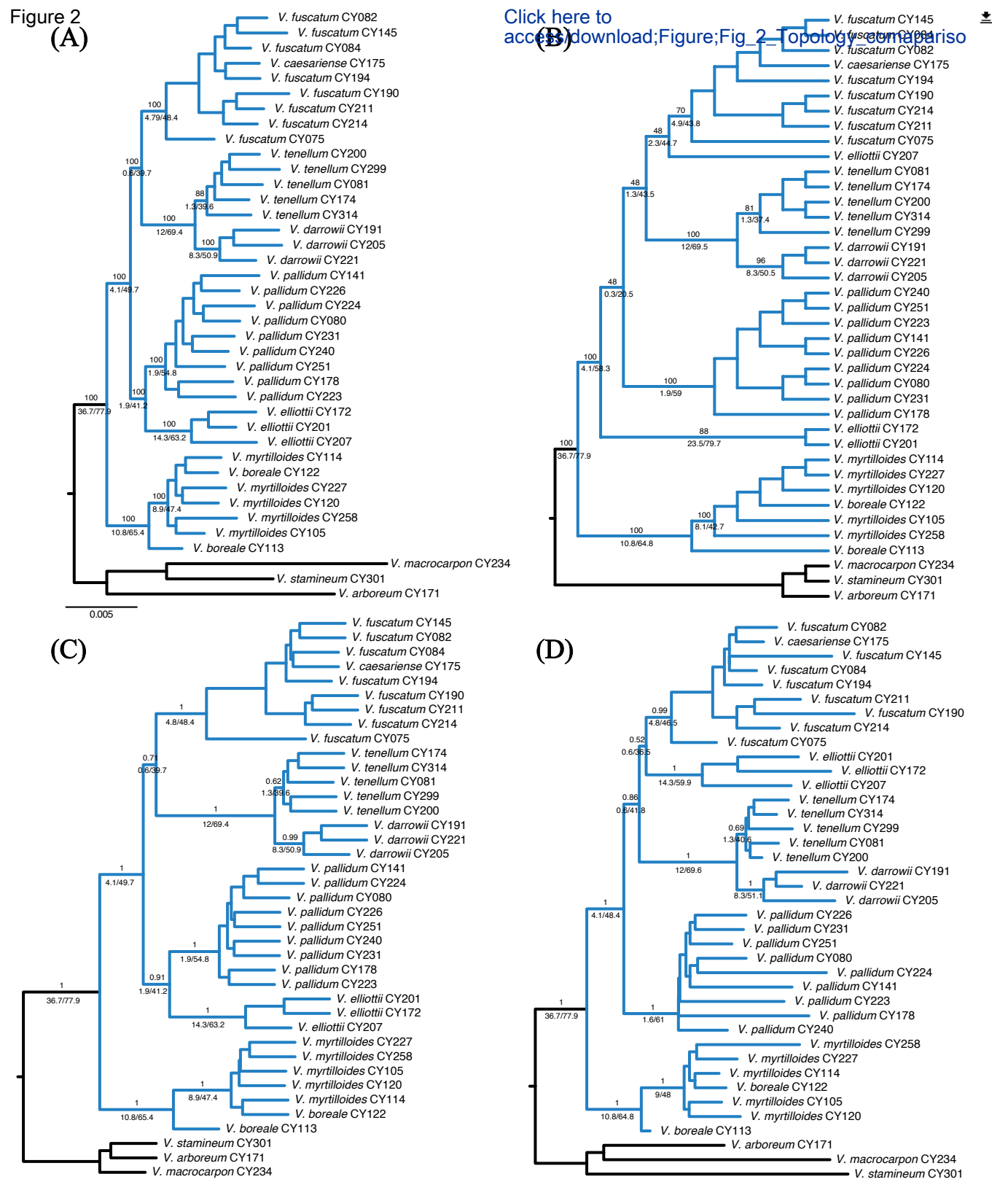
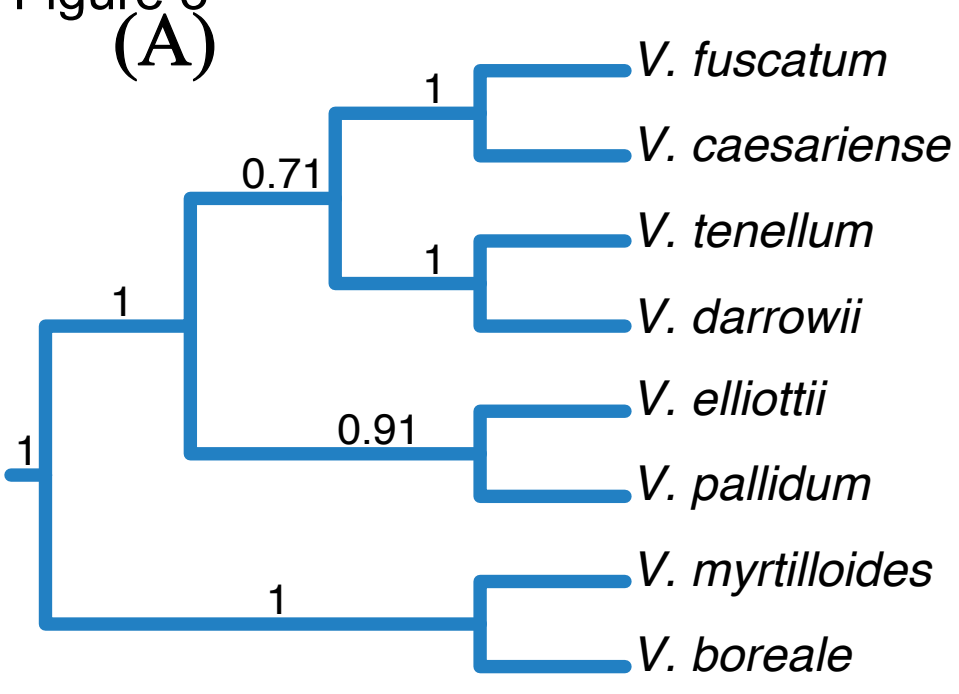


Figure 3



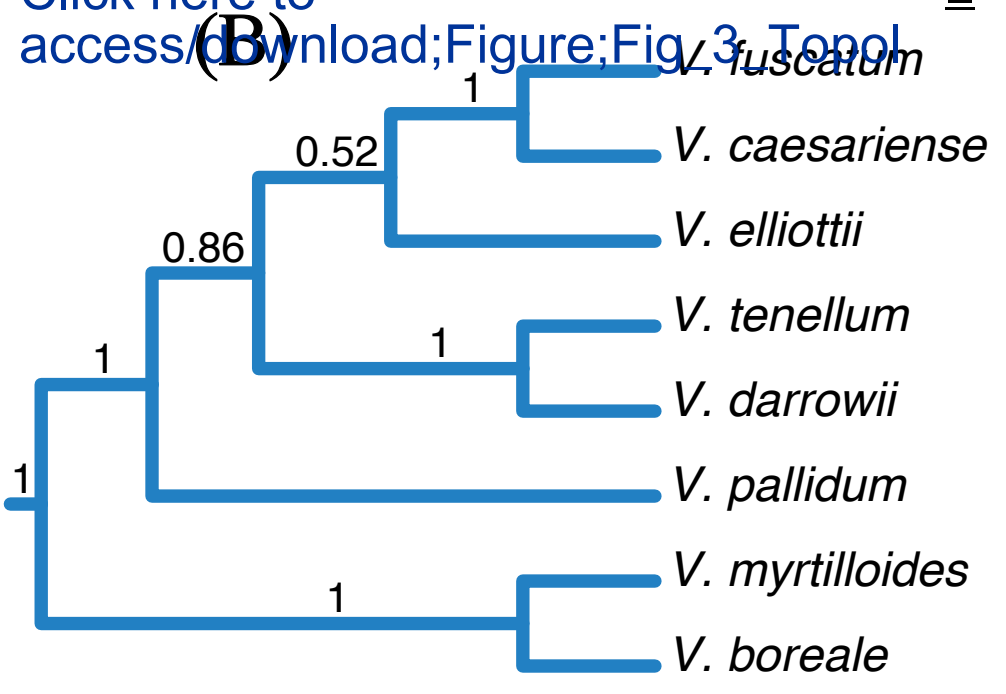
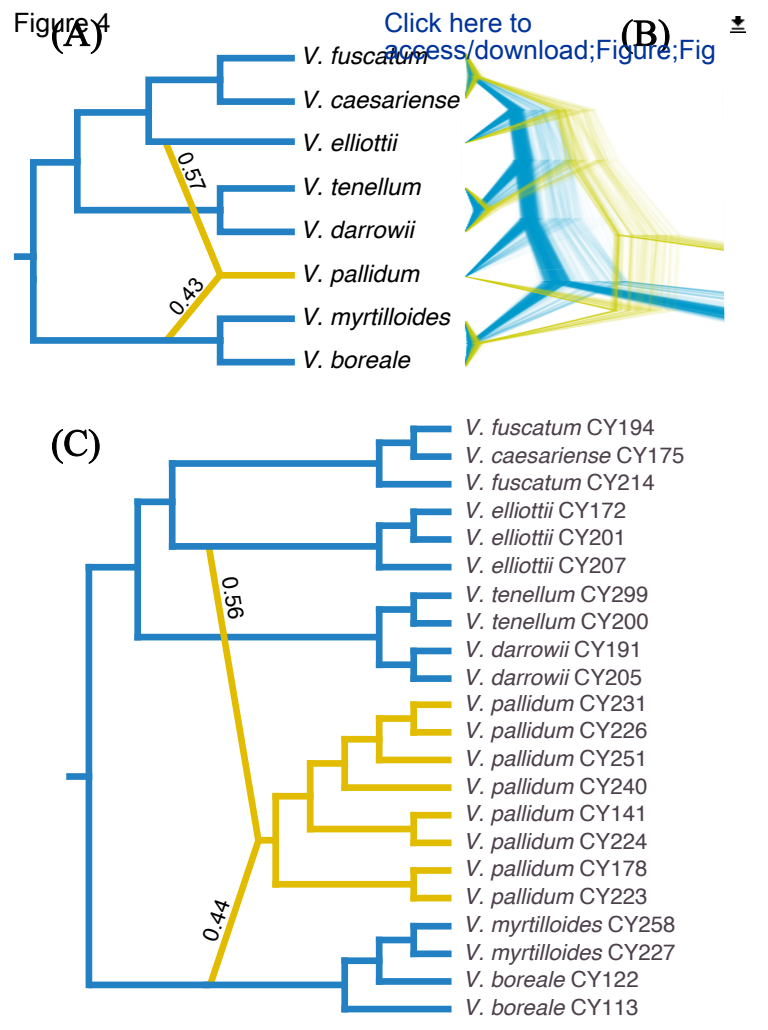


Figure 4



Appendix S1. Voucher table.

Number	Determination	Author	2C genome size (pg)*	Ploidy
PM-CY-075	Vaccinium fuscatum	Aiton	1.45	*2x*
PM-CY-080	Vaccinium pallidum	Aiton	1.38	*2x*
PM-CY-081	Vaccinium tenellum	Aiton	1.35	*2x*
PM-CY-082	Vaccinium fuscatum	Aiton	1.35	*2x*
PM-CY-084	Vaccinium fuscatum	Aiton	1.38	*2x*
PM-CY-105	Vaccinium myrtilloides	Michaux	1.43	*2x*
PM-CY-113	Vaccinium boreale	Hall & Aalders	1.08	*2x*
PM-CY-114	Vaccinium myrtilloides	Michaux	1.34	*2x*
PM-CY-120	Vaccinium myrtilloides	Michaux	1.45	*2x*
PM-CY-122	Vaccinium boreale	Hall & Aalders	1.35	*2x*
PM-CY-141	Vaccinium pallidum	Aiton	1.36	*2x*
PM-CY-145	Vaccinium fuscatum	Aiton	1.43	*2x*
PM-CY-171	Vaccinium arboreum	Marshall	-	-
PM-CY-172	Vaccinium elliottii	Chapman	1.21	*2x*
PM-CY-174	Vaccinium tenellum	Aiton	1.21	*2x*
PM-CY-175	Vaccinium caesariense	Mackenzie	1.30	*2x*
PM-CY-178	Vaccinium pallidum	Aiton	1.25	*2x*
PM-CY-190	Vaccinium fuscatum	Aiton	1.45	*2x*
PM-CY-191	Vaccinium darrowii	Camp	1.37	*2x*
PM-CY-194	Vaccinium fuscatum	Aiton	1.41	*2x*
PM-CY-200	Vaccinium tenellum	Aiton	1.39	*2x*
PM-CY-201	Vaccinium elliottii	Chapman	1.40	*2x*
PM-CY-205	Vaccinium darrowii	Camp	1.31	*2x*
PM-CY-207	Vaccinium elliottii	Chapman	1.32	*2x*
PM-CY-211	Vaccinium fuscatum	Aiton	1.37	2x
PM-CY-214	Vaccinium fuscatum	Aiton	1.44	2x
PM-CY-221	Vaccinium darrowii	Camp	1.39	*2x*
PM-CY-223	Vaccinium pallidum	Aiton	1.60	*2x*
PM-CY-224	Vaccinium pallidum	Aiton	1.65	*2x*
PM-CY-226	Vaccinium pallidum	Aiton	1.37	*2x*
PM-CY-227	Vaccinium myrtilloides	Michaux	-	-
PM-CY-231	Vaccinium pallidum	Aiton	-	-
PM-CY-234	Vaccinium macrocarpon	Aiton	1.38	*2x*
PM-CY-240	Vaccinium pallidum	Aiton	-	-
PM-CY-251	Vaccinium pallidum	Aiton	1.33	*2x*
PM-CY-258	Vaccinium myrtilloides	Michaux	1.29	*2x*

PM-CY-299	Vaccinium tenellum	Aiton	1.41	*2x*
PM-CY-301	Vaccinium stamineum	L.	-	-
PM-CY-314	Vaccinium tenellum	Aiton	1.32	*2x*

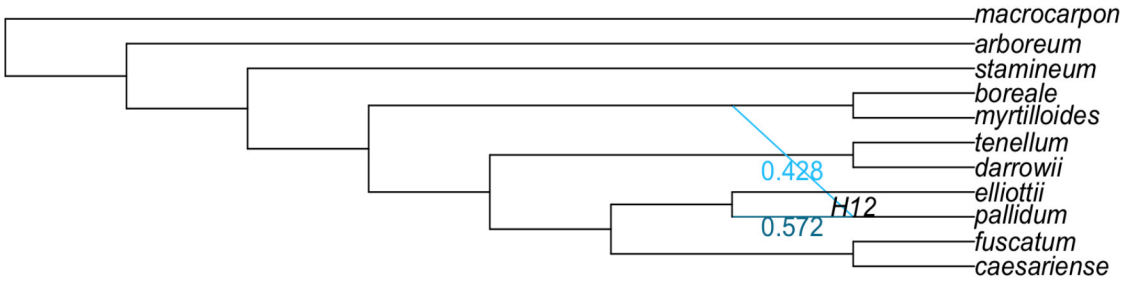
***The 2C genome size values reported here are averages of two independent runs**

Section	Location	Latitude
Cyanococcus	NC; Pilot Mountain, seep streamside, Grindstone trail, low elevatio	36.347191
Cyanococcus	TN; Cherohala Skyway (Rt. 165); 0.5km E of Hemlock Rd turnoff; roa	35.362685
Cyanococcus	NC; Duke Forest off of Gate 10 entrance.	36.022586
Cyanococcus	NC; Duke Forest off of Gate 10 entrance.	36.022586
Cyanococcus	NC; Duke Forest off of Gate 10 entrance. Hairless	36.022586
Cyanococcus	NH; White Mountains; below Silver Cascade Falls	44.206797
Cyanococcus	ME; Mt Desert Island; Cox Protectorate	44.402011
Cyanococcus	ME; Mt Desert Island; Cox Protectorate	44.402011
Cyanococcus	NH; White Mountains; north of Echo Lake along trail to Artists Bluff	44.182038
Cyanococcus	NH; Mount Lafayette, NH, ridge trail	44.158272
Cyanococcus	NJ; Cheesequake State Park; trail to Hooks Creek Lake, yellow trail	40.437405
Cyanococcus	NJ; Cheesequake State Park; trail to Hooks Creek Lake, yellow trail	40.437405
Batodendron	NC; Raven Rock State Park; Raven Rock loop trail	35.466053
Cyanococcus	NC; Raven Rock State Park; Raven Rock loop trail	35.466053
Cyanococcus	NC; Raven Rock State Park; Raven Rock loop trail	35.466053
Cyanococcus	NC; Raven Rock State Park; Raven Rock loop trail	35.466053
Cyanococcus	NC; Raven Rock State Park; Raven Rock loop trail	35.466053
Cyanococcus	FL; along Gainesville-Hawthorn trail	29.591233
Cyanococcus	FL; Gainesville; woods next to Walt Judd's house	29.571185
Cyanococcus	SC; Dr Humphries Rd just before junction with Rt. 34	34.234087
Cyanococcus	SC; Peachtree Rock Preserve, common along trail to the rock	33.830945
Cyanococcus	GA; Cochran, Red Dog Farm Rd (dirt road) near junction with Magn	32.449065
Cyanococcus	FL; Apalachicola NF, along Hwy 65, across from NF Rd 105 pullout.	30.28174
Cyanococcus	FL; Telogia, along Hwy 65; 100m North of Telogia Baptist Church	30.354447
Cyanococcus	FL; Racetrack Rd near intersection with FL-9B	30.105055
Cyanococcus	FL; Yulee; Mentoria Rd near junction with Rt. 200	30.617185
Cyanococcus	FL; Port Charlotte; Tippecanoe Environmental Park	26.994556
Cyanococcus	VA; along Blue Ridge Parkway	37.927431
Cyanococcus	VA; Blue Ridge Parkway, Ravens Roost Overlook	37.933354
Cyanococcus	VA; Riven Rock Park, Harrisonburg; along Rawley Pike Rd.	38.517555
Cyanococcus	WV; Canaan Valley; Freeland boardwalk	39.024692
Cyanococcus	OH; West Branch State Park; along Aliance Rd	41.125812
Oxycoccus	OH; Triangle Lake Bog State Nature Preserve	41.118853
Cyanococcus	NC; Bull Pen road, North Carolina, Slick Rock	
Cyanococcus	NC; Trail from Shortoff Mt to Cole Gap	35.109372
Cyanococcus	MI; Upper Peninsula; UNDERC Field Station; Tender Bog	46.230041

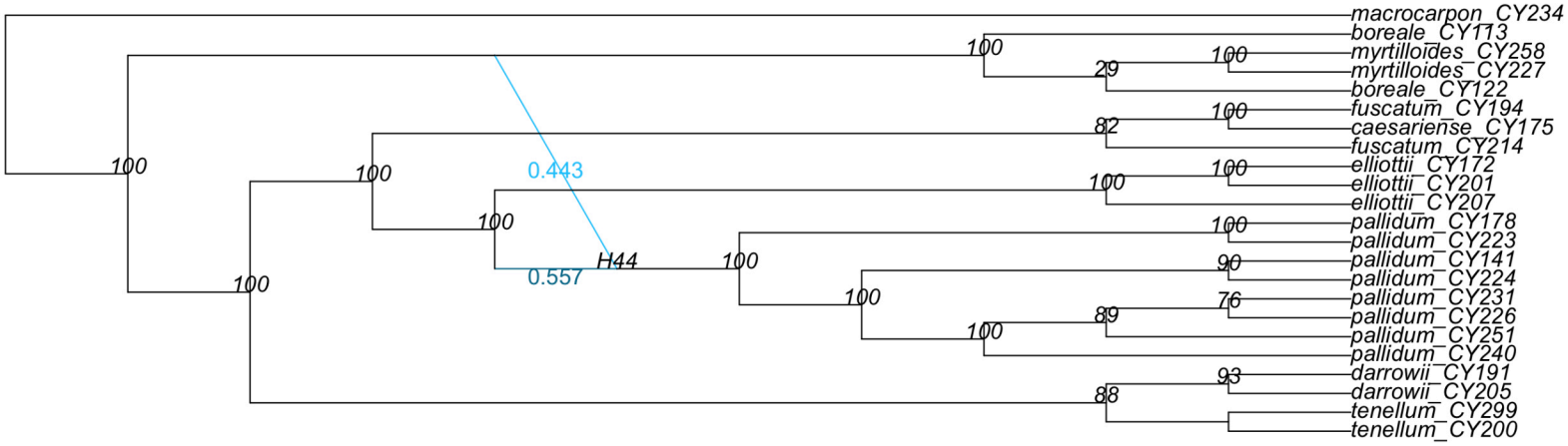
Appendix S2. SNaQ results.

Likelihood values are given for each model tested from three datasets.
The best network selected (indicated with an asterisks) is shown below each table.

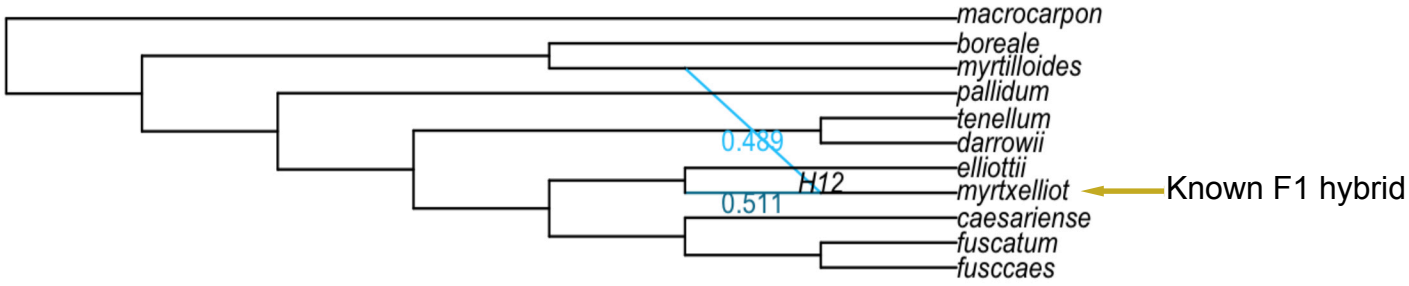
dataset: all diploids (alleles)	-loglik
net0	142.493988
net1*	65.6625044
net2	59.9362875
net3	59.9362875



dataset: all pallidum pops (IUPAC)	-loglik
net0	5292.03115
net1*	2919.78883
net2	2855.42376
net3	2848.98776



dataset: F1 myrtilloides x elliottii (alleles)	-loglik
net0	230.960307
net1*	103.464253
net2	90.8944061
net3	90.8944061

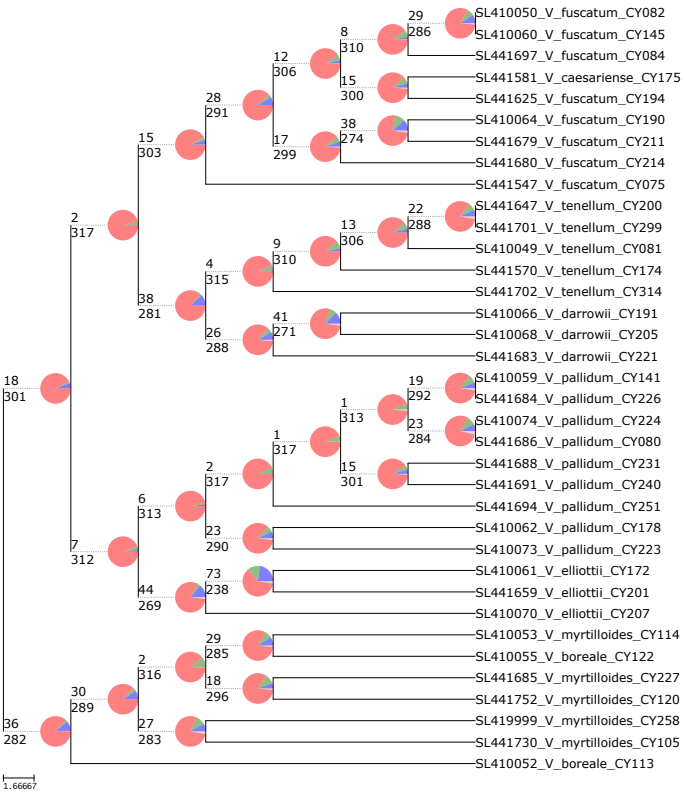


*Network shown for each dataset.

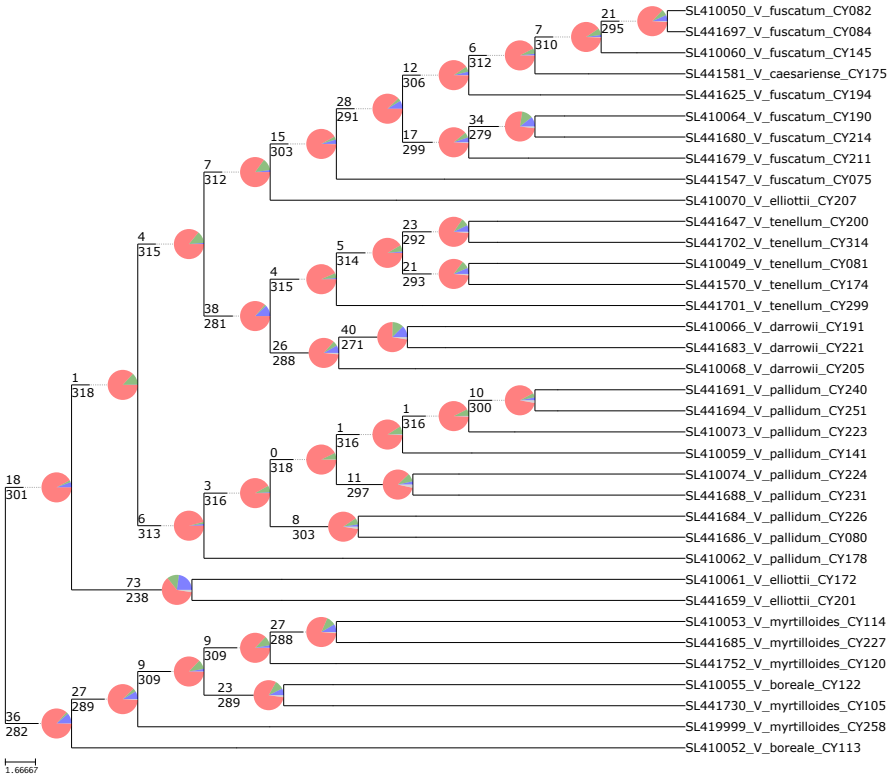
Appendix S3. PhyParts results.

The best individual IUPAC gene trees inferred from IQ-TREE were used as input to visualize discordance for the four main topologies (A-D) recovered with concatenated and species tree analyses (see also Fig. 2).

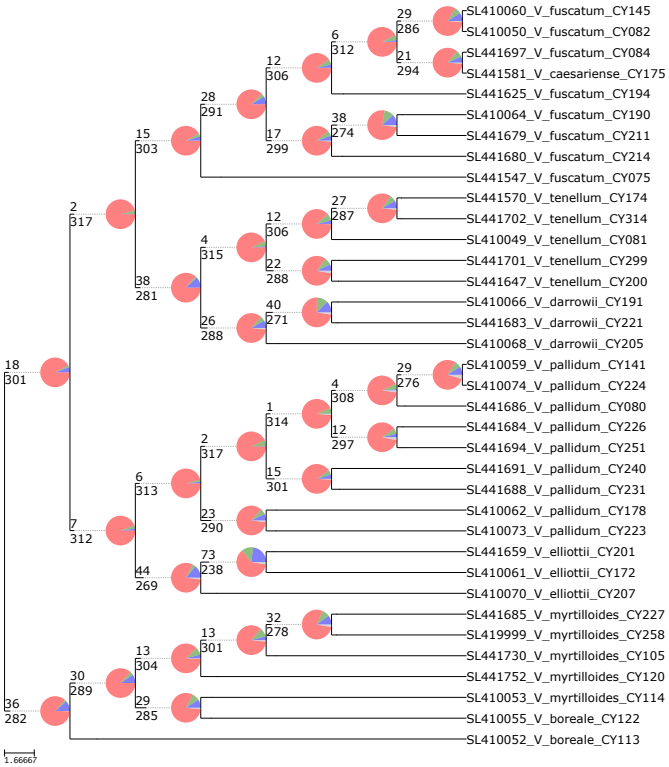
(A) iqtree (IUPAC)



(B) SVDQuartets (IUPAC)



(C) ASTRAL (IUPAC)



(D) ASTRAL (alleles)