Using target sequence capture to improve the phylogenetic resolution of a rapid radiation in New Zealand *Veronica*

Anne E. Thomas[1,5] iD , Javier Igea[1], Heidi M. Meudt[2], Dirk C. Albach[3], William G. Lee[4], and Andrew J. Tanentzap[1]

PREMISE: Recent, rapid radiations present a challenge for phylogenetic reconstruction. Fast successive speciation events typically lead to low sequence divergence and poorly resolved relationships with standard phylogenetic markers. Target sequence capture of many independent nuclear loci has the potential to improve phylogenetic resolution for rapid radiations.

METHODS: Here we applied target sequence capture with 353 protein-coding genes (Angiosperms353 bait kit) to *Veronica* sect. *Hebe* (common name hebe) to determine its utility for improving the phylogenetic resolution of rapid radiations. *Veronica* section *Hebe* originated 5–10 million years ago in New Zealand, forming a monophyletic radiation of ca 130 extant species.

RESULTS: We obtained approximately 150 kbp of 353 protein-coding exons and an additional 200 kbp of flanking noncoding sequences for each of 77 hebe and two outgroup species. When comparing coding, noncoding, and combined data sets, we found that the latter provided the best overall phylogenetic resolution. While some deep nodes in the radiation remained unresolved, our phylogeny provided broad and often improved support for subclades identified by both morphology and standard markers in previous studies. Gene-tree discordance was nonetheless widespread, indicating that additional methods are needed to disentangle fully the history of the radiation.

CONCLUSIONS: Phylogenomic target capture data sets both increase phylogenetic signal and deliver new insights into the complex evolutionary history of rapid radiations as compared with traditional markers. Improving methods to resolve remaining discordance among loci from target sequence capture is now important to facilitate the further study of rapid radiations.

KEY WORDS discordance; noncoding sequences; phylogenomics; Plantaginaceae; polyploidy.

Recent, rapid species radiations present rich opportunities for improving our understanding of the mechanisms involved in lineage diversification (Abe and Lieberman, 2009; Drummond et al., 2012; Hughes and Atchison, 2015). However, rapid radiations also present a challenge for phylogenetics. Fast successive speciation events often lead to low sequence divergence and poorly resolved relationships with traditional phylogenetic markers, which are typically slow-evolving plastid or nuclear ribosomal (nrDNA) genes (Nicholls et al., 2015;

Fernández-Mazuecos et al., 2018; Bagley et al., 2020; Larridon et al., 2020). High levels of introgression and incomplete lineage sorting (ILS, where the coalescence of alleles into a common ancestor is older than previous speciation events) further complicate the phylogenetic reconstruction of rapid radiations, and a small number of markers are often unlikely to recover the true species history amidst this kind of noise (Degnan and Rosenberg, 2009; Fernández-Mazuecos et al., 2018; Jones et al., 2019; Larridon et al., 2020; Stubbs et al., 2020).

Target sequence capture of large numbers of independent nuclear loci now has the potential to improve phylogenetic signal for rapid radiations compared with traditional markers. This high-throughput approach using RNA or DNA baits to extract genomic loci can provide orders of magnitude more genetic data, significantly boosting the number of phylogenetically informative sites (Nicholls et al., 2015). Nuclear loci also tend to be more variable than plastid loci, particularly when noncoding regions are also captured alongside protein-coding genes (McKain et al., 2018). The independence of nuclear loci also contrasts with nonrecombining plastid loci, which are inherited as a unit (Nicholls et al., 2015). All of these features can potentially help to counteract the effects of both low phylogenetic signal and signal-obscuring complexities like ILS and hybridization in rapid radiations. Another advantage of target capture is its effectiveness with degraded DNA, such as from herbarium specimens, due to the specificity of baits and short length of targeted sequences, thereby expanding sources for taxon sampling (Hart et al., 2016; McKain et al., 2018; Brewer et al., 2019; Shee et al., 2020). Finally, universal probe kits for targeted sequencing, such as the Angiosperms353 bait kit (hereafter Angiosperms353), have further increased the cost-effectiveness, feasibility, and standardization of the target capture method. Angiosperms353, which captures 353 putatively single-copy, protein-coding genes common to all angiosperms (Johnson et al., 2018), has been shown to be as robust as taxon-specific kits (Larridon et al., 2020). Angiosperms353 has been used to improve the resolution of phylogenies for rapid radiations including Papuasian *Schefflera* (Shee et al., 2020), *Cyperus* (Larridon et al., 2020), and *Nepenthes* (Murphy et al., 2020).

Despite its promise, target capture-based phylogenetics also faces several challenges. Although the genes targeted are putatively low- or single-copy, it is difficult to avoid paralogy, particularly in polyploid plants (Morales-Briones et al., 2018; Jones et al., 2019; Stubbs et al., 2020; Siniscalchi et al., 2021). The inclusion of paralogs can increase conflicting phylogenetic signal among genes. Even when paralogs are screened out where detected, the inherent variability among nuclear loci that can sometimes boost phylogenetic signal can also increase noise if gene trees are discordant due to factors such as ILS, causing well-supported but conflicting topologies and hampering phylogenetic resolution (Degnan and Rosenberg, 2009; Gernandt et al., 2018; Morales-Briones et al., 2018; Bagley et al., 2020). Ideally, this potential complexity can also be leveraged to gain better insight into evolutionary history (Murphy et al., 2020; Stubbs et al., 2020). Metrics such as concordance factors, which quantify how well a species tree is supported by individual genes and identify where alternative topologies may be strongly supported, can help in that regard. Another challenge occurs when individual loci have low information content. Although many loci still collectively contain more information than standard markers do, the true phylogenetic signal in this information can potentially be swamped by stochasticity resulting from nonphylogenetic signal such as homoplasy, model error, and methodological artifacts (Philippe et al., 2011; Townsend et al., 2012; Mclean et al., 2019). Thus, screening loci based on alignment

quality, information content, and, less commonly used, concordance with the best-supported species tree topology may ultimately help to improve the overall resolution of a target capture-derived phylogeny (Herrando-Moraira et al., 2018; Jones et al., 2019).

Here, our general aim was to quantify the phylogenetic resolution that universal target capture methods provide for a rapid radiation and determine how best to prepare and leverage data to improve these outcomes. We focused our analyses on *Veronica* sect. *Hebe* (Plantaginaceae), the largest plant radiation in New Zealand. The group is well suited for our analyses because it originated 5–10 million years ago, forming a monophyletic radiation of ca 130 extant species (Meudt et al., 2015). These species are highly diverse in morphology and ecological niche, presenting a useful group for exploring additional evolutionary and biogeographical questions (Albach and Meudt, 2010). Recent molecular phylogenetic studies have shown that several previously separate New Zealand genera are included within *Veronica*, whereby the following informal names were proposed for these subclades within section *Hebe* (Fig. 1; Wagstaff and Garnock-Jones, 1998; Albach and Meudt, 2010): the core hebes (originally genus *Hebe*, 88 species sensu Bayly and Kellow, 2006) that range from subshrubs with "whipcord" habit (scale-like leaves) to small trees; speedwell hebes (previously *Parahebe*, 13 New Zealand species and 11 from New Guinea not included in this analysis, Garnock-Jones and Lloyd, 2004), herbs or subshrubs closer in morphology to the northern hemisphere herbaceous *Veronica* habit; sun hebes (previously *Heliohebe,* 5 species, Garnock-Jones, 1993), decumbent subshrubs; snow hebes (previously *Chionohebe* and some *Parahebe*, 10 species; Meudt, 2008; Meudt and Bayly, 2008), alpine cushion plants; and semi-whipcord hebes (previously *Leonohebe,* 4 species sensu Bayly and Kellow, 2006), also with a whipcord habit. Past studies based on chloroplast and ribosomal internal transcribed spacer (ITS) markers have achieved high bootstrap support (BS) for monophyly of each subclade with limited phylogenetic resolution within these groups due to lack of variation in the markers (Albach and Meudt, 2010; Meudt et al., 2015). Previous efforts to develop low-copy nuclear markers for *Veronica* failed to generate markers optimized for phylogeny reconstruction (Mayland-Quellhorst et al., 2016). Target capture sequencing with universal probes, while not tailored to the genus, is increasingly well validated and promises a higher output of genetic data. No study of *Veronica* has attempted to explore the phylogenetic resolution of the section or the larger genus using phylogenomic data.

In this study, we compared the phylogenetic resolution and support of trees based on the Angiosperms353 bait kit versus traditional markers for the rapid radiation of *Veronica* sect. *Hebe* in New Zealand. Using the Angiosperms353 bait kit, we aimed to identify which subsets of the data provided the best phylogenetic signal by comparing coding, noncoding, and combined sequences and by filtering individual genes. We identified levels of conflict among the loci using concordance factors. Finally, we determined how well target capture loci were able to resolve relationships at different taxonomic scales and among previously identified hebe subclades.

MATERIALS AND METHODS

Tree estimation with existing phylogenetic data

To compare the effectiveness of target capture markers to traditional markers, we collated previously published genetic data for

FIGURE 1. Examples of *Veronica* species from subclades mentioned in the text. (A) Snow hebe: *Veronica chionohebe,* CC BY-NC-ND 4.0. Te Papa (WELT SP084043). (B) Sun hebe: *Veronica raoulii*, CC BY-NC Alex Fergus. (C) Speedwell hebe: *Veronica colostylis,* CC BY-NC-ND 4.0. Te Papa (WELT SP107460). (D) Semi-whipcord hebe: *Veronica quadrifaria*, CC BY 4.0. Te Papa (WELT SP104000). (E–G) Core hebes: *Veronica flavida*, CC BY 4.0. Te Papa (WELT SP103974); *Veronica pinguifolia*, CC BY 4.0. Te Papa (WELT SP103950); *Veronica propinqua* (a whipcord hebe), CC BY 4.0. Te Papa (WELT SP103821).

Veronica sect. *Hebe* from GenBank (https://www.ncbi.nlm.nih.gov/genbank/). We searched GenBank in November 2018 and selected the markers that were available for more than 15% of the 82 *Veronica* sect. *Hebe* species with published sequences, including ITS, chloroplast markers *trnL-trnF*, *rpoB-trnC*, and *rps16*, and a nuclear marker, *CYCLOIDEA* (Appendix 1). The remaining poorly represented markers, including *rbcL*, were not included in the final data set. Three species with sequences in GenBank but absent in our target capture data set were also removed, for a total of 79 species in this analysis. We used PyPHLAWD (Smith and Walker, 2019), a phylogenetic data set-building software, to retrieve the markers.

The GenBank sequences retrieved by PyPHLAWD were aligned for each marker using MAFFT v7.419 (Katoh and Standley, 2013) with the --auto setting. Alignments were trimmed with TrimAl v1.4 (Capella-Gutierrez et al., 2009) to remove sites with less than 30% occupancy, and the sequences were then concatenated. IQ-TREE v1.6 (Nguyen et al., 2015) with ModelFinder (Kalyaanamoorthy et al., 2017) was used to identify the best-fitting models of sequence evolution for the individual marker partitions, followed by the optimal partitioning scheme based on the Bayesian information criterion by merging partitions with the -MF-MERGE option. Combining partitions, if suitable, helps prevent model overspecification (Chernomor et al., 2016; Kalyaanamoorthy et al., 2017).

IQ-TREE was then used to estimate a maximum-likelihood phylogeny from the best partition scheme, which consisted of separate partitions for ITS and *CYCLOIDEA* and a single partition for the chloroplast markers. Support was assessed with 100 standard bootstraps.

Plant samples

Tissue samples from 77 of 130 New Zealand *Veronica* species and one Australian species (*Veronica perfoliata* from subgen. *Pseudoveronica* sect. *Labiatoides*, to be used as an outgroup) were secured from the herbarium collections of the Museum of New Zealand Te Papa Tongarewa (WELT) and University of Oldenburg (Carl-von-Ossietzky Universität Oldenburg; OLD). Samples dried in silica gel subsequent to field collection ($n = 66$) were prioritized for DNA extraction and sequencing due to better chances of high-quality yield. Remaining samples ($n = 12$) were taken from herbarium sheets at WELT. The median collection year of silica gel samples was 2014, and the oldest was from 1998, while the median collection year of the herbarium sheet samples was 2001, and the oldest was from 1997 (Appendix 1). One sample, a northern hemisphere outgroup species, *Veronica chamaedrys* (subgen. *Chamaedrys*), was collected in the field in Cambridge, UK, in 2019 and dried in silica gel.

DNA extraction

Leaf samples of 1 to 2 cm² were homogenized using a TissueLyser II bead mill with 2 mm or 3 mm steel beads. The DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) was used to extract DNA following the manufacturer's protocol with the following changes: polyvinylpyrrolidone was added during the lysis step to remove phenolic compounds, and lysis incubation time was extended to at least 15 min. DNA concentration was quantified using a Qubit Fluorometer 3.0 (Fisher Scientific, Loughborough, UK), and the level of fragmentation was assessed with electrophoresis in 1.25× Tris-borate-EDTA (TBE) agarose gels.

Target enrichment: library preparation

Libraries for target enrichment were prepared by first sonicating DNA samples diluted in 1× TE buffer to approximately 350-bp fragments using a Covaris E220 Focused-ultrasonicator with Covaris microTUBES AFA Fiber Pre-Slit Snap-Cap (Covaris, Woburn, MA, USA), based on the manufacturer's protocol. The NEBNext Ultra II DNA Library Prep Kit for Illumina was used with NEBNext Multiplex Oligos for Illumina (New England BioLabs, Ipswich, MA, USA) to prepare single-indexed DNA libraries for multiplexed sequencing. During library prep, we size-selected samples for 300–400 bp with purification beads according to the protocol, except for 13 low-input samples which were not size-selected to maintain library complexity. PCR amplification was done at half-volume, requiring for most libraries an additional three cycles beyond the recommended 3–7 cycles. Otherwise, we followed the manufacturer's protocol for PCR conditions. Libraries were quantified using Qubit and assessed for size distribution and quality using the Agilent 2100 Bioanalyzer High-Sensitivity DNA assay (Agilent Technologies, Santa Clara, CA, USA).

Target enrichment: hybridization

The Daicel Arbor Biosciences (Ann Arbor, MI, USA) myBaits Target Capture Kit, Angiosperms353 v1, was used following the manufacturer's protocol (ver. 4) to enrich the indexed libraries. We performed hybridization and sequencing in two batches following the same protocol. Libraries were pooled and concentrated using a ThermoScientific Savant DNA 120 SpeedVac (Fisher Scientific). As baits have been found to be equally effective at lower volumes than suggested by the manufacturer (Hale et al., 2020), hybridization reactions were performed with baits diluted 1:1 relative to the recommended concentration in nuclease-free water. The reactions were incubated in a SimpliAmp Thermal Cycler at 65°C for 23–24 h. Amplification was performed for 18 cycles with KAPA HiFi 2X HotStart ReadyMix PCR Kit (Roche, Basel, Switzerland) and Illumina adaptor reamplification primers (Meyer and Kircher, 2010; IS5_reamp.P5 5′-AATGATACGGCGACCACCGA-3′, IS6_reamp.P7 5′-CAAGCAGAAGACGGCATACGA-3′). PCR products were purified using the Monarch DNA and PCR Cleanup Kit (New England BioLabs). An additional Bioanalyzer assay confirmed successful enrichment and amplification.

Sequencing and sequence data analysis

Enriched libraries were multiplexed and sequenced on an Illumina NextSeq 500 with a v 2.5 Mid-Output kit for 2 × 150-bp paired-end reads. In addition to trimming adapters, we removed low-quality bases and reads using Trimmomatic following the method of Johnson et al. (2018; settings: Leading: 20, Trailing: 20, Sliding Window: 4:20, Minlen: 50). We used the Hyb-Seq (Weitemier et al., 2014) bioinformatics pipeline HybPiper v1.2 with BWA v0.7.17 (Li and Durbin, 2009) to map and align reads to the Angiosperms353 loci, assemble loci with SPAdes v3.13.0 (Bankevich et al., 2012), and extract gene sequences with exonerate v2.4.0 (Slater and Birney, 2005; Johnson et al., 2016). In addition to retrieving the default output of HybPiper, which consists of assembled protein-coding sequences (exons) in the form of both nucleotides and amino acids, we extracted flanking noncoding intron and intergenic sequences with the HybPiper script intronerate.py. We combined exons and introns into a concatenated sequence for each gene with the supercontig setting of the script retrieve_sequences.py. We also identified potential paralogs with the HybPiper script paralog_investigator.py, which flags genes with multiple assembled contigs that were at least 85% of the reference sequence length.

As an exploratory analysis to uncover additional heterozygosity potentially due to paralogs and masked by consensus sequences, we applied the allele phasing pipeline described by Kates et al. (2018; available at https://github.com/mossmatters/phyloscripts/tree/master/alleles_workflow). This pipeline maps the raw reads to the supercontig consensus sequences with BWA-MEM v0.7.17 (Li, 2013 [Preprint]), identifies variant sites with GATK v3.7.0 HaplotypeCaller (McKenna et al., 2010), matches overlapping variant reads to produce separate sequences for the top two haplotypes with WhatsHap v0.18 (Patterson et al., 2015) and the pipeline script haplonerate.py, and aligns each haplotype across samples with MACSE v1.2 (Ranwez et al., 2011) and MAFFT. We estimated gene trees for these alignments with FastTree v2.1.3 (Price et al., 2010) to determine whether haplotypes from individual samples grouped together or whether separate clades formed, indicating paralogous lineages.

Filtering and preparing alignments

Downstream analyses were separately performed on the exon-only, intron-only, and combined supercontig subsets (referred to as "gene subsets") under several filtering schemes (see below).

The sequences retrieved by HybPiper underwent several filtering and cleaning steps before and after alignment to improve data coverage and reduce noise (Fig. 2). Before alignment, genes with less than 50% occupancy across the taxa (n = 41 loci) and with paralog warnings for >10% of the samples (n = 42 loci) were excluded. The remaining 270 genes were aligned using MAFFT --auto. Nucleotide and amino acid alignments for protein-coding exons were combined using PAL2NAL (Suyama et al., 2006), resulting in a codon-guided "inframe" alignment. All alignments were trimmed using TrimAl: for inframe exons, sites with less than 50% occupancy or less than 0.001 site identity were trimmed, while for the noisier supercontig and intron alignments, sites with less than 70% occupancy

FIGURE 2. Steps taken in alignment, cleaning, and filtering of loci. The main filtering schemes compared in the analysis are shown (full, intersection, SortaDate_BP, and SortaDate_TL). Exons, introns, and supercontigs as assembled by HybPiper were trimmed and filtered separately after alignment and subjected to the same filtering thresholds, except in the case of gap trimming where introns and supercontigs had a higher threshold due to higher alignment noise.
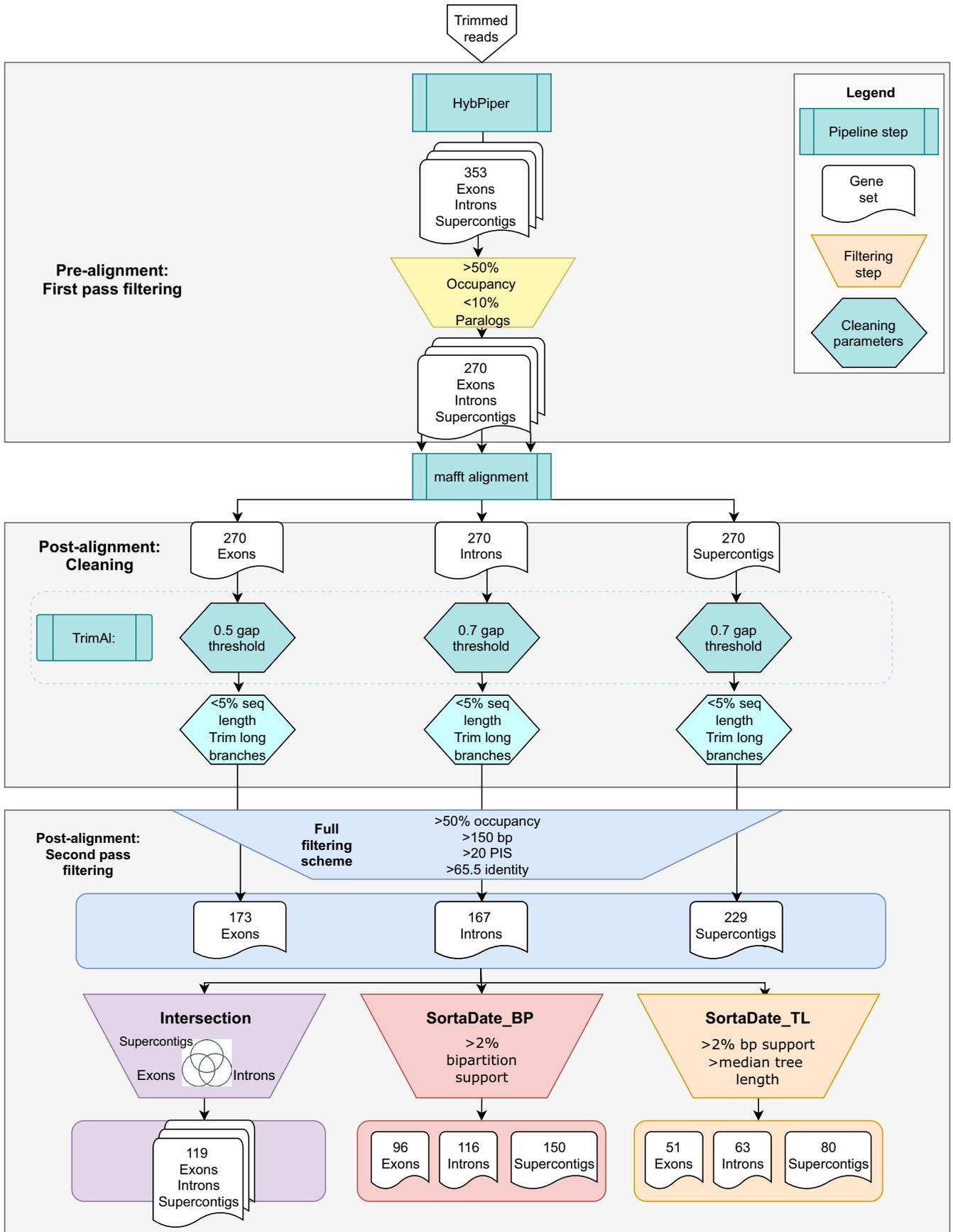
were trimmed. Sequences with less than 5% of alignment length were then removed. For identifying extremely divergent sequences within alignments, gene trees were estimated using IQ-TREE. Branches that were 10 times longer than their sister branch or longer than 0.5 substitutions/site were trimmed using the trimtips. py script from the Phylogenomic Dataset Construction pipeline (Yang and Smith, 2014), which we modified to prune both internal branches (including subtending tips) and terminal branches. The sequences corresponding to the trimmed branches were also removed from the original alignments used to generate the gene trees before tip trimming. Final gene trees were re-estimated from the reduced alignments using IQ-TREE.

Another set of filtering thresholds was applied to each gene subset separately after alignment, which produced the "full" filtering scheme data set (Fig. 2). To improve alignment quality, we calculated the average percentage identity of the alignments using TrimAl and removed genes with a score of less than 65.5%, as these genes tended to be poorly aligned with large gaps (Villaverde et al., 2018). To filter out genes with little phylogenetic information, we removed genes with alignment length less than 150 bp and with fewer than 20 parsimony informative sites (PIS), as calculated with the R package phyloch (Heibl, 2008). We also removed genes with taxon occupancy reduced to less than 50% by the alignment cleaning. To test the effect of retaining genes with low (<10%) rates of paralog warnings, we created an additional data set with these genes removed from the full supercontig data set, leaving 195 supercontig sequences.

Additional filtering schemes

Three additional filtering schemes were applied to the "full" data set (Fig. 2). First, differences in alignment trimming and information content of sequences in exons, introns, and supercontigs resulted in a slightly different set of genes in each gene subset of the "full" filtered data set. To ensure that differences between trees were not strongly influenced by this difference, we applied an "intersection" filtering scheme that included only the genes found in every gene subset in the "full" data set. We also generated two additional filters from individual gene tree statistics calculated with SortaDate (Smith et al., 2018), including bipartition support (the percentage of bipartitions shared between the gene tree and species tree) and tree length. The SortaDate_BP filtering scheme removed genes from the "full" data set with less than 2% bipartition support and below a minimal tree length threshold (0.1 substitutions per site for exons, 0.5 for introns and supercontigs). The stricter SortaDate_TL filtering scheme removed trees with both low bipartition support and length below the median value for the full version of each subset (0.64 substitutions per site for exons, 1.11 for introns, and 0.79 for supercontigs).

Tree building

For each gene subset and filtering scheme, individual alignments were concatenated and initially partitioned by gene. IQ-TREE's ModelFinder was used with the -MF-MERGE option as above to generate optimal partitioning schemes (Kalyaanamoorthy et al., 2017). The search space for the possible partition combinations was reduced to 10% with the -rcluster option for computational feasibility (Lanfear et al., 2012). IQ-TREE v1.6 was then used to estimate maximum-likelihood phylogenies from these partition schemes. Initial runs with ultrafast bootstrap (UFBoot2; Hoang et al., 2018) in IQ-TREE v1.6 did not converge after thousands of iterations, so we performed a standard bootstrap test with 100 bootstraps instead. For the exon data set under the full filtering scheme, some unmerged exon partitions were too uninformative on their own, leading to numerical instability in the bootstraps. Single-locus partitions with fewer than 50 PIS were removed prior to final phylogeny estimation, reducing the exon data set by two genes.

ASTRAL-III, a summary method consistent with the multispecies coalescent (MSC) process (Zhang et al., 2018), was also used to estimate species trees for the target capture data sets. ASTRAL-III was run with gene trees estimated with IQ-TREE v1.6 for all gene subsets and filtering schemes. Local posterior probabilities (LPP) for quartet support and the normalized quartet score for the species tree were subsequently calculated. For comparison, SVDquartets (Chifman and Kubatko, 2014), a MSC-based method that estimates species quartets from individual sites rather than gene trees, was run for the full supercontig data set. Multilocus nonparametric bootstraps were calculated to determine quartet support.

To visualize conflict in the full supercontig data set, we calculated and visualized a ConsensusNetwork from splits occurring in 10% or more of the IQ-TREE bootstrap trees and mean edge weights using SplitsTree4 v4.16.2 (Huson and Bryant, 2006).

Tree metrics

To supplement BS, we calculated gene and site concordance factors (gCF and sCF, respectively) implemented in IQ-TREE v2 (Minh et al., 2020a, b). The gCF represents how many gene trees share a given bipartition in the species tree out of those that possibly could contain it ("decisive" gene trees). A gCF of 0 corresponds to no gene trees supporting a bipartition. The sCF represents how many sites in the gene alignments support a particular quartet arrangement in the species tree. A sCF of around 33% indicates that all of the quartets are equally likely, meaning that the alignments are uninformative, and values lower than ~30% are unlikely, although possible in the "anomaly zone", where the species tree has a bipartition that is not the one supported by the most sites. IQ-TREE also calculates discordance factors for both genes and sites (gDF and sDF, respectively). The gDF quantifies support for the two nearest-neighbor interchange bipartitions (gDF1 and gDF2) and for all other possible topologies (gDFP, as these are paraphyletic relative to the species tree bipartition). The sDF metrics quantify support among sites for the two possible alternative quartets (sDF1 and sDF2). Low gDF1 and gDF2 values or high gDFP values suggest the gene trees or alignments lack a clear signal, as do sDF values close to 33% (Minh et al., 2020a).

To compare branch support values in different sections of the trees, we calculated node depth on trees rooted with the root function of the R package ape v5.4 (Paradis and Schliep, 2019) in R v4.0.2 using *Veronica chamaedrys* and *Veronica perfoliata* as outgroups. We set the edgelabel option to TRUE to ensure bipartition support corresponded to the correct branches after rooting (Czech et al., 2017). Internal nodes were classified as shallow if they were a parent of a terminal node, whereas all other nodes were considered deep.

Tree similarity was measured using the mutual clustering information (MCI) metric implemented in the R package TreeDist v1.1.1 (Smith, 2020), an alternative to the Robinson–Foulds distance. The MCI metric quantifies how much information is shared between

bipartitions in the trees being compared rather than requiring bipartitions to be identical to count as shared. TreeDist also normalizes the metric from 0 to 1, with 1 corresponding to identical trees and 0 corresponding to trees with no bipartitions in common.

Most figures were plotted using ggplot2 v3.3 (Wickham, 2016) and ggpubr v0.4 (Kassambara, 2020) in R. Trees were plotted using ape v5.4 in R or with Dendroscope v3.7.4 (Fig. 5 inset) (Huson and Scornavacca, 2012) and labels adjusted in Inkscape v1.0.1, and the analytical flowchart (Fig. 2) was created in the web-based tool diagrams.net.

RESULTS

Target capture results

Overall, we obtained approximately 150 kbp of 353 protein-coding exons and an additional 220 kbp of flanking noncoding sequences for each of the 77 hebes and two *Veronica* outgroups. The mean number of reads mapping to sequences was 580,000 reads or a mean of 20% of total reads on target (on par with other studies using this kit, e.g., Johnson et al., 2018). Of the 353 genes, 345 were recovered for one or more samples. The overall mean number of loci with recovered sequences was 314 per sample and did not differ between herbarium and silica-dried samples (two-sample *t*-test, $t = 0.68$, df = 77, $P = 0.501$). Allele phasing did not reveal a clear pattern in the grouping of haplotypes, with some grouping by individual and others spread throughout the gene tree.

Improved phylogenetic support of target capture over traditional markers

All the Angiosperms353 target capture data sets increased phylogenetic information and branch support for *Veronica* sect. *Hebe* relative to traditional markers (Table 1). The final alignments from the target capture data sets had two to three orders of magnitude more PIS than the final GenBank alignment (Table 1). Excluding one data set with low support (i.e., exons under the SortaDate_TL filtering scheme), target capture trees had between 9.2% and 34.2% more bipartitions in IQ-TREE with high BS (at least 80%) than the GenBank tree (Table 1).

Phylogenetic support from different subsets of target capture data

Comparison of gene subsets—Across all filtering schemes, the combined sequence information in the supercontig data set outperformed exons or introns alone in BS with IQ-TREE. Supercontigs had between 21.5 and 33% higher median bootstrap values than exons and between 5.5 and 23% higher median bootstrap values than introns (Fig. 3A). However, supercontig gCF and sCF values were overall comparable to the other subsets, with medians differing by 1% or less (Fig. 3C, D). In the ASTRAL-III supercontig trees, the upper quartile of median LPP was up to 25% higher than exons and 21% higher than introns. However, median LPP was less distinct overall between gene subsets, up to 11.5% higher than in exons and up to 7.5% higher than in introns (Appendix S1). ASTRAL-III gCF scores were similar across all data sets (Appendix S1).

Comparison of filtering schemes—Filtering gene subsets with different thresholds produced slightly varying data set sizes and information content. The intersection filtering scheme reduced the full data set to 119 genes common to each subset with little change in % PIS (Table 1). The number of genes retained by the Sortadate_BP filtering scheme (>2% bipartition support) varied across gene subsets, but % PIS again remained comparable. The SortaDate_TL filter (median tree length threshold) removed the most genes, reducing the full data set by between 62 and 71%, and more consistently increased % PIS (Table 1).

Filtering genes based on bipartition support (SortaDate_BP) or overlap between gene subsets (intersection) did not visibly improve branch support relative to the full filtering scheme in either IQ-TREE or ASTRAL-III. Branch support values were similar within gene subsets across the full, SortaDate_BP, and intersection data sets, with IQ-TREE medians differing by at most 8.5% in bootstraps, 1.0% in gCF, and 0.8% in sCF (Fig. 3). The percentage of bipartitions with at least 80% BS also remained similar among these three data sets in exons, but support varied among introns and supercontigs depending on the filtering scheme (Table 1). The additional filtering by gene tree length (SortaDate_TL) yielded similar gCF and sCF values to the other data sets, but notably decreased BS in exons and introns, suggesting that low-information genes still contribute to resolution within these species tree topologies (Fig. 3). Similar patterns were observed in ASTRAL-III trees (Appendix S1).

TABLE 1. Angiosperms353 target capture markers deliver more phylogenetic information and support than traditional markers for *Veronica* sect. *Hebe*. Comparison of GenBank markers and target capture markers for the three gene subsets (exons, introns, supercontigs) and four filtering schemes (full, intersection, SortaDate_BP, and SortaDate_TL) by number of markers, total sites, parsimony informative sites (PIS), and percentage of bipartitions with ≥80% bootstrap support (BS) in IQ-TREE.

| Gene subset	Filtering scheme	No. markers	Total no. sites (bp)	No. PIS	% PIS	% Bipartitions ≥80% BS
GenBank	NA	5	4184	266	6.4	19.7
Exons	Full	173	107,112	15,129	14.1	27.6
	Intersection	119	74,784	10,347	13.8	28.9
	SortaDate_BP	96	66,495	9632	14.5	28.9
	SortaDate_TL	51	39,738	6944	17.5	19.7
Introns	Full	167	85,644	21,677	25.3	35.5
	Intersection	119	71,340	18,734	26.3	34.2
	SortaDate_BP	116	65,631	17,117	26.1	46.1
	SortaDate_TL	63	43,927	13,076	29.8	27.6
Supercontigs	Full	229	206,654	36,937	17.9	53.9
	Intersection	119	139,076	24,818	17.8	46.1
	SortaDate_BP	150	160,786	30,146	18.7	51.3
	SortaDate_TL	80	95,859	20,384	21.3	44.7
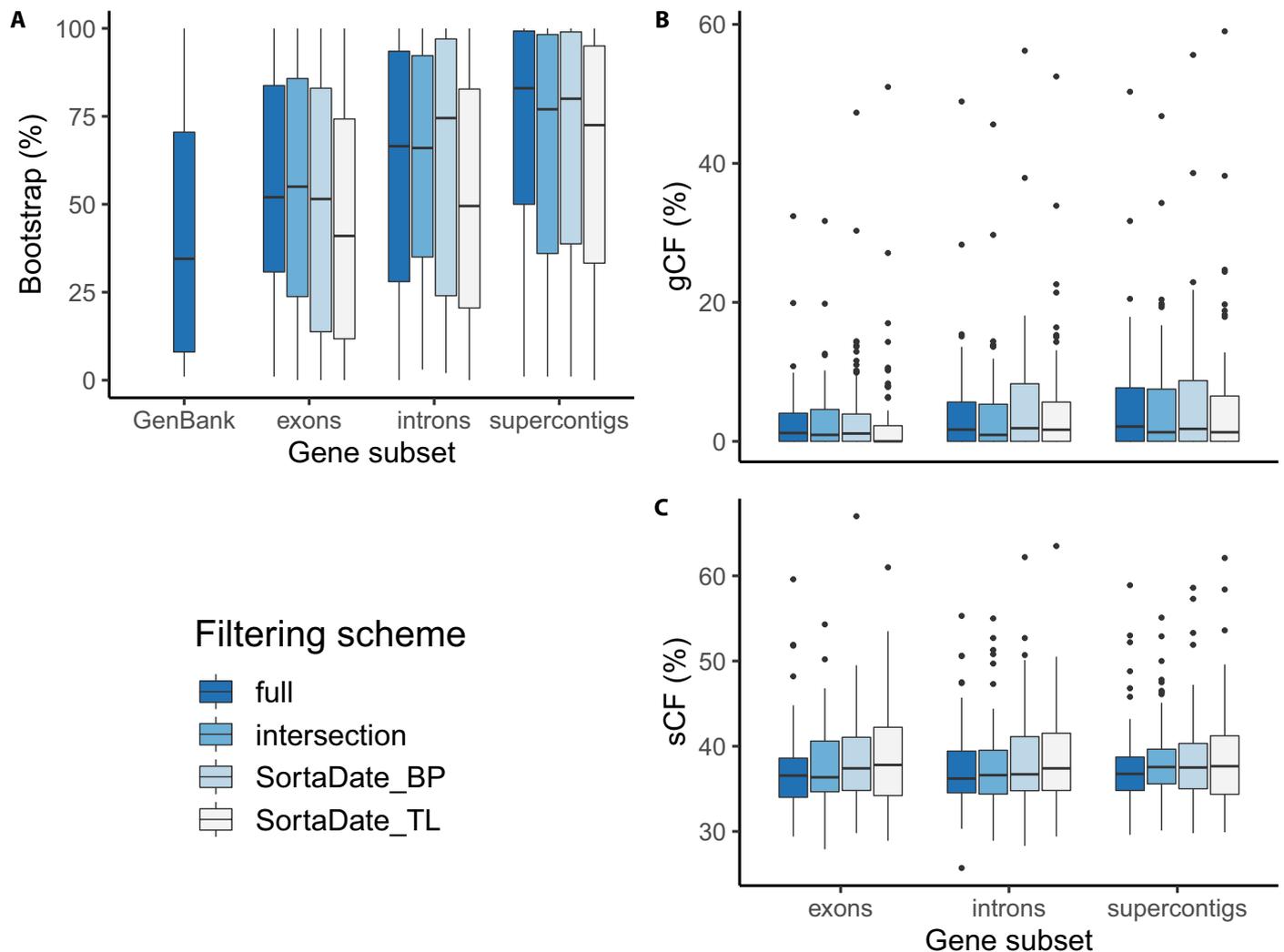
FIGURE 3. Supercontigs provide highest bootstrap support (BS) irrespective of filtering scheme without improving concordance. Phylogenies were estimated using exons, introns, and supercontigs (combined) gene subsets using the original full data set, intersection (genes present in each gene subset), SortaDate_BP (filtered by bipartition support) and Sortadate_TL (filtered by bipartition support and tree length) filtering schemes (see Table 1 for details). Boxplots of branch support values for all bipartitions in each combination of filtering scheme and gene subset show (A) BS, (B) gene concordance factor (gCF), and (C) site concordance factor (sCF). Boxes show quartiles; whiskers show minimum and maximum excluding outliers that are outside of 1.5-times interquartile range, shown by points.

Compared to IQ-TREE, ASTRAL-III trees had equivalent or lower branch support scores and an average overall normalized quartet score of 0.42 (Appendices S1, S2). SVDquartets also had lower BS than IQ-TREE (Appendix S3). Additionally, removing genes with low rates of paralog warnings from the full supercontig data set did not change the distribution of concordance factors or BS values except to decrease the lower quartile of BS (Appendix S4). Given the similarity among support values for filtering schemes and methods, the remaining analyses focused on the full IQ-TREE data set for exons, introns, and supercontigs.

Concordance factors vs. bootstrap support—Regardless of the data set, gene and site concordance factors were generally low, indicating poor concordance between gene trees despite relatively high IQ-TREE bootstrap values in supercontigs data sets. Median gCF ranged from 0% to 2.13% across the data sets (Fig. 3B), reflecting bipartitions with few gene trees sharing the same topology.

The maximum gCF value in any data set was 59%. While low gCF values can be due either to strong support of alternative topologies or to a lack of clear signal in the trees, the discordance factors pointed to the latter. The gDF1 and gDF2 values were even lower than gCF, but gDFP for all other possible topologies was extremely high (median 96% for the full supercontigs tree), showing that in most cases there was no clear alternative topology (Appendix S5). Similarly, sCF values close to neutral (33% = no informative decisive sites) pointed to a lack of clear signal among sites. Site concordance factors were low across data sets, with medians between 36.2 and 37.8%, just above the neutral value (Fig. 3C). In the full data set supercontig phylogeny, high bootstrap, gCF, and sCF values were correlated ($r = 0.49$ for gCF and bootstrap, $P < 0.001$; $r = 0.54$ for sCF and bootstrap, $P < 0.001$), but high bootstrap values also occurred with low sCF and gCF values, illustrating that bootstrap values do not entirely capture the variation inherent in the data set (Fig. 4).
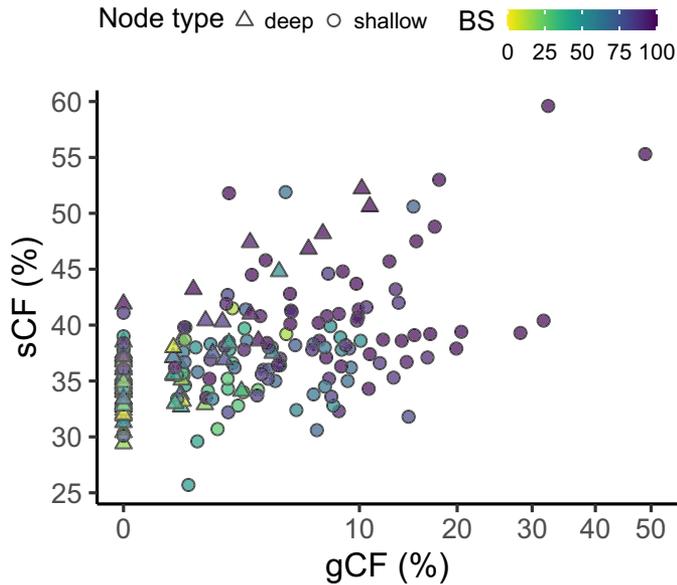
FIGURE 4. Bootstrap support (BS), gene concordance factor (gCF), and site concordance factor (sCF) are moderately correlated, but BS is high for some low concordance factors. Most high BS and gCF values were in shallow nodes, i.e., parents of terminal nodes. Points show each bipartition in the full data set supercontig phylogeny.

Phylogenetic support from target capture loci in different regions of the phylogeny

Support for subclades—The Angiosperms353 sequences largely supported the subclades identified by past genetic and morphological studies, with the highest median support for the small subclades (i.e., <12 species). In the supercontigs tree, the sun hebe and snow hebe subclades were strongly monophyletic, while the remaining groups were mostly monophyletic: 10 of 12 sampled speedwell hebe species formed a monophyletic group; the core hebe subclade (48 species sampled) was monophyletic when excluding *Veronica macrantha*, which has had uncertain taxonomic placement in past studies (Albach and Meudt, 2010; Meudt et al., 2015); and the semi-whipcord hebes (3 species) were monophyletic if the similarly uncertain *Veronica cupressoides* was excluded (Albach and Meudt, 2010; Fig. 5). The supercontigs subset showed the best overall support for most subclades (Fig. 7). In the supercontigs subset, the monophyletic groups (including the monophyletic portion of speedwell hebes) had BS values of 99% or 100% and were also preserved by the ConsensusNetwork analysis (Fig. 7). While Angiosperms353 recovered the main subclades with strong support and improved on GenBank markers for many nodes within subclades, resolution and support within subclades were variable, with a mix of strongly supported and uncertain relationships (Figs. 5, 6).

The relationships in the core hebes, the largest subclade and the most speciose part of the radiation, showed the most variation in support (Fig. 6). The best-supported relationship in the entire tree was found in this subclade, between the sister Chatham Island species *V. dieffenbachii* and *V. chathamica*, with the highest gCF in the tree at 50.3%, the highest sCF at 58.9%, and a bootstrap value of 100% (compared to a value of 55% BS in the GenBank tree) (Fig. 5). A third Chatham Island species, *V. barkeri,* grouped with them with a gCF of 17.59% and a sCF of 53.0%. There were several other instances of high support (Fig. 5). In many relationships, however, the gCF and sCF values in the core hebes exhibited high levels of uncertainty (Fig. 4; Appendix S6).

Uncertainty was exemplified in the ConsensusNetwork analysis, which showed many conflicting splits among the core hebes (Fig. 7).

Topologies also showed incongruence between gene subsets. The normalized MCI tree similarity (0 = no shared information between bipartitions, 1 = identical) was 0.55 between exons and introns, 0.60 between exons and supercontigs, and 0.63 between introns and supercontigs. The relationships between the subclades varied for every gene subset: the exons placed the core hebes as sister to the semi-whipcord, snow, and speedwell hebes collectively; the introns placed the core hebes as sister to the sun hebes; and the supercontigs placed the core hebes as sister to the semi-whipcord hebes. BS for almost all of these placements was below 80% and gCF was below 5%. The speedwell and snow hebes were sisters in every gene subset phylogeny, but BS was below 80% and several speedwell species were placed inconsistently outside the subclade. For example, in the speedwell hebes, the supercontigs tree placed *V. lilliputiana* as sister to the snow hebes with 78% BS, while the exons tree placed *V. lilliputiana* within the monophyletic speedwell hebes with 75% BS, and the introns tree additionally placed *V. planopetiolata*, *V. colostylis*, and *V. linifolia* in a paraphyletic group sister to the snow hebes (Fig. 5; Appendices S7, S8).

Shallow vs. deep node resolution—Our data showed that, overall, Angiosperms353 sequences were better suited to resolve shallower nodes (defined here as parents of terminal nodes) rather than deeper relationships of our rapid radiation. We specifically found that for target capture loci under the full filtering scheme, median gCF was 2.6% to 4.6% higher and median BS was 35% to 43% higher among the shallow internal nodes (51 nodes in the supercontig tree) than the deeper nodes of the tree (26 nodes) (Figs. 4, 8A, 8B). This difference was largely driven by uncertainty in the deepest nodes representing the backbone of the tree (Fig. 5), as several subclade nodes were well supported. The difference in support between shallow and deeper nodes was less pronounced in sCF for all gene subsets of the full data set (Fig. 8C). The GenBank tree had similar levels of BS for shallow versus deeper relationships, although the median bootstrap was still 18% higher for shallow nodes than for deep nodes (Fig. 4A). A greater contrast was seen in the lower quartiles of BS. Therein, the difference between shallow and deep nodes was only 5% for GenBank and between 29% and 44.5% for target capture trees, showing that lower bootstrap values were more likely to be found at any depth of the GenBank tree (Fig. 4A; Appendix S9).

Terminal vs internal branch lengths—Target capture markers accumulated more divergence, as represented by branch length, in the terminal than internal branches of the trees. The median length of terminal branches (excluding the outgroup *V. chamaedrys*) in the supercontig tree under the full filtering scheme was 0.02 substitutions per site, while the median length of internal branches was 0.003 substitutions per site (Fig. 8D). Other subsets showed a similar pattern (Fig. 8). The GenBank tree also had a higher median terminal branch length, 0.002 substitutions per site, than internal branch length, 0.001 substitutions per site, but in contrast to the target capture trees, the ranges of terminal and internal branch lengths overlapped considerably (Fig. 8D).

DISCUSSION

By using the Angiosperms353 bait kit to generate new phylogenomic data, our study improved the phylogenetic resolution of the rapid
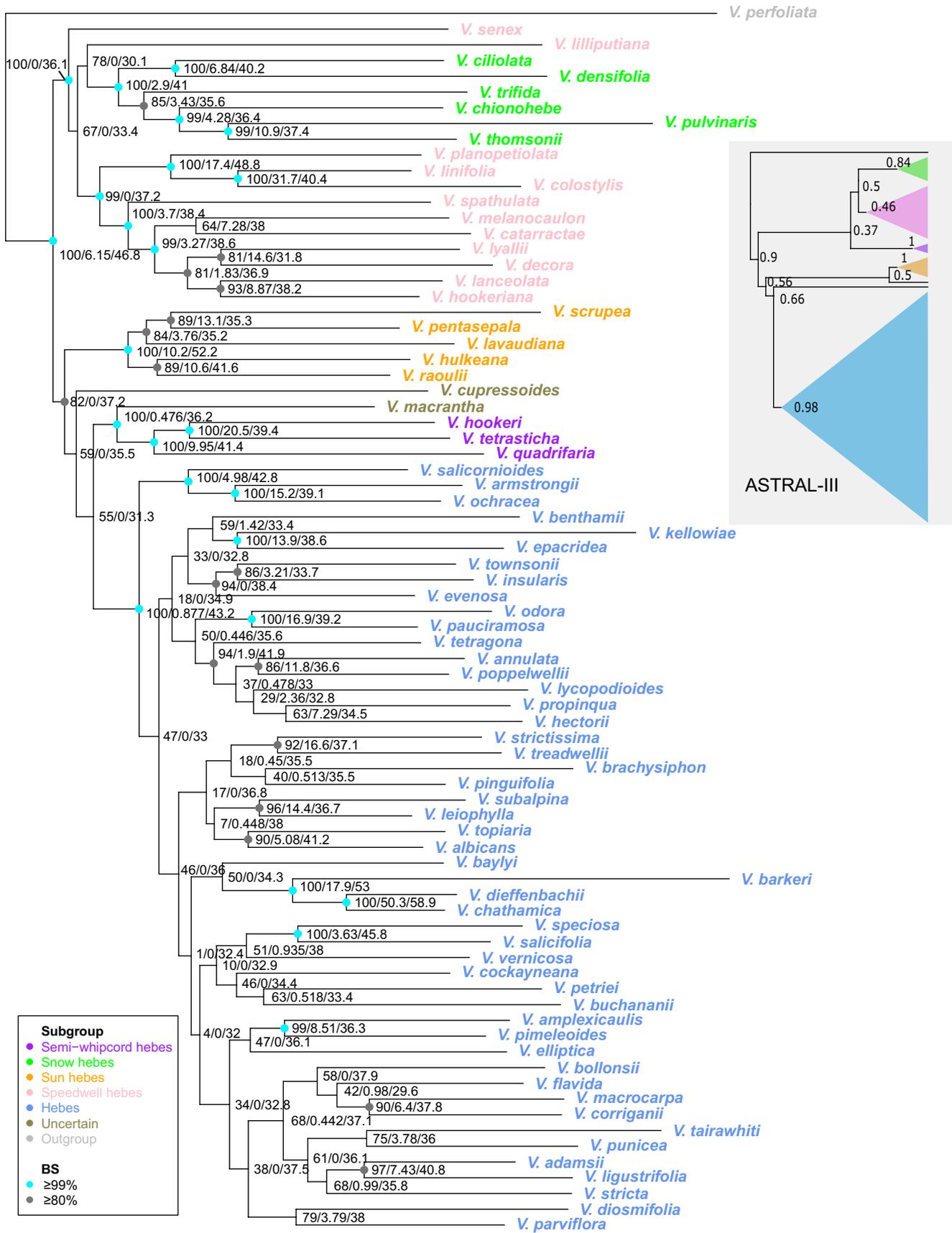
FIGURE 5. Subclades are mostly monophyletic in a phylogeny for *Veronica* subgen. *Pseudoveronica* sect. *Hebe*. Main figure: The phylogeny was estimated with IQ-TREE from 229 supercontig sequences with the full filtering scheme. V. *perfoliata* from subgen. *Pseudoveronica* sect. *Labiatoides* was used as an outgroup (in addition to *V. chamaedrys* from subgen. *Chamaedrys,* not shown). Nodes labelled with bootstrap support (BS)/gene concordance factor/site concordance factor support values. Subclades are color-coded (Albach and Meudt, 2010), and nodes with BS values at or above 80 and 99 are colored. Inset summarizes topology of subclades in tree estimated by ASTRAL-III with LPP support; details in Appendix S2.
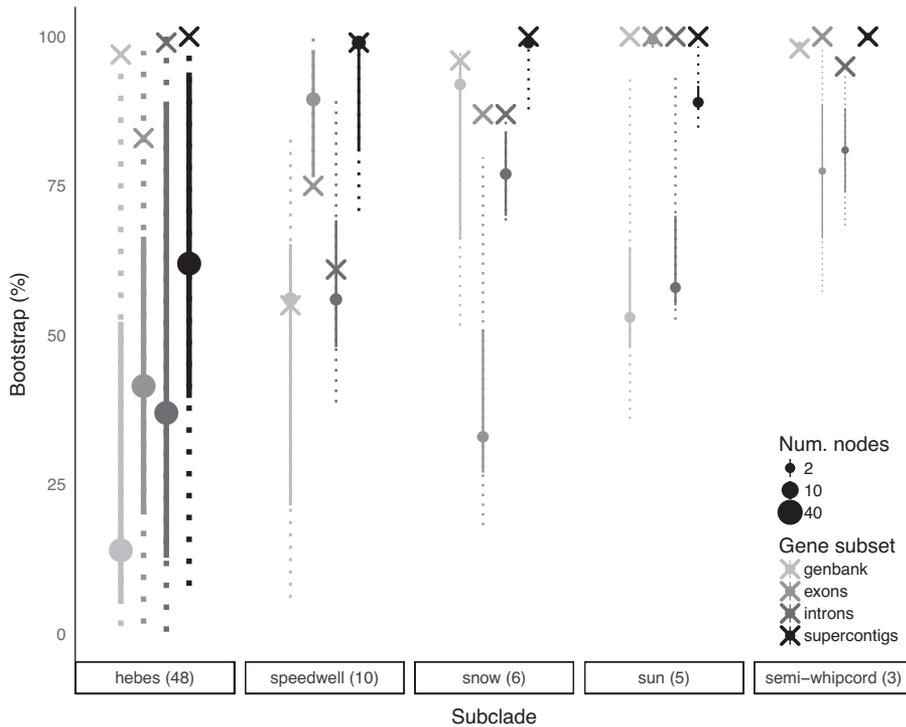

FIGURE 6. Small *Veronica* subclades have best overall support, mostly from supercontigs. Distribution of bootstrap support for monophyletic subclades in the full data set phylogeny, showing median, 25% and 75% quantiles with solid lines, and 5% and 95% quantiles with dotted lines. Support value of overall bipartition for each subclade marked by "x." Size of bars and points scaled by number of nodes in subclade, which is one less than number of species. Number of species in parentheses alongside subclade names and excludes those with uncertain placement ($n = 2$) and nonmonophyletic speedwell hebes ($n = 3$).

studies that have compared supercontigs to exons and introns, i.e., *Nepenthes* with Angiosperms353 (Murphy et al., 2020) and *Burmeistera* with a custom target capture data set (Bagley et al., 2020). A contrasting result of Villaverde et al. (2018) can be attributed to poor enrichment success of introns. The improved support provided by supercontigs in multiple cases suggests that, in addition to providing more phylogenetic information, supercontigs may compensate for the weaknesses of either exons or introns alone. Exons typically have lower evolutionary rates relative to flanking introns (Bagley et al., 2020), which likely explains why introns outperformed exons alone in shallow and deep nodes in our data set. Introns, however, may exhibit increased noise due to higher variability and potentially higher missing data (Murphy et al., 2020). Concurrently, introns have higher evolutionary rates and so may better resolve shallower parts of the tree than exons, which are under stronger selective constraints and may be more adequate to resolve older divergences (Bagley et al., 2020). While conflict between exons and introns is possible (and present in our data set), this conflict did not decrease the support provided by supercontigs relative to the two subsets alone.

While magnifying phylogenetic signal and improving some BS values, phylogenomic data sets may reveal complexity (e.g., gene tree discordance) that challenges attempts to resolve rapid radiations. This issue was apparent in our data set, where we found improved BS for subclades supported by past studies (Wagstaff and Garnock-Jones, 1998; Albach and Meudt, 2010; Meudt et al., 2015) and for shallow species relationships, but also incongruence between gene subsets for the backbone of our *Veronica* sect. *Hebe* phylogeny.

Furthermore, gCF and sCF values below 5–10% were common even for branches with high BS. As a statistical test of sampling variance, BS tends to increase with sample size, i.e., number of loci, which could explain its improvement with target capture simply because it generated magnitudes more data than traditional marker studies (Minh et al., 2020a). Concordance factors are thus valuable as a complementary measure of support that is independent of sample size above a certain threshold. Additional complementary metrics quantifying incongruence, such as internode certainty (Zhou et al., 2020), can be considered in future studies.

radiation of *Veronica* sect. *Hebe* as compared with traditional marker genes. A major benefit of the target sequence capture was the recovery of both targeted coding sequences and off-target noncoding sequences, which together provided the best phylogenetic support. However, we also found widespread gene tree and site discordance despite high bootstrap values. This discrepancy is common when reconstructing phylogenies of rapid radiations with target capture data due to processes like ILS, introgression, and stochastic error (Murphy et al., 2020; Shee et al., 2020; Stubbs et al., 2020). Our results now suggest that these conflicts cannot be readily resolved by filtering phylogenomic data sets. Improved development and application of discordance-aware methods will provide a way to resolve these conflicts and deepen our understanding of rapid radiations.

Suitability of supercontigs for phylogenetic reconstruction

The combined signal of exons and introns outweighed noise for improving overall BS. This outcome was also found in the few
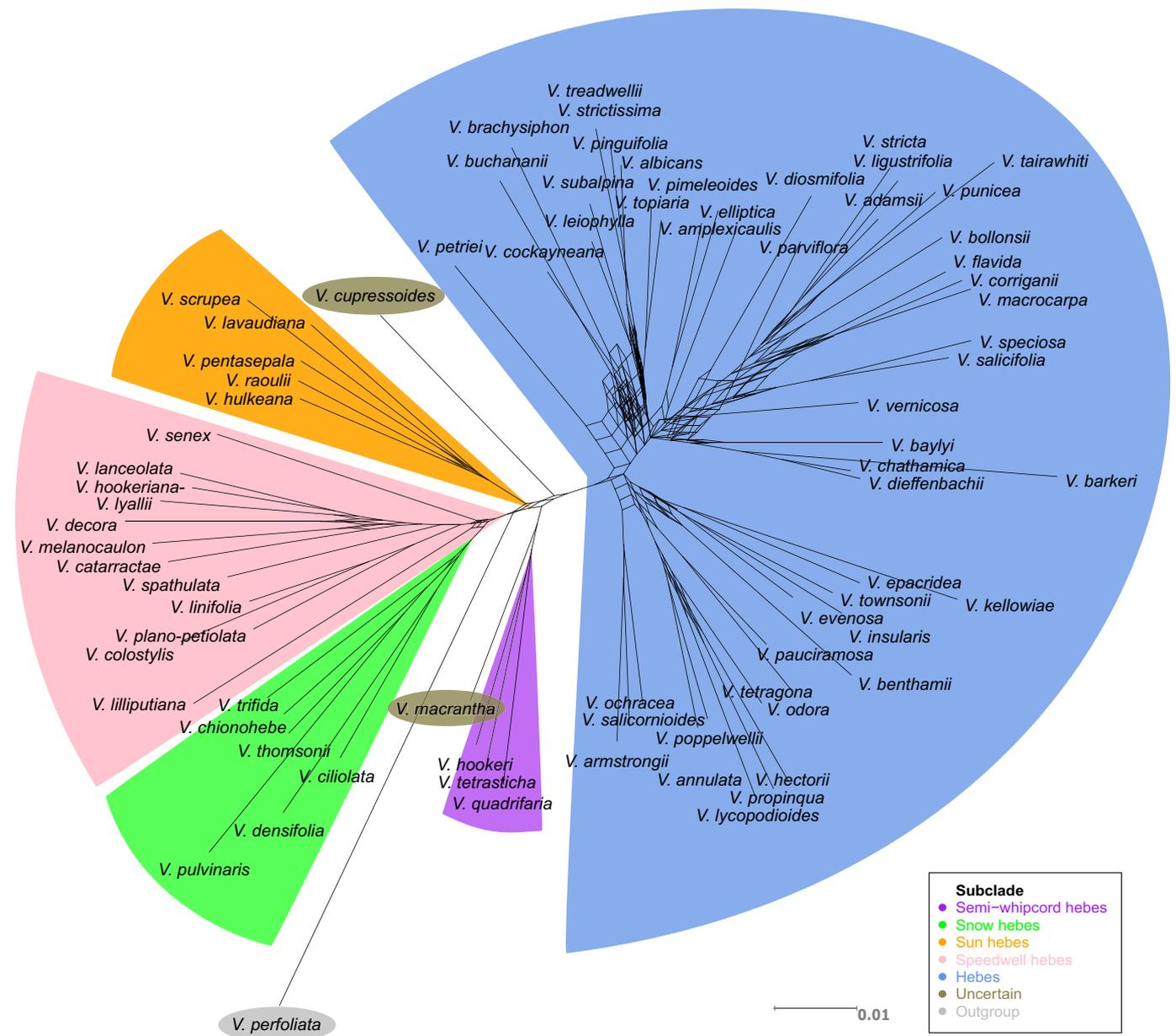
FIGURE 7. ConsensusNetwork of phylogenetic splits preserves subclades but shows conflict within core hebes. ConsensusNetwork calculated from splits present in 10% or more of IQ-TREE bootstrap trees from 229 supercontigs with SplitsTree4. Subclades color-coded.

Beyond discordance: opportunities for improving the application of target sequence capture to rapid radiations

Our finding of widespread discordance despite high BS was consistent with other phylogenomic studies of rapid radiations (Bagley et al., 2020; Murphy et al., 2020; Shee et al., 2020; Stubbs et al., 2020). Shorter branch lengths and low support values deeper in the tree, as seen in our data set, are also common in rapid radiations with short basal divergences that are more susceptible to homoplasy and ILS (Townsend et al., 2012; Bagley et al., 2020). Long terminal branches, a pattern also seen in the *Nepenthes* radiation sequenced with Angiosperms353 baits (Murphy et al., 2020), may further obscure signal with non-informative variation (Townsend et al., 2012). These factors, together with differences between coding and

noncoding regions or other sources of noise, may have also contributed to incongruence between gene subsets in the backbone of the tree. The lack of improvement in concordance factors for trees estimated with the MSC method ASTRAL-III and the low resolution of the tree estimated with SVDquartets also highlight the high level of discordance present in this data set (although missing data for some taxa could affect SVDquartets inference; see Nute et al., 2018). The ConsensusNetwork analysis visually demonstrates the conflict, especially in the core hebe subclade (Fig. 7). Notably, however, gene trees did not reveal strongly supported competing topologies, as shown by high gDFP values and neutral sCF values for IQ-TREE. This lack of strong signal among the genes and sites in some places in the trees may explain why support values were not
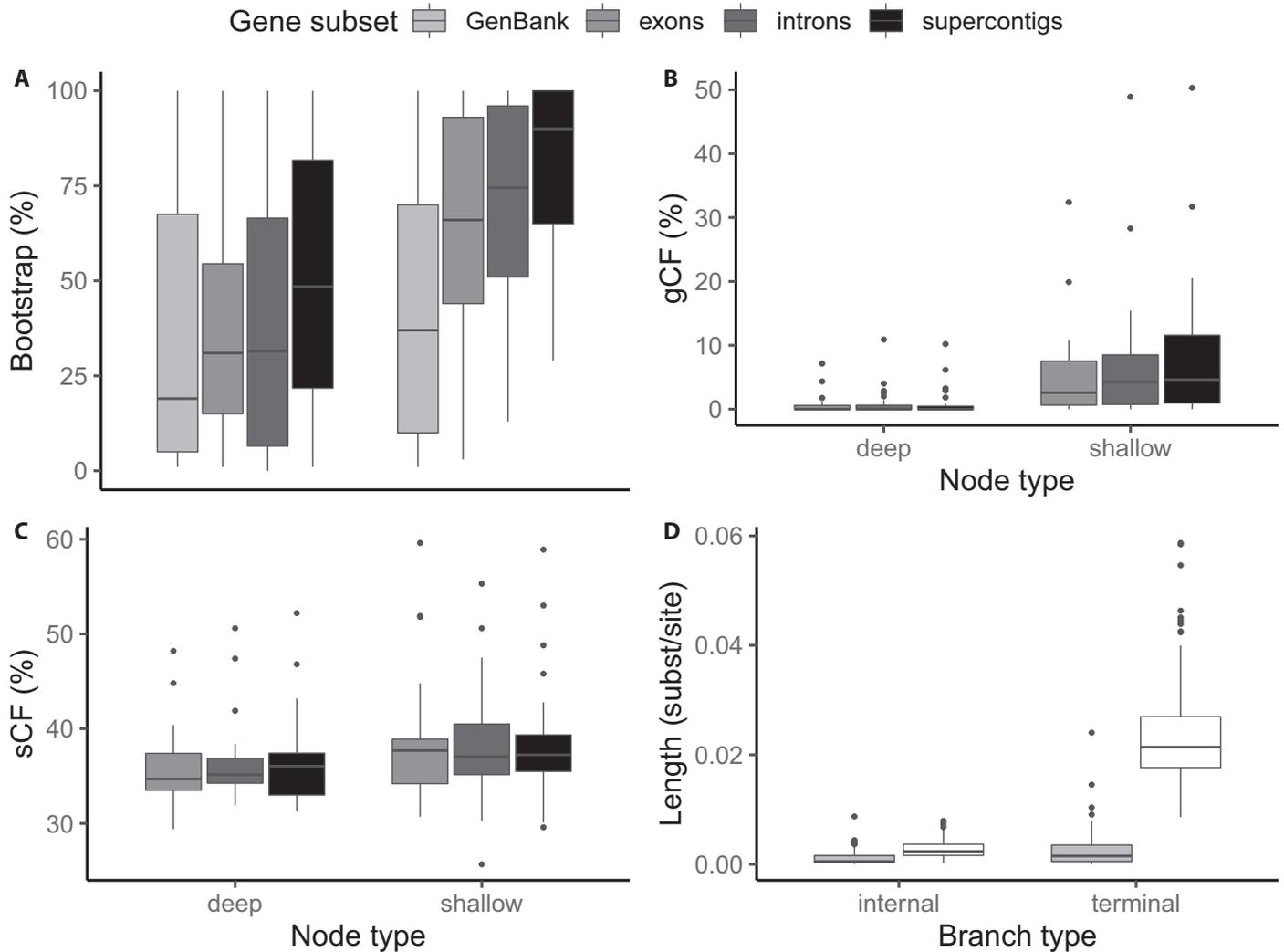
FIGURE 8. Target capture markers improved resolution for shallow, recent divergences in *Veronica* subgen. *Pseudoveronica* sect. *Hebe*. For parents of terminal nodes and deeper nodes, boxplots show (A) BS, (B) gCF, and (C) sCF values for different gene subsets under the full filtering scheme. gCF and sCF were not calculated for the GenBank data set. We also considered (D) branch lengths for internal and terminal nodes in GenBank and supercontig phylogenies. Boxplot symbols as in Fig. 2.

improved by filtering based on gene tree-derived concordance metrics. Rather, discordance was so widespread in our data set that it could not be mitigated by removing specific genes. Median gene tree bipartition support values were so close to 0 that most of the genes removed likely contributed virtually nothing to phylogenetic signal, even strong conflicting signal. The reasons for this kind of discordance could include both evolutionary processes–such as ILS, hybridization, inconsistent paralog retention by polyploids, and introgression–and nonphylogenetic artifacts such as alignment error resulting from inherent variation (Townsend et al., 2012; Minh et al., 2020a).

Methodological improvements could help to address discordance in rapid radiations. One explanation for the low resolution found by the summary MSC methods we applied is that they may be biased by poorly resolved gene trees (Gatesy and Springer, 2014; but see Jiang et al., 2020). Bayesian MSC methods that coestimate gene and species trees can outperform two-step methods such as ASTRAL, particularly for highly discordant data sets (Heled and Drummond, 2010), but are limited by the computational resources required for large data sets like ours. However, these MSC methods only account for ILS and not other processes like gene duplication and hybridization that also generate discordance in rapid radiations. New MSC methods are being developed that account for gene duplication and loss, e.g., ASTRAL-Pro (Zhang et al., 2020a). Evolutionary network models are a promising alternative for resolving rapid radiations with hybridization, as they incorporate gene flow into the MSC model, but, again, remain limited by model complexity and high computational costs (Blair and Ané, 2020; Jiang et al., 2020). More complete taxon sampling may be a straightforward way to resolve incongruence in the presence of gene tree discordance (Hedtke et al., 2006; Gardner et al., 2020).

Better methods for addressing polyploidy, which is commonly associated with plant radiations (Seehausen, 2004; Soltis et al., 2009), will also help to reduce phylogenetic discordance. In *Veronica* sect. *Hebe*, the hypothesized ancestral ploidy level is 6*x* relative to diploid *Veronica* outgroups, and 34 of the 80 core hebe species have

ploidy levels between 12*x* and 18*x* (Albach et al., 2008). Bait hybridization and assembly methods that produce chimeric sequences or inconsistently represent haplotypes from different genome duplications can therefore contribute to discordance and phylogenetic noise in these cases if paralogs are undetected (Nicholls et al., 2015). HybPiper assembled and flagged relatively few paralogs for our data set. Allele phasing with the available pipelines (i.e., WhatsHap) also did not prove effective for identifying paralogs hidden by chimeric sequences. This pipeline was designed for diploids (Kates et al., 2018), limiting how much heterozygosity or paralogy we could examine in our polyploid species. Improved methods for detecting hidden paralogs in polyploid phylogenomic data sets (e.g., Freyman et al., 2020; Nauheimer et al., 2021) will thus be crucial for disentangling discordance. It is also possible that genome duplication events in *Veronica* sect. *Hebe* are based on autopolyploid or allopolyploid events that were recent enough that the resulting gene copies have not diverged enough for assembly and mapping methods to detect differences, as found in other sections of *Veronica* (Padilla-García et al., 2018). Alternatively, the copies for these genes could have been quickly lost from the genome, possibly reflecting Angiosperms353 design, which targeted mostly single-copy genes in the plants used to design the bait set (Johnson et al., 2018). Nevertheless, the role of polyploidy in this data set remains unclear and requires further study. More applicable methods for haplotype phasing in polyploid genomes are now in development, with entire genomes successfully phased for some model organisms (Zhang et al., 2020b) and flexible tree-based methods becoming available for assigning phylogenomic loci to subgenomes (e.g., Freyman et al., 2020). As longer reads and higher sequencing depth become more accessible, polyploid haplotype phasing may further disentangle phylogenetic complexity and help reveal the role of polyploidy in the diversification of ecologically diverse radiations such as *Veronica* sect. *Hebe* (Soltis et al., 2009).

CONCLUSIONS

Phylogenomic target capture data sets deliver valuable insight into the complex evolutionary history of rapid radiations but can be difficult to analyze with existing computational approaches. Our test of universal target capture offered by Angiosperms353 for *Veronica* sect. *Hebe* was no exception. We found increased BS for many clades, further improved by including both coding and noncoding sequences, despite widespread gene tree discordance and some topological incongruence between data sets. The resulting data set therefore offers an important advance toward reconstructing the phylogenetic history of *Veronica* sect. *Hebe* and enabling its extensive ecological diversity to inform broader evolutionary questions. Future work in *Veronica* sect. *Hebe* will build on this advance, including increased taxon sampling and further examination of paralogs, polyploidy, and diversification. Our results also stress that the development and evaluation of appropriate phylogenetic methods has hardly tracked the increased availability of sequencing data, although advances are occurring. Appropriate methods for disentangling discordance, such as gene-tree based measures of concordance, examination of paralogs (e.g., Freyman et al., 2020), and modeling coalescence, should continue to be explored to use target capture efficiently for resolving the phylogenetic history of rapid radiations. Nevertheless, target capture can be a highly cost-effective and tractable method for studying rapid radiations. Although target capture

loci are not immune to phylogenetic complexity, they constitute a valuable foothold in the process of understanding this complexity.

AUTHOR CONTRIBUTIONS

A.T. obtained funding, performed lab work and bioinformatic analyses, and wrote the manuscript. J.I. provided conceptual input, supervised bioinformatic analyses, and edited the manuscript. H.M.M. supervised curation of herbarium specimens, organized and provided samples, and provided conceptual input. D.C.A. provided samples and conceptual input. W.G.L. provided conceptual input. A.J.T. obtained funding, provided conceptual input, supervised analyses, and edited the manuscript.

DATA AVAILABILITY STATEMENT

Illumina sequencing reads have been deposited in the NCBI Sequence Read Archive (submission pending https://submit.ncbi.nlm.nih.gov/subs/sra/SUB9172517/). Scripts are available at https://github.com/annethomas/veronica_phylo.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. Supercontigs provide marginally higher ASTRAL III posterior values without improving concordance.

APPENDIX S2. ASTRAL III phylogeny for *Veronica* subgen. *Pseudoveronica* sect. *Hebe*.

APPENDIX S3. SVDquartets phylogeny for *Veronica* subgen. *Pseudoveronica* sect. *Hebe*.

APPENDIX S4. Removal of retained genes with low rates (<10% of samples) of paralog warnings does not improve support compared to full IQ-TREE supercontig data set.

APPENDIX S5. Gene discordance factor due to paraphyly (gDFP) much higher than gCF or two nearest neighbor interchange topologies under full filtering scheme with IQ-TREE.

APPENDIX S6. *Veronica* subclades have similar overall concordance factors.

APPENDIX S7. IQ-TREE phylogeny for *Veronica* subgen. *Pseudoveronica* sect. *Hebe* estimated from 173 exon-only sequences.

APPENDIX S8. IQ-TREE phylogeny for *Veronica* subgen. *Pseudoveronica* sect. *Hebe* estimated from 167 intron-only sequences.

APPENDIX S9. IQ-TREE phylogeny for *Veronica* subgen. *Pseudoveronica* sect. *Hebe* estimated from ITS, *trnL-trnL-trnF*, *rps16*, *rpoB-trnC*, and *CYCLOIDEA2*.

LITERATURE CITED

Abe, F. R., and B. S. Lieberman. 2009. The nature of evolutionary radiations: a case study involving Devonian trilobites. *Evolutionary Biology* 36: 225–234.

Albach, D. C., M. M. Martínez-Ortega, L. Delgado, H. Weiss-Schneeweiss, F. Özgökce, and M. A. Fischer. 2008. Chromosome numbers in Veroniceae (Plantaginaceae): review and several new counts. *Annals of the Missouri Botanical Garden* 95: 543–566.

Albach, D. C., and H. M. Meudt. 2010. Phylogeny of *Veronica* in the Southern and Northern Hemispheres based on plastid, nuclear ribosomal and nuclear low-copy DNA. *Molecular Phylogenetics and Evolution* 54: 457–471.

Bagley, J. C., S. Uribe-Convers, M. M. Carlsen, and N. Muchhala. 2020. Utility of targeted sequence capture for phylogenomics in rapid, recent angiosperm radiations: Neotropical *Burmeistera* bellflowers as a case study. *Molecular Phylogenetics and Evolution* 152: 106769.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477.

Bayly, M., and A. Kellow. 2006. An illustrated guide to New Zealand hebes. Te Papa Press, Wellington, New Zealand.

Blair, C., and C. Ané. 2020. Phylogenetic trees and networks can serve as powerful and complementary approaches for analysis of genomic data. *Systematic Biology* 69: 593–601.

Brewer, G. E., J. J. Clarkson, O. Maurin, A. R. Zuntini, V. Barber, S. Bellot, N. Biggs, et al. 2019. Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Frontiers in Plant Science* 10: 1102.

Capella-Gutierrez, S., J. M. Silla-Martinez, and T. Gabaldon. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973.

Chernomor, O., A. von Haeseler, and B. Q. Minh. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology* 65: 997–1008.

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

Czech, L., J. Huerta-Cepas, and A. Stamatakis. 2017. A critical review on the use of support values in tree viewers and bioinformatics toolkits. *Molecular Biology and Evolution* 34: 1535–1542.

Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24: 332–340.

Drummond, C. S., R. J. Eastwood, S. T. S. Miotto, and C. E. Hughes. 2012. Multiple continental radiations and correlates of diversification in *Lupinus* (leguminosae): testing for key innovation with incomplete taxon sampling. *Systematic Biology* 61: 443–460.

Fernández-Mazuecos, M., G. Mellers, B. Vigalondo, L. Sáez, P. Vargas, and B. J. Glover. 2018. Resolving recent plant radiations: Power and robustness of genotyping-by-sequencing. *Systematic Biology* 67: 250–268.

Freyman, W. A., M. G. Johnson, and C. J. Rothfels. 2020. homologizer: Phylogenetic phasing of gene copies into polyploid subgenomes. *bioRxiv*. https://doi.org/10.1101/2020.10.22.351486

Gardner, E. M., M. G. Johnson, J. T. Pereira, A. S. A. Puad, D. Arifiani, Sahromi, S., N. J. Wickett, and N. J. C. Zerega. 2020. Paralogs and off-target sequences improve phylogenetic resolution in a densely-sampled study of the breadfruit genus (*Artocarpus*, Moraceae). *Systematic Biology* 70: 558–575.

Garnock-Jones, P. J. 1993. Phylogeny of the *Hebe* complex (Scrophulariaceae: Veroniceae). *Australian Systematic Botany* 6: 457–479.

Garnock-Jones, P. J., and D. G. Lloyd. 2004. A taxonomic revision of *Parahebe* (Plantaginaceae) in New Zealand. *New Zealand Journal of Botany* 42: 181–232.

Gatesy, J., and M. S. Springer. 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular Phylogenetics and Evolution* 80: 231–266.

Gernandt, D. S., X. A. Dugua, A. Vázquez-Lobo, A. Willyard, A. M. Letelier, J. A. P. de la Rosa, D. Piñero, and A. Liston. 2018. Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection *Australes*. *American Journal of Botany* 105: 711–725.

Hale, H., E. M. Gardner, J. Viruel, L. Pokorny, and M. G. Johnson. 2020. Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. *Applications in Plant Sciences* 8: e11337.

Hart, M. L., L. L. Forrest, J. A. Nicholls, and C. A. Kidner. 2016. Retrieval of hundreds of nuclear loci from herbarium specimens. *Taxon* 65: 1081–1092.

Hedtke, S. M., T. M. Townsend, and D. M. Hillis. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Systematic Biology* 55: 522–529.

Heibl, C. 2008. PHYLOCH: R language tree plotting tools and interfaces to diverse phylogenetic software packages. Available at: http://www.christophh eibl.de/Rpackages.html.

Heled, J., and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* 27: 570–580.

Herrando-Moraira, S., J. A. Calleja, P. Carnicero, K. Fujikawa, M. Galbany-Casals, N. Garcia-Jacas, H.-T. Im, et al. 2018. Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe Cardueae (Compositae). *Molecular Phylogenetics and Evolution* 128: 69–87.

Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35: 518–522.

Hughes, C. E., and G. W. Atchison. 2015. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytologist* 207: 275–282.

Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.

Huson, D. H., and C. Scornavacca. 2012. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Systematic Biology* 61: 1061–1067.

Jiang, X., S. V. Edwards, and L. Liu. 2020. The multispecies coalescent model outperforms concatenation across diverse phylogenomic data sets. *Systematic Biology* 69: 795–812.

Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. C. Zerega, and N. J. Wickett. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4: 1600016.

Johnson, M. G., L. Pokorny, S. Dodsworth, L. R. Botigué, R. S. Cowan, A. Devault, W. L. Eiserhardt, et al. 2018. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* 68: 594–606.

Jones, K. E., T. Fér, R. E. Schmickl, R. B. Dikow, V. A. Funk, S. Herrando-Moraira, P. R. Johnston, et al. 2019. An empirical assessment of a single family-wide hybrid capture locus set at multiple evolutionary timescales in Asteraceae. *Applications in Plant Sciences* 7: e11295.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–589.

Kassambara, A. 2020. ggpubr: 'ggplot2' based publication ready plots. Website: https://CRAN.R-project.org/package=ggpubr.

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.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

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.

Larridon, I., T. Villaverde, A. R. Zuntini, L. Pokorny, G. E. Brewer, N. Epitawalage, I. Fairlie, et al. 2020. Tackling rapid radiations with targeted sequencing. *Frontiers in Plant Science* 10: 1655.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio] [Preprint].

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.

Mayland-Quellhorst, E., H. M. Meudt, and D. C. Albach. 2016. Transcriptomic resources and marker validation for diploid and polyploid *Veronica* (Plantaginaceae) from New Zealand and Europe. *Applications in Plant Sciences* 4: 1600091.

McKain, M. R., M. G. Johnson, S. Uribe-Convers, D. Eaton, and Y. Yang. 2018. Practical considerations for plant phylogenomics. *Applications in Plant Sciences* 6: e1038.

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.

Mclean, B. S., K. C. Bell, J. M. Allen, K. M. Helgen, and J. A. Cook. 2019. Impacts of inference method and data set filtering on phylogenomic resolution in a rapid radiation of ground squirrels (Xerinae: Marmotini). *Systematic Biology* 68: 298–316.

Meudt, H. M. 2008. Taxonomic revision of Australasian snow hebes (*Veronica*, Plantaginaceae). *Australian Systematic Botany* 21: 387–421.

Meudt, H. M., and M. J. Bayly. 2008. Phylogeographic patterns in the Australasian genus *Chionohebe* (*Veronica* s.l., Plantaginaceae) based on AFLP and chloroplast DNA sequences. *Molecular Phylogenetics and Evolution* 47: 319–338.

Meudt, H. M., B. M. Rojas-Andrés, J. M. Prebble, E. Low, P. J. Garnock-Jones, and D. C. Albach. 2015. Is genome downsizing associated with diversification in polyploid lineages of *Veronica*? *Botanical Journal of the Linnean Society* 178: 243–266.

Meyer, M., and M. Kircher. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* 2010: pdb.prot5448.

Minh, B. Q., M. W. Hahn, and R. Lanfear. 2020a. New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution* 37: 2727–2733.

Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, and R. Lanfear. 2020b. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37: 1530–1534.

Morales-Briones, D. F., A. Liston, and D. C. Tank. 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytologist* 218: 1668–1684.

Murphy, B., F. Forest, T. Barraclough, J. Rosindell, S. Bellot, R. Cowan, M. Golos, et al. 2020. A phylogenomic analysis of *Nepenthes* (Nepenthaceae). *Molecular Phylogenetics and Evolution* 144: 106668.

Nauheimer, L., N. Weigner, E. Joyce, D. Crayn, C. Clarke, and K. Nargar. 2021. HybPhaser: a workflow for the detection and phasing of hybrids in target capture datasets. *Applications in Plant Sciences* 9 (in press).

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.

Nicholls, J. A., R. T. Pennington, E. J. Koenen, C. E. Hughes, J. Hearn, L. Bunnefeld, K. G. Dexter, et al. 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Science* 6: 710.

Nute, M., J. Chou, E. K. Molloy, and T. Warnow. 2018. The performance of coalescent-based species tree estimation methods under models of missing data. *BMC Genomics* 19: 286.

Padilla-García, N., B. M. Rojas-Andrés, N. López-González, M. Castro, S. Castro, J. Loureiro, D. C. Albach, et al. 2018. The challenge of species delimitation in the diploid–polyploid complex *Veronica* subsection *Pentasepalae*. *Molecular Phylogenetics and Evolution* 119: 196–209.

Paradis, E., and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–528.

Patterson, M., T. Marschall, and N. Pisanti. 2015. Whatshap: weighted haplotype assembly for future-generation sequencing reads. *Journal of Computational Biology* 22: 498–509.

Philippe, H., H. Brinkmann, D. V. Lavrov, D. T. J. Littlewood, M. Manuel, G. Wörheide, and D. Baurain. 2011. Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biology* 9: e1000602.

Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS One* 5: e9490.

Ranwez, V., S. Harispe, F. Delsuc, and E. J. P. Douzery. 2011. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One* 6: e22594.

Seehausen, O. 2004. Hybridization and adaptive radiation. *Trends in Ecology & Evolution* 19: 198–207.

Shee, Z. Q., D. G. Frodin, R. Cámara-Leret, and L. Pokorny. 2020. Reconstructing the complex evolutionary history of the Papuasian S*chefflera* radiation through herbariomics. *Frontiers in Plant Science* 11: 258.

Siniscalchi, C. M., O. Hidalgo, L. Palazzesi, J. Pellicer, L. Pokorny, O. Maurin, I. J. Leitch, et al. 2021. Lineage-specific vs. universal: a comparison of the Compositae1061 and Angiosperms353 enrichment panels in the sunflower family. *Applications in Plant Sciences* 9 (in press). https://doi.org/10.1002/aps3.11419

Slater, G. S. C., and E. Birney. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6: 31.

Smith, M. R. 2020. Information theoretic Generalized Robinson-Foulds metrics for comparing phylogenetic trees. *Bioinformatics* 36: 5007–5013.

Smith, S. A., J. W. Brown, and J. F. Walker. 2018. So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *PLoS One* 13: e0197433.

Smith, S. A., and J. F. Walker. 2019. PyPHLAWD: a python tool for phylogenetic dataset construction. *Methods in Ecology and Evolution* 10: 104–108.

Soltis, D. E., V. A. Albert, J. Leebens-Mack, C. D. Bell, A. H. Paterson, C. Zheng, D. Sankoff, et al. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336–348.

Stubbs, R. L., R. A. Folk, C.-L. Xiang, S. Chen, D. E. Soltis, and N. Cellinese. 2020. A phylogenomic perspective on evolution and discordance in the alpine-arctic plant clade *Micranthes* (Saxifragaceae). *Frontiers in Plant Science* 10: 1773.

Suyama, M., D. Torrents, and P. Bork. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* 34: W609–W612.

Townsend, J. P., Z. Su, and Y. I. Tekle. 2012. Phylogenetic signal and noise: Predicting the power of a data set to resolve phylogeny. *Systematic Biology* 61: 835–835.

Villaverde, T., L. Pokorny, S. Olsson, M. Rincón-Barrado, M. G. Johnson, E. M. Gardner, N. J. Wickett, et al. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist* 220: 636–650.

Wagstaff, S. J., and P. J. Garnock-Jones. 1998. Evolution and biogeography of the *Hebe* complex (Scrophulariaceae) inferred from ITS sequences. *New Zealand Journal of Botany* 36: 425–437.

Weitemier, K., S. C. K. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: 1400042.

Wickham, H. 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag, NY, NY, USA.

Yang, Y., and S. A. Smith. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution* 31: 3081–3092.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 153.

Zhang, C., C. Scornavacca, E. K. Molloy, and S. Mirarab. 2020a. ASTRAL-Pro: Quartet-based species-tree inference despite paralogy. *Molecular Biology and Evolution* 37: 3292–3307.

Zhang, X., R. Wu, Y. Wang, J. Yu, and H. Tang. 2020b. Unzipping haplotypes in diploid and polyploid genomes. *Computational and Structural Biotechnology Journal* 18: 66–72.

Zhou, X., S. Lutteropp, L. Czech, A. Stamatakis, M. V. Looz, and A. Rokas. 2020. Quartet-based computations of internode certainty provide robust measures of phylogenetic incongruence. *Systematic Biology* 69: 308–324.

APPENDIX 1. GenBank accession numbers for five markers from previous studies and specimen vouchers for Angiosperms353 samples used in this study.

Species			GenBank					Angiosperms353 sequencing		
Species	Subclade	Chromosome no. (*n* = 20/21)	CYC2	ITS	rpoB-trnC	rps16	trnL-trnL-trnF	Voucher	Source	Year collected
Veronica_adamsii	hebes	80	NA	KJ630622	NA	NA	KJ630723	WELT SP103455	silica	2013
Veronica_albicans	hebes	40/80[1]	NA	AF037373	NA	NA	KJ630724	WELT SP103942	silica	2015
Veronica_amplexicaulis	hebes	40	NA	KJ630624	NA	NA	KJ630725	WELT SP103965	silica	2015
Veronica_barkeri	hebes	40/80	NA	AF037374	NA	NA	NA	WELT SP080537	herbarium	1997
Veronica_baylyi	hebes	116	NA	KJ630626	NA	NA	KJ630727	WELT SP103983	silica	2015
Veronica_benthamii	hebes	40	NA	AF229041	NA	NA	NA	WELT SP102778	herbarium	2013
Veronica_bollonsii	hebes	40	NA	KJ630627	NA	NA	KJ630728	WELT SP103972	silica	2015
Veronica_brachysiphon	hebes	120	NA	KJ630628	NA	NA	KJ630729	WELT SP103452	silica	2013
Veronica_buchananii	hebes	40/80	NA	KJ630629	NA	NA	KJ630730	OLD 00021	silica	2013
Veronica_chathamica	hebes	40	NA	AF037387	NA	NA	NA	WELT SP103984	silica	2015
Veronica_cockayneana	hebes	120	NA	AF037399	NA	NA	NA	WELT SP103994	silica	2015
Veronica_corriganii	hebes	80	NA	AF037384	NA	NA	KJ630733	WELT SP103453	silica	2013
Veronica_dieffenbachii	hebes	40	NA	AY034852	NA	NA	KJ630736	WELT SP080545	herbarium	1998
Veronica_diosmifolia	hebes	40/80	NA	KJ630636	NA	NA	KJ630737	WELT SP103969	silica	2015
Veronica_elliptica	hebes	40	NA	AF037392	NA	FJ848242	AY540883	OLD 00039	silica	2013
Veronica_epacridea	hebes	42	NA	AF037389	NA	NA	NA	WELT SP103988	silica	2015
Veronica_evenosa	hebes	120	NA	KJ630638	NA	NA	KJ630739	WELT SP103456	silica	2013
Veronica_flavida	hebes	40	NA	KJ630639	NA	NA	KJ630740	WELT SP103974	silica	2015
Veronica_insularis	hebes	40	NA	EF635486	NA	NA	AF486406	WELT SP083581	herbarium	2001
Veronica_kellowiae	hebes	42	NA	AY034856	NA	NA	NA	WELT SP083615	herbarium	2003
Veronica_leiophylla	hebes	80	NA	KJ630646	NA	NA	KJ630747	OLD 00127	silica	2013
Veronica_ligustrifolia	hebes	40	NA	KJ630647	NA	NA	KJ630748	WELT SP103962	silica	2015
Veronica_macrantha	hebes	42	FJ848315	AY034853	FJ848197	FJ848244	FJ848048	WELT SP103475	silica	2013
Veronica_macrocarpa	hebes	80/120	NA	KJ630651	NA	NA	KJ630752	WELT SP103976	silica	2015
Veronica_odora	hebes	42/84	FJ848317	AF037388	NA	FJ848245	AY540882	WELT SP103823	silica	2014
Veronica_parviflora	hebes	80	NA	AY034854	NA	NA	NA	WELT SP103975	silica	2015
Veronica_pauciramosa	hebes	42	NA	AF069466	NA	NA	NA	WELT SP103948	silica	2015
Veronica_petriei	hebes	42	NA	AF229042	NA	NA	NA	WELT SP083580	herbarium	2001
Veronica_pimeleoides	hebes	40/80	NA	AY034855	NA	NA	KJ630755	OLD 00128	silica	2013
Veronica_pinguifolia	hebes	40/80	NA	KJ630655	NA	NA	KJ630756	WELT SP103950	silica	2015
Veronica_punicea	hebes	118	NA	KJ630656	NA	NA	KJ630758	WELT SP103454	silica	2013
Veronica_salicifolia	hebes	40	FJ848320	AF037386	FJ848199	FJ848248	FJ848049	WELT SP080808	herbarium	1998
Veronica_speciosa	hebes	40	NA	KJ630660	NA	NA	KJ630762	OLD 00050	silica	2013
Veronica_stricta	hebes	40/80	NA	KJ630662	NA	NA	KJ630764	WELT SP103967	silica	2015
Veronica_strictissima	hebes	80	NA	KJ630663	NA	NA	KJ630765	OLD 00038	silica	2013
Veronica_subalpina	hebes	80	NA	KJ630664	NA	NA	KJ630766	WELT SP103939	silica	2015
Veronica_tairawhiti	hebes	80	NA	KJ630665	NA	NA	KJ630767	WELT SP088197	herbarium	2009
Veronica_topiaria	hebes	122	NA	KJ630666	NA	NA	KJ630768	OLD 00028	silica	2013
Veronica_townsonii	hebes	40	NA	AY034857	NA	NA	KJ630769	WELT SP103954	silica	2015
Veronica_treadwellii	hebes	40	NA	KJ630668	NA	NA	KJ630770	WELT SP103457	silica	2013
Veronica_vernicosa	hebes	42	NA	AY034858	NA	NA	KJ630771	WELT SP103477	silica	2013
Veronica_annulata	hebes (whipcord)	42	NA	AF069464	NA	NA	NA	WELT SP104001	silica	2015
Veronica_armstrongii	hebes (whipcord)	84	NA	AF069463	NA	NA	KJ630726	WELT SP103459	silica	2013
Veronica_hectorii	hebes (whipcord)	40	NA	AF069461	NA	NA	KJ630741	WELT SP103820	silica	2014

APPENDIX 1. (Continued)

Species		Chromosome no. (n = 20/21)	GenBank					Angiosperms353 sequencing		
Species	Subclade		CYC2	ITS	rpoB-trnC	rps16	trnL-trnL-trnF	Voucher	Source	Year collected
Veronica_lycopodioides	hebes (whipcord)	40	NA	AF069456	NA	NA	KJ630750	WELT SP103944	silica	2015
Veronica_ochracea	hebes (whipcord)	124	NA	AF069462	NA	NA	KJ630754	WELT SP103461	silica	2013
Veronica_poppelwellii	hebes (whipcord)	40	NA	AF069454	NA	NA	NA	WELT SP103951	silica	2015
Veronica_propinqua	hebes (whipcord)	40	NA	AF069458	NA	NA	NA	WELT SP103946	silica	2015
Veronica_salicornioides	hebes (whipcord)	42	FJ848318	AF069465	FJ848198	FJ848246	AY540879	OLD 00125	silica	2013
Veronica_tetragona	hebes (whipcord)	40	NA	AF069457	NA	NA	NA	WELT SP104003	silica	2015
Veronica_cupressoides	semi-whipcord hebes	42	FJ848322	AF037378	FJ848202	FJ848251	AY540880	WELT SP103949	silica	2015
Veronica_hookeri	semi-whipcord hebes	42	NA	AY034851	EU349506	FJ848250	FJ848050	WELT SP104004	silica	2015
Veronica_quadrifaria	semi-whipcord hebes	42	FJ848321	AF037377	FJ848201	FJ848249	AY540886	WELT SP104000	silica	2015
Veronica_tetrasticha	semi-whipcord hebes	42	FJ848323	NA	FJ848203	FJ848252	FJ848051	WELT SP102875	silica	2014
Veronica_chionohebe	snow hebes	42	FJ848307	FJ848070	EU349548	FJ848233	FJ848044	WELT SP084028/A	silica	2005
Veronica_ciliolata	snow hebes	42	FJ848304	AF229036	EU349518	FJ848230	FJ848041	WELT SP084037	silica	2005
Veronica_densifolia	snow hebes	42	FJ848305	AF037375	EU349531	FJ848232	FJ848042	WELT SP102867	silica	2014
Veronica_pulvinaris	snow hebes	42	FJ848309	AF229038	EU349560	FJ848235	FJ848046	WELT SP103902/A	silica	2014
Veronica_thomsonii	snow hebes	42	FJ848310	AF229039	EU349578	FJ848234	FJ848045	WELT SP102858	silica	2014
Veronica_trifida	snow hebes	42	FJ848335	AF037376	FJ848210	NA	NA	WELT SP102861	silica	2014
Veronica_catarractae	speedwell hebes	42	FJ848324	AY034859	FJ848204	FJ848253	KJ630731	OLD 00052	silica	2013
Veronica_colostylis	speedwell hebes	42	NA	AF229045	NA	NA	KJ630732	WELT SP103966	silica	2015
Veronica_decora	speedwell hebes	40	FJ848325	AF229047	FJ848205	FJ848255	AY540877	WELT SP103987	silica	2015
Veronica_hookeriana	speedwell hebes	42	FJ848326	KJ630641	EU349511	FJ848256	KJ630742	WELT SP090421	silica	2008
Veronica_lanceolata	speedwell hebes	42	FJ848327	KJ630644	FJ848206	FJ848257	FJ848055	WELT SP103920	silica	2015
Veronica_lilliputiana	speedwell hebes	42	NA	AF037394	NA	NA	FJ848052	WELT SP103989	silica	2015
Veronica_linifolia	speedwell hebes	42	FJ848328	AF229048	FJ848207	FJ848258	FJ848056	WELT SP090408	silica	2008
Veronica_lyallii	speedwell hebes	42	FJ848329	AF037395	EU349516	FJ848259	FJ848057	OLD 00069	silica	2013
Veronica_ melanocaulon	speedwell hebes	42	FJ848330	KJ630652	EU349509	FJ848260	FJ848058	WELT SP098804	herbarium	1997
Veronica_ planopetiolata	speedwell hebes	84	FJ848331	AF229050	EU349512	FJ848261	FJ848059	WELT SP091593	silica	2012
Veronica_senex	speedwell hebes	42	FJ848332	FJ848082	FJ848208	FJ848262	FJ848060	WELT SP086370	silica	1998
Veronica_spathulata	speedwell hebes	84	FJ848333	AY034861	EU349517	FJ848263	FJ848061	WELT SP101322	herbarium	1997
Veronica_hulkeana	sun hebes	42	FJ848312	AF037379	FJ848194	FJ848238	KJ630744	WELT SP103447	silica	2012
Veronica_lavaudiana	sun hebes	42	NA	KJ630645	NA	FJ848239	KJ630746	WELT SP101443	herbarium	2005
Veronica_pentasepala	sun hebes	42	FJ848313	FJ848076	FJ848195	FJ848240	NA	WELT SP090563	silica	2007
Veronica_raoulii	sun hebes	42	FJ848314	AF037380	FJ848196	FJ848241	AY540885	WELT SP103472	silica	2013
Veronica_scrupea	sun hebes	42	FJ848311	FJ848074	FJ848193	FJ848237	NA	WELT SP101324	silica	2007
Veronica_chamaedrys	outgroup (N. hemisphere)	16 (32)	FJ848296	DQ227329	FJ848187	AY218814	AY673632	CGE00032992	silica	2019
Veronica_perfoliata	outgroup (Australia)	40	NA	JX196844	FJ848192	FJ848228	FJ848040	WELT SP101339	herbarium	2009