The quest to resolve recent radiations: Plastid phylogenomics of extinct and endangered Hawaiian endemic mints (Lamiaceae)

Andreanna J. Welch[a,b,*], Katherine Collins[a], Aakrosh Ratan[c], Daniela I. Drautz-Moses[d], Stephan C. Schuster[d], Charlotte Lindqvist[a,*]

[a]*Department of Biological Sciences, University at Buffalo (SUNY), Buffalo, NY, 14260, USA*

[b]*Current affiliation: School of Biological and Biomedical Sciences, Durham University, South Road, Durham, DH1 3LE, UK*

[c]*Department of Public Health Sciences and Center for Public Health Genomics, University of Virginia, Charlottesville 22908, VA, USA*

[d]*Singapore Centre on Environmental Life Sciences Engineering, Nanyang Technological University, Singapore 637551*

*Corresponding authors:E-mail addresses: andreanna05@gmail.com (A.J. Welch); cl243@buffalo.edu (C. Lindqvist)

Abstract

The Hawaiian mints (Lamiaceae), one of the largest endemic plant lineages in the archipelago, provide an excellent system to study rapid diversification of a lineage with a remote, likely paleohybrid origin. Since their divergence from New World mints 4-5 million years ago the members of this lineage have diversified greatly and represent a remarkable array of vegetative and reproductive phenotypes. Today many members of this group are endangered or already extinct, and molecular phylogenetic work relies largely on herbarium samples collected during the last century. So far a gene-by-gene approach has been utilized, but the recent radiation of the Hawaiian mints has resulted in minimal sequence divergence and hence poor phylogenetic resolution. In our quest to trace the reticulate evolutionary history of the lineage, a resolved maternal phylogeny is necessary. We applied a high-throughput approach to sequence 12 complete or nearly complete plastid genomes from multiple Hawaiian mint species and relatives, including extinct and rare taxa. We also targeted 108 hypervariable regions from throughout the chloroplast genomes in nearly all of the remaining Hawaiian species, and relatives, using a next-generation amplicon sequencing approach. This procedure generated ~20Kb of sequence data for each taxon and considerably increased the total number of variable sites over previous analyses. Our results demonstrate the potential of high-throughput sequencing of historic material for evolutionary studies in rapidly evolving lineages. Our study, however, also highlights the challenges of resolving relationships within recent radiations even at the genomic level.

Keywords: Hawaii; Lamiaceae; phylogenomics; plastid genomes; radiation

1. Introduction

While molecular phylogenetic work over the last few decades has addressed relationships within and among major organismal groups, an on-going challenge is how to approach more recent evolutionary radiations. In particular, phylogenetic inference poses considerable limitations for resolving reticulate, recent relationships that may not exhibit a bifurcating, hierarchical evolutionary history. Remote island archipelagoes serve as a "natural laboratory" for studies of such recent evolutionary radiations. The isolation of these islands (both from the mainland and from one another), plus the wealth of different habitats available, provides ideal settings for rapid diversification and adaptation (Emerson, 2002). The Hawaiian archipelago, which lies greater than 3700 km from the nearest continental landmass, provides several famous examples of organismal radiations, including *Drosophila* (Kambysellis et al., 1995), spiders (Gillespie et al., 1994), birds (Lerner et al., 2011), and plants, such as the silversword alliance (Witter and Carr, 1988), and lobelioids (Givnish et al., 2013).

Another example is that of the endemic Hawaiian mints (Lamiaceae), one of the most species-rich plant lineages in the archipelago (Baldwin and Wagner, 2010). The ancestor of this monophyletic lineage arrived in the Hawaiian Islands 4-5 million years ago and rapidly diversified into three genera: *Haplostachys*, *Phyllostegia*, and *Stenogyne* (Roy et al., 2013). The 59 extant members of this lineage exhibit remarkable phenotypical and ecological diversity. For example, *Haplostachys* have fragrant white flowers associated with an insect-pollination syndrome and produce dry fruits, whereas both *Phyllostegia* and *Stenogyne* produce fleshy fruits, and *Stenogyne* have flowers associated with bird pollination. Species are often endemic to one or two islands and can be found in a variety of habitats, ranging from xeric shrubland to wet forests, and at a wide range of elevations (Lindqvist and Albert, 2002). Similar to many other island endemics, however, many Hawaiian mint species have been impacted by anthropogenic influences. At least 13 species are extinct, and the majority of the extant taxa are threatened or critically imperiled, with several species having fewer than 100 individuals remaining (Wagner et al. 1999). Taxonomic knowledge for the majority of the species in this lineage comes from collections in herbaria and botanical gardens, and these collections represent primary sources of material in phylogenetic analyses of their evolutionary history (Lindqvist and Albert, 2002; Lindqvist et al., 2003).

Recent molecular phylogenetic work in the subfamily Lamioideae has established the phylogenetic position of the Hawaiian lineage within tribe Stachydeae, a largely heterogeneous tribe spanning a wide array of morphological, cytological, and biogeographical diversity comprising of at least 12 genera, including the subcosmopolitan genus *Stachys* (Bendiksby et al., 2011; Lindqvist and Albert, 2002; Scheen et al., 2010). The Hawaiian genera are nested within a strongly supported lineage predominantly composed of New World *Stachys*. Intriguingly, molecular phylogenetic studies show that while chloroplast DNA (cpDNA) sequence data group the monophyletic Hawaiian mints with Meso-South American (SA) *Stachys* (Fig. 1a), nuclear ribosomal DNA sequences suggest that they are most closely related to temperate North American (NA) *Stachys* (Fig. 1b) (Lindqvist and Albert, 2002; Lindqvist et al., 2003; Roy et al., 2013). Interestingly, the Californian-Pacific northwestern species *Stachys chamissonis* groups is closely related to the Hawaiian clade in both cases. This cyto-nuclear phylogenetic incongruence is corroborated with low-copy nuclear loci that indicate a polyphyletic placement of the Hawaiian species, with some members grouping with temperate NA *Stachys* and others grouping with Meso-SA *Stachys* (Fig. 1c) (Roy et al., 2015). These results, in combination with karyotypic

and morphological data and multi-locus coalescence species tree reconstruction, convey a scenario for the Hawaiian mint lineage that is best explained by a hybrid origin involving at least two separate American *Stachys* lineages, descendants of which colonized the Hawaiian Islands (Lindqvist and Albert, 2002; Roy et al., 2013; Roy et al., 2015). A number of other Hawaiian angiosperm radiations of American origin appear to have a hybrid ancestry, which may have been an important factor in insular colonizations and diversification of species-rich lineages (Baldwin and Wagner, 2010).

Paradoxically, in contrast to their extensive morphological and ecological diversity, DNA sequence variation among radiating species is often remarkably low. Phylogenies of the Hawaiian mints have been reconstructed from datasets composed of commonly used DNA sequence markers, including cpDNA and nuclear ribosomal 5S non-transcribed spacer and external transcribed spacer regions (Lindqvist and Albert, 2002; Roy et al., 2013), which are rapidly evolving in many other species (Baldwin and Markos, 1998; Shaw et al., 2005; Small et al., 1998), but they fail to resolve relationships among the Hawaiian mints. A similar pattern of low genetic diversity has been noted in other recent radiations of Hawaiian plants (Appelhans et al., 2014; Baldwin and Sanderson, 1998; Cronk et al., 2005; Givnish et al., 2009; Givnish et al., 2013). Given this low diversity in recent radiations, large amounts of data are clearly required to increase resolution. Moreover, to resolve reticulate relationships, robust phylogenies based on both uniparental and biparental markers are required. Since standard cpDNA markers commonly used in plant molecular phylogenetics are insufficient to trace the ancestry of the Hawaiian lineage, we undertook a phylogenomic approach.

With the development of next-generation sequencing technologies (NGS) it is now possible to generate large amounts of sequence data for non-model taxa (Ekblom and Galindo, 2011). Sequences of complete chloroplast genomes are now becoming available for many contemporary species (Moore et al., 2006), and phylogenetic analyses of complete chloroplast genomes have been found to substantially increase resolution in taxa that have undergone rapid radiations (Parks et al., 2009), bypassing previous trade-offs between the number of taxa and the number of nucleotides utilized (Martin et al., 2005; Soltis et al., 2004). However, while generating large amounts of data has become relatively easy, sequencing and assembling complete chloroplast genomes, particularly from historical and ancient material, can still be challenging. DNA obtained from herbarium samples is highly degraded (Staats et al., 2011; Weiss et al., 2015). While its fragmented nature is well suited for the relatively short read lengths produced by second generation sequencing technologies (Knapp and Hofreiter, 2010; Staats et al., 2013), it can be difficult to work with. For example, it has been demonstrated that the decay rate in herbarium samples is about ten times faster than the rate in bones (Weiss et al., 2015), and that 96 – 99% of endogenous DNA may be lost during long term tissue storage (Staats et al., 2011). In addition, ancient material often contains DNA from environmental contaminants with genomic sequencing typically limited to low fractions of endogenous DNA. For example, in 70 year-old herbarium material, less than 5% of reads were from the target species (Staats et al., 2013). This can complicate high quality assembly of complete genomes, unless extensive sequencing is conducted to recover as much of the endogenous genome as possible. Furthermore, mapping reads to a reference genome can be complicated by divergence between the reference and species of interest (Ruffalo et al., 2011). *De novo* assembly, on the other hand, can be time consuming and computationally intensive (Miller et al., 2010a), and may provide only limited success for herbarium samples (Staats et al., 2013). Comparative, reference-guided assemblers

(Ratan, 2009) take an intermediate strategy, and can provide a powerful approach for taxa like the Hawaiian mints, for which no close reference sequence is available.

The aims of this study were to 1) test the application of NGS to Hawaiian mint herbarium material and sequence the first complete chloroplast genome of an Hawaiian mint to facilitate resequencing and assembly of additional chloroplast genomes from representatives of Stachydeae, including members of all three genera of Hawaiian mints; 2) conduct plastid phylogenomic analysis with extensive taxonomic sampling to provide a well-resolved maternal phylogeny, which is imperative for assessments of hybrid origins and the evolutionary history of a major Hawaiian radiation; and 3) examine the pattern of variation among these chloroplast genomes to evaluate levels of variability among Hawaiian mints and their close relatives, including regions traditionally used in plant molecular phylogenetics. This is the first NGS study of a major Hawaiian radiation of plants. By using a shotgun next-generation sequencing approach, we were able to assemble a high-coverage, complete chloroplast genome from an extinct Hawaiian species and generate extensive data from other extinct and endangered taxa from up to ~100 year-old herbarium specimens. This also represents the first chloroplast genome in the subfamily Lamioideae, the second-largest subfamily within the ecologically and economically important Lamiaceae, for which only few chloroplast genomes have been sequenced so far. Our study highlights the feasibility of a genome-scale approach to take advantage of cataloged, but rare and possibly extinct historical collections. It also accentuates that even a genomic approach may not be able to resolve recently evolved reticulate lineages.

2. Materials and methods

2.1. Taxon sampling

Eighteen samples were selected for shotgun sequencing and assembly of complete chloroplast genome sequences (Table 1). These included 12 Hawaiian mint species (five from contemporary sample collections and seven from herbarium collections ranging up to ~100 years old), of which two extinct species were represented by two accessions each. We also sequenced four *Stachys* species, representing both close and more distantly related taxa. To conduct an extensive phylogenetic analysis of this lineage we selected 96 additional taxa for targeted amplicon sequencing, including 48 out of ~59 described Hawaiian species, as well as representative taxa among close Old and New World relatives from all major recognized lineages (Table 2). Specifically, we conducted targeted sequencing for 2 *Haplostachys*, 29 *Phyllostegia*, and 22 *Stenogyne* accessions, plus 46 *Stachys* and 1 *Suzukia* outgroup taxa. Samples ranged in age from 4 – 105 years, with an average age of 32 years (Supplementary Fig. 1). The contemporary samples have been stored in silica gel at room temperature after sampling.

2.2. Library construction and next-generation sequencing

Approximately 100 mg dried leaf tissue for each sample was homogenized using the TissueLyser system (Qiagen), and DNA was extracted using the DNeasy plant mini kit (Qiagen). DNA isolated from herbarium samples for shotgun Illumina sequencing was processed separately from contemporary samples in a PCR hood using dedicated supplies and reagents.

Overnight UV decontamination of pipettes, working surfaces, plastics, and buffers, as well as extraction and negative controls were used to minimize and detect any potential contamination. To confirm the presence of target DNA in herbarium DNA extracts, we amplified a short chloroplast region, the *trnL* intron, following the protocol described in Lindqvist & Albert (2002).

For contemporary samples, DNA was sheared to a size range between 200 – 600 bp using ultrasonication in a Covaris S220, and then indexed, shotgun genomic sequencing libraries were prepared following the standard Illumina TruSeq protocol. For herbarium samples, we used the NEBNext DNA library prep master mix kit (New England Biolabs). Samples were quantified using the PicoGreen high sensitivity assay, then pooled, and sequences were produced on the Illumina MiSeq (150 bp paired-end reads) and HiSeq 2000 platforms (48 bp single-end reads and 100 bp paired-end reads).

Reads were demultiplexed and adapter sequences were trimmed using the program AdapterRemoval (Lindgreen, 2012). To assess DNA damage in old herbarium specimens, we examined the patterns of C→T and G→A misincorporation in the sequencing reads (from shotgun sequenced samples) using the program mapDamage2.0 (Jónsson et al., 2013). For this analysis, we mapped reads to a chloroplast genome sequence assembled from a contemporary specimen of *Haplostachys haplostachya* (see below). We investigated the C→T misincorporations at the first position of the 5' end of the aligned reads, and compared the results to data from ancient mammal remains up to ~2500 years in age described by Sawyer et al. (2012). We also investigated plots of the C→T and G→A misincorporation rates at both the 5' and 3' end of the reads.

2.3. Assembly of the Hawaiian mint reference chloroplast genome

Because no chloroplast genome sequences from closely related species were available at the time, we implemented a combined reference-guided and *de novo* approach (Cronn et al., 2008) to assemble the sequence for *Stenogyne haliakalae*, the Hawaiian mint species with the greatest number of reads. First, we conducted a reference-guided assembly using YASRA 2.32 (Ratan, 2009) with olive (*Olea europaea,* NC_013707) (Mariotti et al., 2010) as the reference, which was the closest chloroplast genome sequence available at the time, and a setting of ≥75% identity between the reads and the reference. We also conducted *de novo* assembly of the shotgun data using default settings in SOAPdenovo v1.05 (Li et al., 2010). Contigs from these analyses were combined into a single dataset, split into overlapping sequences of 400 bp, and a second YASRA analysis performed using the same settings as above. The order of the resulting contigs, as compared to the olive sequence, was determined using BLAST searches.

In order to close the gaps between contigs, we designed 38 pairs of PCR primers using Primer3 (Rozen and Skaletsky, 2000), with primers located on the ends of each contig (see Supplementary Table 1 in the companion data paper by Welch et al., 2016). Since the length of the gaps between the contigs was not precisely known, we used DNA extracted from a high-quality contemporary sample of *S. bifida*. Amplifications were conducted using 1X AmpliTaq PCR Buffer II, 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 5 mM TMACl, 0.04% BSA, 0.4 μM each primer, 1 unit AmpliTaq DNA Polymerase (Applied Biosystems), and 1 μL DNA extract in 25 μL total reaction volume. Reactions were incubated at 95°C for 2 min, and 35 cycles of 95°C for 30 sec, 50°C for 30 sec, and 72°C for 1 min, followed by a 4 min incubation at 72°C. PCR products were Sanger sequenced in both directions at the University of Washington High

Throughput Genomics Center in Seattle, Washington. Sanger sequences were trimmed, edited, and then aligned to contig sequences in Sequencher 4.7 (Gene Codes). A pseudo-reference sequence was made of the resulting assembly, as described in Cronn et al. (2008).

Reads from *S. haliakalae* were mapped to the pseudo-reference sequence using the program BWA v. 0.6.2 (Li and Durbin, 2009). The assembly was viewed using the program Tablet (Milne et al., 2013), and then further refined through Sanger sequencing of eight additional regions (see Supplementary Table 1 in Welch et al., 2016, for primer information) where sequence coverage or mapping quality was low (e.g., border of the inverted repeat and small single copy region, in tandem repeat regions, and homopolymer runs). Reads were mapped to the final assembly sequence, PCR duplicates were flagged using the MarkDuplicates tool of the Picard command-line tools software suite (http://picard.sourceforge.net/index.shtml), and a consensus was called using SAMtools v. 0.1.19 (Li et al., 2009). The reference genome assembly was initially annotated using DOGMA (Wyman et al., 2004), with the annotations refined using tRNAscan-SE (Schattner et al., 2005) and additional BLAST searches. The inverted repeats were analyzed using the program Inverted Repeats Finder (Warburton et al., 2004) with T7 set to 500,000 and a maximum loop size of 1,000,000. A diagram of the chloroplast genome assembly was created using Circos (Krzywinski et al., 2009).

2.4. Assembly of additional mint chloroplast genome sequences

We also assembled chloroplast genome sequences for 11 additional mint species, for which sufficient data were available (Table 1), using methods similar to those described above. For the Hawaiian mints we conducted reference-guided assembly using YASRA for each species with *S. haliakalae* as the reference sequence. A consensus sequence was called, and reads were mapped again using BWA. A final species-specific consensus sequence was called using SAMtools. To assemble the chloroplast genomes for the *Stachys* species, we first assembled the sequence for *Stachys chamissonis*, as this represents the species that is most closely related to the Hawaiian mints in phylogenies constructed from short chloroplast sequences (Lindqvist and Albert, 2002; Roy et al., 2013). We conducted five separate YASRA runs, using recently available chloroplast genome sequences as references: *S. haliakalae*, *Olea europaea*, *Sesamum indicum* (NC_016433) (Yi et al., 2012), *Origanum vulgare* (JX880022) (Lukas and Novak, 2013), and *Salvia miltiorrhiza* (NC_020431) (Qian et al., 2013). The contigs from the runs with each reference were aligned and an interim consensus sequence called. The reads were then mapped to this interim sequence, and a final consensus sequence was created with SAMtools. The *Stachys chamissonis* sequence was used as the reference for YASRA assemblies of the remaining *Stachys* species. However, for *Stachys byzantina*, the most divergent species of our taxa, additional YASRA runs were conducted using *Olea europaea* and *Sesamum indicum* as references. Genome sequences were annotated as described above.

We used Mauve 2.3.1 (Darling et al., 2010) to compare the structure and similarity of the mint genome sequences to each other and to 13 available Lamiales chloroplast genomes: *Ajuga reptans* (NC_023102), *Andrographis paniculata* (NC_022451), *Boea hygrometrica* (NC_016468), *Jasminum nudiflorum* (NC_008407), *Lindenbergia philippensis* (NC_022859), *Olea_europaea* (NC_013707), *Origanum vulgare* (JX880022), *Pinguicula ehlersiae* (NC_023463), *Salvia miltiorrhiza* (NC_020431), *Schwalbea americana* (NC_023115), *Sesamum_indicum* (NC_016433), *Tectona grandis* (NC_020098), and *Utricularia gibba* (NC_021449). For all analyses we set the gap opening penalty to -200 and the gap extension penalty to -30. For comparisons of mint sequences to each other, we assumed the chloroplast

genomes were collinear and set the seed weight to 9. Similarly, to compare mint chloroplast genomes with other Lamiales without rearrangements, we assumed the sequences were collinear, and set the seed weight to 7. For comparisons of mint sequences to other Lamiales that demonstrate inversions and rearrangements, we trimmed off one of the inverted repeats at coordinates suggested by Inverted Repeats Finder so that homology could be determined in the remaining region. Seed weight was set to 19 for this analysis.

2.5. Investigation of variable sites and regions

To further investigate the distribution of variable sites across the chloroplast genomes of the Hawaiian mints and the *Stachys* outgroups, we identified single nucleotide polymorphisms (SNPs) among six high quality chloroplast genomes: *Stenogyne haliakalae, Stenogyne bifida, Haplostachys haplostachya, Stachys chamissonis, Stachys coccinea,* and *Stachys sylvatica*. These species represent all of the main lineages within our samples, and for which chloroplast genomes were sequenced at > 10X depth. We used BWA to map the reads for each of these species onto the *Stenogyne haliakalae* reference genome as described above. The programs SAMtools and BCFtools were used to call SNPs with a SNP quality score > 30. All SNPs were manually examined in Tablet and SNPs on edges of indels and repetitive regions were excluded. Annotations of the *S. haliakalae* reference genome were transferred to the locations of the SNPs for comparisons of variability with the most commonly used plastid regions for phylogenetic analyses (Shaw et al., 2005; Shaw et al., 2007). We compared the levels of chloroplast genome diversity among the six genomes by identifying unique, variable positions, which we refer to as potentially informative characters (PICs) (Shaw et al., 2005; Shaw et al., 2007). Note that the definition of PIC used here excludes indels and inversions, which have been included in other studies of locus variability (Shaw et al., 2005; Shaw et al., 2007).

We conducted several analyses of the PIC datasets to examine patterns of variation in the Hawaiian mints and the *Stachys* outgroups. For coding, intron, and intergenic spacer and pseudogene regions we compared the number of PICs per locus. However, this approach does not take into account the length of the locus, and very long loci will appear to have more PICs than shorter loci. Therefore, we divided the number of PICs by the total length of the region to give the percent PICs per locus. However, very short regions could still appear to have a high percentage of variable sites, when in fact only a small number of the sites were variable. We show both measures in the companion data paper (Welch et al., 2016). We also investigated variability across the genomes by calculating these measures for 1,000 bp non-overlapping sliding windows across the entire chloroplast genome sequences.

2.6. Targeted resequencing of variable regions

We used a multiplex, tailed amplicon sequencing approach to target the most variable regions of the chloroplast genome for sequencing and phylogenetic analyses in an extensive set of 105 mint taxa (Fig. 2). To identify variable regions, reads from 15 taxa subjected to shotgun sequencing (including the partial genomes from all of the historical samples, except *Phyllostegia variabilis*) were mapped to the *Stenogyne haliakalae* reference sequence using BWA. These taxa represent members from all three genera of Hawaiian mints and close and more distant *Stachys* relatives. SNPs were called with SAMtools and BCFtools, filtering out those with SNP quality < 30. From among these SNPs we selected a total of 108 variable loci identified from the single copy regions, including (1) all the regions that had a variant position among the Hawaiian mints

(except where every individual had an alternative allele compared to the reference sequence) and (2) additional regions that had variant positions among at least two of the *Stachys* species. 100 bp of flanking sequence on either side of the SNPs was retrieved from the reference genome, and PCR primers were designed using BatchPrimer3 (You et al., 2008) and examined manually for quality control. Sequences complementary to the Illumina sequencing adapters were appended to the end of each primer (see Table 9 in the companion data paper by Welch et al., 2016).

Multiplex PCR amplification was conducted using 12-13 loci per reaction, for a total of 8 reactions per sample. Reactions were prepared on ice in 50 µL total volume with 1X Amplitaq Gold PCR buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 µM each forward and reverse primer, 0.04% bovine serum albumin (BSA), 5% dimethyl sulfoxide (DMSO), 5 mM tetramethylammonium chloride (TMACl), 1 unit of Amplitaq Gold DNA polymerase and 30 ng total DNA. Amplification was conducted using the following thermal cycling protocol: 94°C for 10 min, then 40 cycles of 94 °C for 30 sec, 50 °C for 30 sec, 72 °C for 30 sec, followed by 72 °C for 7 min. Reactions were pooled and purified using 0.65x Agencourt AMPure XP magnetic beads according to the manufacturer protocol, except using 80% ethanol for wash steps and eluting in 1X TE buffer. A second, indexing PCR was conducted to add sample specific indices and extend the sequencing adapters to full length. Reactions were conducted in 25 µL volumes, with 1X Amplitaq Gold PCR buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 2.5 µL of P5 Index and 2.5 µL of P7 Index 2 from the Illumina Nextera Index Kit, 0.04% BSA, 5% DMSO, and 5 mM TMACl, 1 unit of AmpliTaq Gold and 10 µL of the cleaned multiplex PCR product. Amplification was conducted using the following thermal cycling conditions: 94 °C for 10 min, then 5 cycles of 94 °C for 30 sec, 63 °C for 30 sec, 72 °C for 30 sec, and a final elongation of 3 min at 72 °C. Reactions were purified using the AMPure XP magnetic beads as above, but eluting in Qiagen EB buffer instead. Libraries were quantified using the Qubit dsDNA HS assay, pooled and 150 bp paired-end sequences were produced on the Illumina MiSeq platform.

Sequencing reads were demultiplexed, mapped to the *S. haliakalae* reference genome and SNPs were called using SAMtools as above. Any variable sites with a SNP quality score < 30, as well as those occurring in the PCR primer locations, were discarded. Four loci with low coverage in the majority of individuals were excluded as well. The remaining SNPs were manually examined in Tablet and those that occurred on edges of gaps and homopolymer repeats were excluded, as above. Fifteen sites assigned a genotype of "0/1" by SAMtools were tagged as potential sites of heteroplasmy or SNP calling errors, and hence were removed from further consideration in phylogenomic analyses (see Supplementary Table 1 for a list of sites). The final SNP dataset contained 750 loci, which represented sequence data from 18,670 bp of the most variable chloroplast regions in 105 mint taxa (alignments are deposited in the TreeBASE data repository, http://purl.org/phylo/treebase/phylows/study/TB2:S18917).

2.7. Phylogenetic and other data analyses

We conducted phylogenomic analyses of the 12 complete mint chloroplast genome sequences described here, and additional analyses of these sequences plus 13 Lamiales sequences available in Genbank: *Ajuga reptans*, *Bartsia inaequalis* (KF922718), *Boea hygrometrica* (NC_016468), *Jasminum nudiflorum*, *Lindenbergia philippensis*, *Olea_europaea*, *Olea_woodiana* (NC_015608), *Origanum vulgare*, *Pinguicula ehlersiae*, *Salvia miltiorrhiza*, *Sesamum_indicum*, *Tectona grandis*, and *Utricularia gibba*. Tobacco (*Nicotiana tabacum,* Z00044) was used as outgroup. Sequences were aligned with MAFFT v7.164b using the FFT-NS-2 progressive alignment option (Katoh et al., 2002; Katoh and Standley, 2013). Maximum

likelihood analyses were carried out on an unpartitioned data set using the RAxML blackbox web-server (Stamatakis et al., 2008) assuming the general time reversible substitution model with rate heterogeneity modeled by a gamma distribution with four discrete rate categories (GTR + G). Branch support was assessed using 1000 bootstrap replicates. Bayesian inference was conducted using MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003) on the CIPRES Science Gateway (Miller et al., 2010b) on the unpartitioned data set. The GTR substitution model was used as before, but with rate heterogeneity modeled by both a gamma distribution and a proportion of invariable sites. The default priors were utilized, along with the default heating scheme (one cold and three heated chains), and runs were conducted for 20 million generations with the first 10% discarded as burn-in. Multiple independent analyses were conducted from random starting seeds. Plots of the likelihood across generations and the effective sample size (ESS) for each parameter (>500 in nearly all cases) were examined in Tracer v1.5 (Drummond and Rambaut, 2007). Trees were viewed in FigTree v1.4.0 (Drummond and Rambaut, 2007).

We also conducted phylogenomic analyses for 105 Hawaiian mint and *Stachys* taxa on the dataset of 750 variable positions identified from sequencing approximately 20,000 bp from the most variable regions from across the chloroplast genome (Fig. 2). This comprehensive analysis represents nearly complete taxonomic sampling for the Hawaiian lineage, including both extinct and extant taxa. Individuals used for shotgun sequencing were also included in this dataset. Analyses were conducted with an unpartitioned data set and with the data partitioned into coding and non-coding sites. We used jModelTest v2.1.6 and the Akaike information criterion (AIC) to determine the best fitting substitution model. The GTR+I+G model was selected for the unpartitioned data, and the GTR and the GTR+I+G models were selected for the coding and non-coding partitions, respectively. Maximum likelihood analyses were carried out using the RAxML blackbox web-server with these substitution models and 1000 bootstrap replicates. Bayesian inference was conducted using MrBayes 3.1.2, as described above. Runs were conducted for 30 million generations. Support for the partitioned vs the unpartitioned model was assessed using the likelihood ratio test and estimation of Bayes Factors in Tracer v1.5.

Since we expect only small genetic distances among the Hawaiian species and a network may better express the multitude of plausible trees, we created a median-joining network using the program SplitsTree v4.12.3, in which alternative potential evolutionary paths are displayed in the form of cycles (Bandelt et al., 1999; Huson and Bryant, 2006). In addition, a principal component analysis was performed using GenAlEx 6.501 (Peakall and Smouse, 2012) and the standardized covariance method with pairwise genetic distances calculated between haplotypes. The first two principal components were plotted.

To test the effect of potential DNA damage in old herbarium specimens, we also conducted phylogenetic analyses using BEAST v1.8.2 implementing two different DNA damage models (Rambaut et al., 2009). We used Bayes Factors estimated in Tracer v1.5 to compare the model without DNA damage to either the model with age dependent damage at all sites or the model with age dependent damage at sites with transitions only. For all runs we utilized the partitioned data set with the substitution models described above, linking the tree and clock models across partitions. The ages of the samples were known from the year they were collected from the field, and therefore no distributions were specified for the prior on the tip dates. The uncorrelated log-normal relaxed clock was implemented with the Birth-Death tree prior, and the remaining priors were left at their default values. Multiple independent runs of 40 $x10^6$

generations were conducted using different random starting seeds, and the results were analyzed in Tracer as described above.

3. Results

3.1. Patterns of DNA damage in herbarium specimens

We obtained between ~1 and 18.8 million shotgun Illumina reads for six species sampled from herbarium collections, ranging in age up to ~100 years (shotgun sequencing reads for all samples, including herbarium and contemporary samples, have been deposited in NCBI's sequence read archive SRP070171). One sample, collected in 1903, failed to produce sufficient reads for inclusion in the analysis. The presence of C→T and G→A misincorporations across the length of the reads, with minor increases at the 5' and 3' ends, respectively, suggests evidence for the presence of damaged sites in the DNA molecules (Supplementary Fig. 2). We compared the frequency of C→T misincorporations at the first position of the 5' end of mint reads to levels observed in reads obtained from the remains of mammals ranging up to ~2500 years in age (Sawyer et al., 2012). Overall, levels of damage observed in the mints were low and within the range expected for samples of this age (Supplementary Fig. 3).

3.2. Stenogyne haliakalae chloroplast genome assembly and characterization

To conduct a high-resolution investigation of the evolutionary history of the Hawaiian mint radiation, we assembled the chloroplast genome for the extinct Hawaiian species *Stenogyne haliakalae*, collected from the island of Maui in 1936. The initial reference-guided assembly, using olive as the reference, resulted in 60 contigs ranging in length from 73 – 17,056 bp with an N50 of 5,125 bp. The initial de novo assembly resulted in 39,212 contigs ≥ 100 bp, ranging up to 1,859 bp in length, with an N50 of 155 bp. These contigs were used as input for a second reference guided assembly, which produced 38 contigs ranging in length from 185 – 17,056 bp, with an N50 of 6,308 bp. In our assembly, contigs stretched across all of the borders between the repeats and single copy regions. Sanger sequencing was used to close the remaining gaps, and to refine and validate the sequence. Mapping the shotgun sequencing reads against the final assembly confirms the location and orientation of the inverted repeats and suggests a complete sequence as no gaps or ambiguities were detected.

The final assembled chloroplast genome sequence for *Stenogyne haliakalae* is 149,736 bp (average of 50X read depth, Fig. 2). The sequence demonstrates the typical quadripartite structure of other angiosperms, with a large single copy (LSC) region of 81,364 bp and a small single copy (SSC) region of 17,499 bp, separated by a pair of inverted repeats (IRa and IRb) of 25,436 bp. The length and structure of the sequence is similar to other plastid genome sequences of members in the order Lamiales (Lukas and Novak, 2013; Mariotti et al., 2010; Qian et al., 2013) (Fig. 3, see also Fig. 1 in the companion data publication Welch et al., 2016). The overall GC content is 38.5%, with the GC content of the LSC, SSC, and IR being 36.7%, 32.5%, and 43.4%, respectively. The higher GC content of the IR reflects the higher GC content of the four ribosomal RNAs present in the IR (Fig. 2).

The gene content and order of the *Stenogyne haliakalae* sequence is also similar to other Lamiales (Fig. 3). The chloroplast genome contains 114 unique genes (see Table 2 in Welch et

al., 2016). Eighty of these are protein coding genes: seven coding for Photosystem I, 15 for Photosystem II, six for Cytochrome b6/f, six for ATP synthase, one for Rubisco, 11 for NADH oxidoreductase, four for RNA polymerase, nine for large subunit ribosomal, and 12 for small subunit ribosomal proteins. Six genes code for proteins with other functions (maturase, protease, translation initiation factor, cytochrome c biogenesis, acetyl-CoA-carboxylase, and envelope member proteins) and three code for proteins with unknown functions. In addition to protein coding genes, there are four ribosomal RNA genes as well as 30 transfer RNA genes, which represents all 20 amino acids. A total of 18 genes occur in the inverted repeats: seven full or partial protein coding genes, all four ribosomal RNA genes, and seven transfer RNA genes. Therefore, a total of 132 predicted functional genes were identified. Eighteen of these genes contain an intron (see Table 3 in Welch et al., 2016), and the rps12 gene is trans-spliced with the 5' end in the LSC and the 3' end duplicated in the IR. Overall, 52.8% of the chloroplast genome codes for proteins, introns represent 13.3% of the genome, ribosomal RNAs account for 6.0%, and transfer RNAs make up 1.8%. The remaining 26.1% is non-coding and consists of intergenic spacers and pseudogenes.

3.3. Additional mint chloroplast genome assemblies and comparison to other Lamiales

Between 1.7 and 20.2 million shotgun Illumina reads were obtained for 15 additional species, from samples ranging up to 100 years in age. We were able to assemble complete or nearly complete chloroplast genomes for 11 species with read depths ranging between 6.7 and 193x (Table 1). Of the 12 chloroplast genomes described here, two complete genome sequences were obtained from samples originally collected in the early to mid 1930s (*Stenogyne haliakalae* and *Stenogyne kanehoana*), and one nearly complete (~89%) genome sequence was obtained from *Haplostachys linearifolia* samples collected in 1910 and 1918. These three species are either extinct or critically endangered today (Wagner, 1999). Substantial portions of the chloroplast genome (35 – 42%) were obtained for three additional extinct or endangered species using herbarium samples up to 100 years old. While the approach was largely successful overall, one sample from the extinct species *Phyllostegia variabilis*, collected in 1903, failed to produce sufficient data.

Chloroplast genome structure was highly conserved among the endemic mints, as expected, and also among the *Stachys* outgroups (see Fig. 2 in the companion data publication Welch et al., 2016). Sequence lengths among all 12 cp genomes sequenced ranged from 149,722 bp in *Stachys byzantina* to 150,674 bp in *Stachys coccinea,* with a mean of 150,124 bp (Table 1). All demonstrated the quadripartite structure typical of angiosperms with an LSC (mean = 81,718 ± 66 bp standard error; see Table 4 in Welch et al., 2016) and SSC (mean = 17,513 ± 28 bp), separated by a pair of inverted repeats (mean = 25,446 ± 5 bp). Most variation in length was due to small indels within the LSC, however there was one indel of 407 bp in the intergenic spacer between tRNA-Cys and petN in *Stenogyne haliakalae*, and three indels of 135, 145, and 111 bp in *Stachys byzantina*, which occurred in the intergenic spacers between rps16 and trn-Q, atpH and atpI, and trn-L and trn-F, respectively. This latter intergenic spacer is a common locus used in phylogenetic inference in plants, and a Sanger sequence of *Stachys byzantina* for this region (Genbank accession KF529898) determined by Salmaki et al. (2013) matched the assembly with 100% identity, including the 111 bp indel. All of the assembled sequences demonstrated the same gene content and order as described for *Stenogyne haliakalae* above, including the gene content of the inverted repeats.

We further compared the structure of *Stenogyne haliakalae* and *Stachys byzantina* (the most distantly related species of those assembled) with additional species of the Lamiales (Fig. 3 and Fig. 1 in Welch et al., 2016). Structures of most Lamiales chloroplast genomes are generally highly conserved (including carnivorous members of Lentibulariaceae, Fig. 1 in Welch et al., 2016), with only short regions differing between them. The most notable exceptions to this are an inverted region within the LSC in *Schwalbea americana*, an inverted and rearranged LSC region of *Jasminum nudiflorum*, and an inverted SSC region in *Andrographis paniculata*, *Schwalbea americana, and Jasminum nudiflorum* (Fig. 1 in Welch et al., 2016). Phylogenomic analyses of complete chloroplast genomes (Fig. 4, Supplementary Fig. 4) confirm the placement of the Hawaiian mints within the Lamiaceae, the monophyly of the Hawaiian species, and that the Hawaiian mints are most closely related to Meso-South American *Stachys*, although there is still poor resolution within the Hawaiian lineage. All chloroplast genome sequences have been deposited in the GenBank database under accession numbers KU724131-KU724141). Alignments and tree files are available in the TreeBASE data repository (http://purl.org/phylo/treebase/phylows/study/TB2:S18917).

3.4. Chloroplast genome sequence variability within a mint radiation

A similarity plot from Mauve of the 11 complete chloroplast genome sequences from Hawaiian and *Stachys* taxa (excluding *Haplostachys linearifolia,* due to missing data) showed a high degree of similarity (see Fig. 2 in the companion data publication Welch et al., 2016), even between *Stenogyne haliakalae* and the most distant *Stachys* species. To conduct a more detailed investigation of chloroplast genome variation in this mint radiation, we compared variable positions, or potentially informative characters (PICs) (Shaw et al., 2007; Shaw et al., 2014), among six complete, high-quality chloroplast genome sequences (i.e. > 10X depth): *Stenogyne haliakalae, Stenogyne bifida, Haplostachys haplostachya, Stachys chamissonis, Stachys coccinea,* and *Stachys sylvatica.* These species represent each of the major maternal lineages in this group (Fig. 6), and we compared the PICs of the three *Stachys* to those of the three Hawaiian mints. A total of 565 PICs were identified among *Stachys* species, while only 104 were variable among the three Hawaiian mint species (Fig. 2).

To examine the distribution of variable sites across the chloroplast genome, we conducted a sliding window analysis of non-overlapping 1,000 bp segments (Fig. 5, Table 5 in Welch et al., 2016). Overall, the large and small single copy regions were more variable than the inverted repeat region, and the three *Stachys* taxa were much more variable than the three Hawaiian taxa. For *Stachys*, the mean # PICs per window was 4.52 (median = 4.00), while for the Hawaiian mints the mean was just 0.83 (median = 0.00). For the Hawaiian taxa, most windows (66 of 125) had no variable sites, but only 13 windows lacked variable sites among the *Stachys*. More conserved regions often flanked windows with higher variability, and therefore no consecutive windows of particularly high variability could be identified.

In order to compare patterns and levels of variability with the Hawaiian mint radiation to those of other closely related mints, as well as more distantly related angiosperm taxa, we investigated if particular genes or non-coding regions demonstrated higher levels of variability. We calculated the number of PICs per locus, as well as the % PICs per locus, to take into account region length. For protein coding genes, among *Stachys* species, 57% of exons longer than 100 bp contained at least one PIC (52/91; see Figs. 3a & 4a and Table 6 in Welch et al., 2016), while for the Hawaiian taxa only 20% had at least one PIC (18/91). As with the sliding window analysis, the *Stachys* taxa demonstrated higher variability (mean % PICs per locus =

0.44) than the Hawaiian lineage (mean % PICs per locus = 0.07). For both groups of taxa, coding regions (which make up about 60% of the chloroplast genome) cumulatively contained ~40% of all the PICs identified (225/565 for *Stachys* and 40/104 for the Hawaiian lineage). The SSC portion of the *ycf1* gene contained the highest overall number of PICs (see Fig. 3a and Table 6 in Welch et al., 2016), however this region is fairly long (~4500 bp) and so has a moderate percentage of PICs. Among *Stachys rps19*, *rpl36*, *rpl32*, *matk*, and *rbcL* had the highest percentage of PICs. These overlap partially with the most variable loci in the Hawaiian lineage, which were *ndhE*, *rbcL*, and *matK* (see Fig. 4a and Table 6 in Welch et al., 2016). Most of these loci are less than about 300 bp in length, except for *rbcL* and *matK*, which are both relatively long and have a higher percentage of PICs.

Unsurprisingly, non-coding regions were more variable than coding regions, with intergenic spacers and pseudogenes having higher variability than introns. Among *Stachys* species, 17/18 intron sequences contained at least one PIC, while the Hawaiian taxa had PICs in only seven introns (see Figs. 3b & 4b and Table 7 in Welch et al., 2016). Again, variability was higher overall among the *Stachys* taxa (mean %PICs per locus = 0.61, median = 0.56) than in the Hawaiian taxa (mean = 0.10, median = 0.00). These regions cumulatively contained ~13% of all PICs identified in the genomes for both taxon sets. The *ndhA* intron contained the largest number of variable sites for all taxa, as well as a relatively high %PICs. The portion of the *trnK-UUU* intron between the first exon and the *matK* gene (which is contained within the intron) had the highest percentage of variable sites for *Stachys*. For intergenic spacers and pseudogenes, 72% of regions longer than 100bp were variable in *Stachys* (65/90), whereas for the Hawaiian species, only 32% were variable (29/90, see Figs. 3c & 4c and Table 8 in Welch et al., 2016). As before, variability was higher among *Stachys* (mean % PICs per locus = 0.94, median = 0.82) than in the Hawaiian lineage (mean = 0.16, median = 0.00). Intergenic spacers cumulatively contained ~43% of PICs for both groups of taxa, although these regions represent only 26% of the entire chloroplast genome sequence. For *Stachys* species, three regions (*rps16-trnQ*, *rpl20-rps12*, and *rpl32-trnL*) had > 10 PICs, and all had over 1% PICs, with the *rpl32-trnL* region having 1.76% PICs. Three different regions had > 2% PICs: *ccsA-ndhD*, *rps18-rpl20*, and *trnH-psbA*. These regions largely do not overlap with the most variable regions in the Hawaiian lineage. Among those species, three regions had more than 2 PICs (*petA-psbJ*, *rpoB-trnC*, and *atpH-atpI*), however the three regions with the highest percentage of PICs were *rpl33-rps18* (0.68%), *rpl16-rps3* (0.71%), and *psbI-trnS* (0.83%). Only the *petA-psbJ* region had a relatively high number of PICs in both *Stachys* (9) and Hawaiian taxa (3), although the percentage PICs was moderate in both (0.85% in *Stachys*, 0.28% in Hawaiian).

3.5. Targeted re-sequencing and phylogeny of the Hawaiian mint radiation

We used an amplicon sequencing approach to target 108 variable regions from across the chloroplast genome of the mint species investigated here (Fig. 2). As opposed to genotyping these 108 identified SNPs, this allowed us to discover numerous additional variable sites in the flanking regions when sequencing a larger number of taxa, and it also mitigates ascertainment bias. Out of the 108 regions (representing 19,273 bp), four were discarded for poor amplification or low sequence coverage (Mint204, Mint771, Mint43421, and Mint 57465), and all of the remaining regions were validated as containing one or more variable loci. Thus, after removing these regions, our targeted sequencing data set represented 18,670 bp of the most variable regions from across the chloroplast genome and phylogenetic analyses were conducted utilizing 750 variable SNP loci sequenced in 105 taxa (Table 2).

Both maximum likelihood and Bayesian inference produced congruent topologies (Fig. 6) with strong support for a clade of Hawaiian taxa. Both approaches demonstrated strong support for the model with data partitioned into coding vs. non-coding sites (ML: unpartitioned AIC = 18535.7 vs. partitioned AIC = 18190.6; BI: 2ln(Bayes Factors) = 36.44 in favor of the partitioned model). Analyses in BEAST incorporating models of age-dependent sequence error, potentially due to DNA degradation in old herbarium samples, were not preferred over the model without sequence error (2ln(BF) = 0.74 in support of no error model over a model of age dependent error for transitions only; 2ln(BF) = 2.09 in support of no error model over a model of age dependent error at all sites) and topologies produced under these models were consistent with topologies without the error models. The uncorrelated lognormal relaxed clock analyses in BEAST did indicate some evidence for rate variation among branches, giving an estimate for the coefficient of variation of 1.315 (95% HPD ranges from 1.01 – 1.686 and does not include zero), but again the topology was consistent with analyses from MrBayes and RAxML, with the exception of the placement of two Old World *Stachys* taxa, which had very low posterior probability in the BEAST analysis (Supplementary Fig. 5).

Overall, the phylogeny (Fig. 6) confirms previous results showing a strongly supported clade comprising mostly Hawaiian and New World *Stachys* taxa in addition to some Old World *Stachys*, including the African species *S. alpigena*, *S. natalensis*, and *S. aethiopica* (Lindqvist and Albert, 2002; Roy et al., 2013). Well-resolved clades of temperate NA *Stachys*, Hawaiian mints, and two Meso-SA American *Stachys* lineages corroborate previous results (Roy et al., 2013), although phylogenetic resolution is overall much increased from these previous analyses based on fewer chloroplast loci. Both phylogenetic analyses and a principal coordinate analysis (Supplementary Fig. 6) confirm a close maternal relationship among Hawaiian and Meso-SA *Stachys* taxa. Resolution within the Hawaiian lineage, however, remains poor and only to a minor degree follows taxonomy, e.g., a strongly supported clade of *Stenogyne microphylla* and *S. rugosa* and their putative hybrids that have previously shown close genetic relationships (Lindqvist et al., 2003). The scattering of individuals across generic boundaries in the median-joining network (Fig. 7) and principal coordinate analyses (Supplementary Fig. 7) further demonstrates the inability to capture overall taxonomic groupings into three separate Hawaiian genera.

4. Discussion

Recent radiations provide excellent opportunities to gain a better understanding of the underlying processes of speciation, but the evolutionary histories of such radiations must first be characterized. In Hawaii, the ancestor of the endemic mints diverged into numerous species over the past few million years, filling new ecological niches throughout the islands, making it one of the most diverse plant lineages in the archipelago. However, as typical in other recent radiations, sequence divergence between these species is low and hence resolving evolutionary relationships has been challenging using traditional gene-by-gene phylogenetic methods. Clear phylogenetic resolution, based on both uniparental and biparental markers, is required for investigating the evolutionary processes underlying these radiations and dissecting putative reticulate origins. This study focused on resolving the maternal evolutionary history of the Hawaiian mint lineage.

4.1. Sequencing chloroplast genomes from extinct and ancient plant specimens

This study represents one of relatively few that have successfully obtained and sequenced ancient DNA from plants using a next-generation approach. Most paleogenomic studies to date have focused on large vertebrates that inhabit cold or temperate areas or the close relatives of crop plants that have abundant genomic information available (Allaby et al., 2015; Jaenicke-Despres et al., 2003), although genome-scale sequencing approaches of herbarium material are increasingly being applied (Bakker et al., 2015; Beck and Semple, 2015; Besnard et al., 2014; Staats et al., 2013; Zedane et al., 2015). The ability to sequence herbarium specimens, which may represent the only samples available for rare, endangered, or extinct taxa, provides great promise towards understanding important issues from biogeographic history to domestication, taxonomy, and conservation.

We used shotgun sequencing and a combination of *de novo* and reference-guided assembly to obtain chloroplast genome sequences of 12 mint species, including extinct taxa that were sampled over a century ago. It is well known that treatment and storage conditions of samples can dramatically reduce and degrade endogenous DNA (Pruvost et al., 2007; Staats et al., 2011). Evidence from our shotgun sequencing data suggests that the old herbarium samples utilized here contain low levels of endogenous DNA and exhibit damage patterns typical of ancient DNA. The percentage of shotgun reads mapped is significantly lower in herbarium samples than in contemporary samples (mean herbarium = 0.4% mapped, mean contemporary = 1.85%, one-way ANOVA $F = 13.71$, $p = 0.0019$). Analyses of potential nucleotide damage patterns suggest that reads from herbarium samples contain C→T and G→A misincorporations that are characteristic of the cytosine deamination process that occurs during DNA degradation (Briggs et al., 2007; Schuenemann et al., 2011). The estimated misincorporation rates were low overall, though, and only marginally higher in the extreme 5' and 3' ends of the reads, making this increase appear less defined than in older vertebrate samples. Overall, the low frequency of misincorporation observed in the mints matched expectations based on the relationship with sample age described by Sawyer et al. (2012) (Supplementary Fig. 3).

We were able to successfully obtain at least 54kb of non-contiguous chloroplast genome sequence from six of the seven species included for shotgun sequencing of herbarium material. Our results demonstrate that next-generation sequencing (NGS) is a useful approach for obtaining data from plant samples collected decades or centuries ago. The short sequencing reads obtained from NGS are well suited to the short DNA fragments typical of old samples, and the large amount of data generated using this approach allows analyses of datasets far richer than those that could feasibly be generated through Sanger sequencing (Miller et al., 2008). The relative abundance of chloroplast DNA in the cell likely facilitated the success of our work. Beyond this, mints may provide an especially good system for ancient DNA studies because the leaves contain relatively few compounds that may inhibit DNA extraction and seem to generally preserve DNA well, as judged from the large number of specimens that have been included here and in earlier molecular systematic studies (Bendiksby et al., 2011; Roy et al., 2013; Roy et al., 2015; Scheen and Albert, 2009; Scheen et al., 2010).

Our study employed a mix of both shotgun and targeted sequencing. The shotgun approach allows huge regions or the entire genome to be sequenced, but in the case of ancient samples it can also yield a large proportion of sequences from environmental contaminants (Noonan et al., 2005). Thus, shotgun sequencing requires either very well preserved samples or much higher sequencing effort, which can lead to trade-offs in the number of samples analyzed. A targeted approach can be particularly useful for focusing sequencing efforts on regions of

particular interest, and is beneficial for lineages exhibiting marginal sequence divergence in which the whole genome sequencing approach may be superfluous. Several targeted enrichment approaches have been developed in recent years (Cronn et al., 2012), but they are often costly and can be time-consuming. An amplicon sequencing approach may be a simple and cheaper alternative, and multiplex PCR reduces the amount of DNA and destructive sampling required from valuable herbarium samples (Krause et al., 2006). However, amplification-based enrichment, the approach utilized here, can also have some drawbacks. For example, multiplex PCR may lead to uneven sequencing coverage among some regions. In our dataset, however, only four loci out of 108 (3.7%) were dropped due to poor amplification or low coverage, so this was not a major issue. Additionally, amplification from the small number of endogenous, and potentially damaged, molecules of ancient DNA can lead to sequencing errors. To investigate the impact that such errors could have on our phylogenetic reconstructions, we built trees in BEAST both with and without models of sequencing error. Investigation of Bayes factors suggested that there was no support for the analyses incorporating models of sequencing errors (for transitions only or for all sites) over the models without it, therefore sequencing errors do not appear to have strongly influenced our results. Another disadvantage of the PCR approach is that targeted fragments must be ~100 bp or larger because of the necessity for PCR priming sites. Most ancient DNA is generally fragmented to sizes less than 100 bp, and so the PCR approach may preclude successful results from poorly preserved samples (Knapp and Hofreiter, 2010). Other enrichment approaches, such as hybridization-based sequence capture (Stull et al., 2013), may circumvent this to some degree, but targeting short sequences is still challenging (Ávila-Arcos et al., 2015).

4.2. Mint chloroplast genome structure and variability

The chloroplast genomes of the Hawaiian mints demonstrate the typical angiosperm quadripartite structure, and the gene content, order, and GC content were consistent with most other members of the Lamiales (Mariotti et al., 2010; Wicke et al., 2014; Zhang et al., 2013). The Hawaiian mint genome did not demonstrate any rearrangements or inversions, which have been found in *Jasmin nudiflorum*, *Schwalbea americana*, and a few other species (Welch et al., 2016) (Lee et al., 2007; Wicke et al., 2013). The lengths of the mint chloroplast genomes, including lengths of the inverted repeat and small single copy region, were similar to other species in this family, such as *Origanum vulgare* and *Salvia miltiorrhiza* (Lukas and Novak, 2013; Qian et al., 2013), with most of the differences in length being due to short indels within the long single copy region.

One instance where the chloroplast genome sequences of the Hawaiian mints may be exceptional is in their remarkably low levels of variability. Similar to most molecular phylogenetic studies at low taxonomic levels, previous analyses of the mints have largely been based on a small number of regions, including the *trnL* intron, *trnL-F*, *trnK* intron, and the *rps16* intron (Bräuchler et al., 2010; Roy et al., 2013), and have suffered from low resolution within the Hawaiian lineage. However, these commonly used regions have been shown to actually be among the least informative (Shaw et al., 2005). We compared variability of the complete chloroplast genomes of three *Stachys* and three Hawaiian species to investigate overall levels of variation and found that in the Hawaiian taxa, the low diversity of traditionally utilized phylogenetic markers mirrors the overall low diversity of the entire genome. Among the Hawaiian mint genera, only 104 high quality variable sites were detected (0.07%), compared to the 565 variable sites (0.4%) identified in three *Stachys* species.

We further examined the mint chloroplast genomes to determine if patterns of variability demonstrated by other lineages hold for the Hawaiian mints, but just at a lower scale. As typical in many other species, intergenic spacers demonstrated the highest number and percentage of variable sites in both *Stachys* and Hawaiian taxa, followed by introns, and finally by coding regions. A recent analysis of chloroplast genomes focusing on non-coding regions from 25, mostly congeneric species pairs suggests that a few regions are generally highly variable across most angiosperms, including *ndhF-rpl32*, *rpl32-trnL*, *ndhC-trnV*, and *5'rps16-trnQ* (Shaw et al., 2014). These regions have also been found to be highly variable in other studies using different sets of taxa and approaches (Byrne and Haninson, 2012; Dong et al., 2012; Shaw et al., 2007). All of these regions were moderately to highly variable among *Stachys* species. On the other hand, the most highly variable regions among the Hawaiian mints, in terms of the absolute number of substitutions, were *petA-psbJ*, *rpoB-trnC*, and *atpH-atpI*, while the most variable regions in terms of proportion of variable sites were *ycf3-trnS*, *rpl33-rps18*, *rpl16-rps3*, and *psbI-trnS*. None of these loci were particularly variable among the *Stachys* species, with the exception of *petA-psbJ*, which was moderately variable. The sliding window analysis suggested moderate agreement between the *Stachys* and the Hawaiian taxa regarding which consecutive areas of the chloroplast genome are most variable. Our results indicate that although the general patterns of variability identified for most angiosperms also hold true for the recent radiation of Hawaiian taxa, there is no single hyper-variable and only few moderately variable chloroplast regions of the mint genomes that would provide sufficient resolution to unequivocally identify phylogenetic relationships in this group. It is clear that a genomic approach is needed to obtain sufficient amounts of data.

4.3. Comprehensive chloroplast genome phylogeny of a mint radiation

We conducted an extensive plastid phylogenomic analysis encompassing nearly complete taxonomic sampling of the Hawaiian mints, plus a wide variety of close relatives and outgroup taxa from the genus *Stachys,* to investigate if a genomic approach can resolve maternal relationships among the Hawaiian mints and their close mainland *Stachys* relatives. Our approach adds considerable resolution to the Stachydeae phylogeny, similarly to trees based on ribosomal DNA data (Roy et al., 2013), and corroborates the findings of four well-supported clades comprising a clade of temperate NA *Stachys* (including a few East Asian *Stachys*), two Meso-SA *Stachys* clades, of which one also includes the Californian-Pacific northwestern species *Stachys chamissonis*, and a clade composed of the Hawaiian taxa (Fig. 6). A close relationship between the Hawaiian taxa and Meso-SA *Stachys* clades is strongly supported, as has been suggested previously with cpDNA data, confirming the cyto-nuclear discordance in this lineage (Fig. 1). Within each of the four main clades, resolution is remarkably improved compared to phylogenies based on many fewer cpDNA markers (Roy et al., 2013). For example, close relationships, also identified from nuclear ribosomal markers, are corroborated between large-flowered SA (*Stachys lamioides*) and Mexican (*Stachys coccinea* and *S. pacifica*) species in the Meso-SA clade II and between Eastern NA *Stachys* species (*S. hispida*, *S. latidens*, and *S. aspera*).

Although resolution within the Hawaiian lineage is enhanced, the three genera are not supported as monophyletic, and only a few strongly supported relationships can be discerned among individual species (Figs. 6 and 7). For example, we recovered a strongly supported clade comprising *Stenogyne rugosa*, *S. microphylla*, *S. angustifolia*, and purported hybrids between the two former species, relationships also reported with nuclear AFLP data (Lindqvist et al., 2003).

A second *S. rugosa* specimen (HI03-65) groups in a strongly supported clade with *S. sessilis*, *S. cranwelliae*, and *S. macrantha*, all species occurring on the island of Hawai'i. This specimen from the Kona District belongs to a distinct *S. rugosa* population described as unusually dense tomentose and has been attributed to a variety of *S. rugosa* (*S. rugosa* var. *mollis*) (Wagner et al., 1999). Our results suggest that this population deserves an elevated status as a new species. Although most suggested relationships based on morphological affinities are not supported here, other close interspecific relationships are confirmed, including that of *Phyllostegia helleri* and *P. electra* (Wagner, 1999), *P. mannii* and *P. racemosa*, and *S. haliakalae* and *S. rotundifolia* (Wagner et al., 1999). The latter two species group together in a strongly supported clade with *S. kamehamehae*, a species characterized by long and strongly falcate corollas, indicating a close relationship among these three species. We also find another close relationship of relatively large-flowered *Stenogyne*, including of *S. calaminthoides* and an accession of *S. macrantha*. A second accession of *S. macrantha* groups with *S. cranwelliae*, *S. sessilis*, and *S. rugosa*, questioning the circumscription of this taxon and highlighting the need for infraspecific assessments. Furthermore, *P. helleri* and *P. electra* belong in a strongly supported clade together with *P. waimeae* and *P. renovans*, indicating a close relationship among these species that all occur on Kaua'i, the oldest extant high island in the Hawaiian archipelago. Close relationships of numerous other species from Kaua'i are suggested, including of *P. kahiliensis* and *P. knudsenii*, as well as *S. kealiae*, *S. purpurea*, and *S. campanulata*. Similarly, several well-supported clades comprise species exclusively from the younger islands, Hawai'i and the former Maui Nui complex (Maui, Moloka'i, and Lana'i), suggesting several cases of intra-island diversifications. The interspersing of species from Kaua'i among species from O'ahu and the younger Hawaiian islands indicate a complex biogeographic history, possibly involving multiple forward and backward dispersal events, a finding also shown with nuclear ribosomal markers (Lindqvist et al., 2003).

5. Conclusions

In large and recent radiations that also display interspecific gene flow and complex diversification patterns, such as the Hawaiian mint lineage, fully resolved uniparental evolutionary histories are important components in dissecting their hybrid ancestries. Our study demonstrates that a plastid phylogenomic approach will increase phylogenetic resolution for recently diverged lineages that are poorly resolved with smaller datasets, including those relying on traditionally utilized cpDNA markers. Our results also illuminate, however, the challenges of resolving relationships within such recent radiations, even when massive datasets are available. An approach to sequence complete chloroplast genomes, instead of targeting variable regions, can possibly capture even more sequence diversity among the Hawaiian mints but is unlikely to generate a fully resolved maternal phylogeny due to the evident overall lack of variation. Similarly, phylogenetic reconstruction of ribosomal and low-copy nuclear biparental markers have demonstrated remarkably poor resolution within the Hawaiian lineage, and incongruence between individual gene trees (Roy et al., 2013; Roy et al., 2015). Intriguingly, low-copy nuclear gene trees have shown the Hawaiian species to be polyphyletic, interspersed with either temperate North American or Meso-South American *Stachys*, while a multilocus coalescence analysis concur with nrDNA results placing them within the temperate North American *Stachys* clade (Roy et al., 2015). Neither of these analyses, including the ones presented here, has been able to resolve the currently recognized generic boundaries of the Hawaiian mints, questioning the taxonomic circumscription of this lineage and the interrelationships of its members. A

phylogenomic approach targeting thousands of independently evolved nuclear loci may be the only way to assess the current taxonomy of this group, and dissect the roles of hybridization and introgression in the origin and rapid diversification of this insular lineage. Phylogenomic methods based on markers identified from, e.g., restriction site associated DNA sequencing (RADseq) (Cruaud et al., 2014; Eaton and Ree, 2013; Smith et al., 2014), targeted enrichment of ultraconserved elements (UCEs) or conserved ortholog sequences (COS) (Mandel et al., 2014; Smith et al., 2014), and RNA-Seq (Wen et al., 2013; Wickett et al., 2014) could be suitable methods to resolve Hawaiian mint relationships. However, these approaches are also associated with challenges, including accurately inferring orthology in high polyploid plants, slow mutation rates in exonic regions, which may be too low to resolve recent radiations (Giarla and Esselstyn, 2015), and the need to obtaining high quality RNA and DNA. Overcoming these challenges and utilizing both uniparental and multilocus nuclear data sets may be the best way forward. These combined with the powerful approach of conducting extensive taxonomic sampling of rare, endangered, and extinct taxa from herbarium collections will contribute to a better understanding of the underlying processes that drive recent radiation in this and other plant lineages.

Acknowledgements

We thank the Bernice P. Bishop Museum, The National Tropical Botanic Garden, the Volcano Rare Plant Facility, and numerous herbaria for access to samples. We are also grateful to Tilo Roy and Victor Albert for helpful discussions, Webb Miller for advice on data analysis, and Gerald Carr for use of the Hawaiian mint image. We thank the Center for Computational Resources at the University of Buffalo for access to computational resources used in analyses. This work was supported by the College of Arts and Sciences, University at Buffalo.

Supplementary Materials

Supplementary Figs. 1-7 and Supplementary Table 1. Additional data are published in the companion paper, Welch et al., 2016.

References

Allaby, R., Gutaker, R., Clarke, A., Pearson, N., Ware, R., Palmer, S.A., Kitchen, J.L., Smith, O., 2015. Using archaeogenomic and computational appraoches to unravel the history of local adapation in crops. Phil. Trans. R. Soc. Lond. B. Biol. Sci. 370, 20130377.

Appelhans, M.S., Wen, J., Wood, K.R., Allan, G.J., Zimmer, E.A., Wagner, W.L., 2014. Molecular phylogenetic analysis of Hawaiian Rutaceae (*Melicope, Platydesma* and *Zanthoxylum*) and their different colonization patterns. Bot. J. Linn. Soc. 174, 425-448.

Ávila-Arcos, M.C., Sandoval-Velasco, M., Schroeder, H., Carpenter, M.L., Malaspinas, A.-S., Wales, N., Peñaloza, F., Bustamante, C.D., Gilbert, M.T.P., 2015. Comparative performance of two whole-genome capture methodologies on ancient DNA Illumina libraries. Methods in Ecology and Evolution 6, 725-734.

Bakker, F.T., Lei, D., Yu, J., Mohammadin, S., Wei, Z., van de Kerke, S., Gravendeel, B., Nieuwenhuis, M., Staats, M., Alquezar-Planas, D.E., Holmer, R., 2015. Herbarium

genomics: plastome sequence assembly from a range of herbarium specimens using an iterative organelle genome assembly pipeline. Biol. J. Linn. Soc. 117, 33-43.
Baldwin, B.G., Markos, S., 1998. Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: Congruence of ETS and ITS trees of *Calycadenia* (Compositae). Mol. Phylogen. Evol. 10, 449-463.
Baldwin, B.G., Sanderson, M.J., 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proc. Natl. Acad. Sci. 95, 9402-9406.
Baldwin, B.G., Wagner, W.L., 2010. Hawaiian angiosperm radiations of North American origin. Ann. Bot. 105, 849-879.
Bandelt, H.J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16, 37-48.
Beck, J.B., Semple, J.C., 2015. Next-generation sampling: Pairing genomes with herbarium specimens provides species-level signal in *Solidago* (Asteraceae). Appl. Plant Sci. 3, 1500014. DOI: 10.3732/apps.1500014.
Bendiksby, M., Thorbek, L., Scheen, A.-C., Lindqvist, C., Ryding, O., 2011. An updated phylogeny and classification of Lamiaceae subfamily Lamioideae. Taxon 60, 471-484.
Besnard, G., Christin, P.-A., Malé, P.-J.G., Lhuillier, E., Lauzeral, C., Coissac, E., Vorontsova, M.S., 2014. From museums to genomics: old herbarium specimens shed light on a C3 to C4 transition. J. Exp. Bot. 65, 6711-6721.
Bräuchler, C., Meimberg, H., Heubl, G., 2010. Molecular phylogeny of Menthinae (Lamiaceae, Nepetoideae, Mentheae) - Taxonomy, biogeography and conflicts. Mol. Phylogen. Evol. 55, 501-523.
Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., Pääbo, S., 2007. Patterns of damamge in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. 104, 14616-14621.
Byrne, M., Haninson, M., 2012. Testing the variability of chloroplast sequences for plant phylogeography. Aust. J. Bot. 60, 569-574.
Cronk, Q.C.B., Kiehn, M., Wagner, W.L., Smith, J.F., 2005. Evolution of Cyrtandra (Gesneriaceae) in the Pacific Ocean: The origin of a suptertramp clade. Am. J. Bot. 92, 1017-1024.
Cronn, R., Liston, A., Parks, M., Gernandt, D.S., Shen, R., Mockler, T., 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Res. 36, e122.
Cronn, R., Knaus, B.J., Liston, A., Maughan, P.J., Parks, M., Syring, J.V., Udall, J., 2012. Targeted enrichment strategies for next-generation plant biology. Am. J. Bot. 99, 291-311.
Cruaud, A., Gautier, M., Galan, M., Foucaud, J., Sauné, L., Genson, G., Dubois, E., Nidelet, S., Deuve, T., Rasplus, J.C.Y., 2014. Empirical assessment of RAD sequencing for interspecific phylogeny. Mol. Biol. Evol. 31, 1271-1274.
Darling, A.E., Mau, B., Perna, N.T., 2010. progressiveMauve: Multiple genome alignment with gene gain, loss, and rearrangement. PLoS One 5, e11147. DOI: 10.1371/journal.pone.0011147.
Dong, W., Liu, J., Yu, J., Wang, L., Zhou, S., 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS One 7, e35071. DOI: 10.1371/journal.pone.0035071.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214. DOI: 10.1186/1471-2148-7-214.
Eaton, D.A., Ree, R.H., 2013. Inferring phylogeny and introgression using RADseq data: An example from flowering plants (Pedicularis: Orobanchaceae). Syst. Biol. 62, 689-706.
Ekblom, R., Galindo, J., 2011. Applications of next-generation sequencing in molecular ecology of non-model organisms. Heredity 107, 1-15.
Emerson, B.C., 2002. Evolution on oceanic islands: Molecular phylogenetic approaches to understanding pattern and process. Mol. Ecol. 11, 951-966.
Giarla, T.C., Esselstyn, J.A., 2015. The challenges of resolving a rapid, recent radiation: Empirical and simulated phylogenomics of Philippine shrews. Syst. Biol. 64, 727-740.
Gillespie, R.G., Croom, H.B., Palumbi, S.R., 1994. Multiple origins of a spider radiation in Hawaii. Proc. Natl. Acad. Sci. 91, 2290-2294.
Givnish, T.J., Millam, K.C., Mast, A.R., Paterson, T.B., Theim, T.J., Hipp, A.L., Henss, J.M., Smith, J.F., Wood, K.R., Sytsma, K.J., 2009. Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). Proc. R. Soc. B. 276, 407-416.
Givnish, T.J., Bean, G.J., Ames, M., Lyon, S.P., Sytsma, K.J., 2013. Phylogeny, floral evolution, and inter-island dispersal in Hawaiian *Clermontia* (Lobeliaceae) based on ISSR variation and plastid spacer sequences. PLoS One 8, e62566. DOI: 10.1371/journal.pone.0062566.
Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254-267.
Jaenicke-Despres, V., Buckler, E.S., Smith, B.D., Gilbert, M.T.P., Cooper, A., Doebley, J., Paabo, S., 2003. Early allelic selection in maize as revealed by ancient DNA. Science 302, 1206-1208.
Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., Orlando, L., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 13, 1682-1684.
Kambysellis, M.P., Ho, K.-F., Craddock, E.M., Piano, F., Parisi, M., Cohen, J., 1995. Pattern of ecological shifts in the diversification of Hawaiian *Drosophila* inferred from a molecular phylogeny. Curr. Biol. 5, 1129-1139.
Katoh, K., Misawa, K., Kuma, K.-I., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059-3066.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772-780.
Knapp, M., Hofreiter, M., 2010. Next generation sequencing of ancient DNA: Requirements, strategies and perspectives. Genes 1, 227-243.
Krause, J., Dear, P.H., Pollack, J.L., Slatkin, M., Spriggs, H., Barnes, I., Lister, A.M., Ebersberger, I., Pääbo, S., Hofreiter, M., 2006. Miltiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. Nature 439, 724-727.
Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009. Circos: An information aesthetic for comparative genomics. Genome Res. 19, 1639-1645.
Lee, H.-L., Jansen, R.K., Chumley, T.W., Kim, K.-J., 2007. Gene relocations within chloroplast genomes of *Jasminum* and *Mendora* (Oleaceae) are due to multiple, overlapping inversions. Mol. Biol. Evol. 24, 1161-1180.

Lerner, H.R.L., Meyer, M., James, H.F., Hofreiter, M., Fleischer, R.C., 2011. Multilocus resolution of phylogengy and timescale in the extant adaptive radiation of Hawaiian honeycreepers. Curr. Biol. 21, 1-7.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25, 2078-2079.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., Wang, J., 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 20, 265-272.

Lindgreen, S., 2012. AdapterRemoval: Easy cleaning of next-generation sequencing reads. BMC Res. Notes 5, 337. DOI: 10.1186/1756-0500-5-337.

Lindqvist, C., Albert, V., 2002. Origin of the Hawaiian endemic mints within North American *Stachys* (Lamiaceae). Am. J. Bot. 89, 1709-1724.

Lindqvist, C., Motley, T.J., Jeffrey, J.J., Albert, V., 2003. Cladogenesis and reticulation in the Hawaiian endemic mints (Lamiaceae). Cladistics 19, 480-495.

Lukas, B., Novak, J., 2013. The complete chloroplast genome of *Origanum vulgare* L. (Lamiaceae). Gene 528, 163-169.

Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., Michelmore, R.W., Rieseberg, L.H., Burke, J.M., 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the compositae. Appl. Plant Sci. 2, 1300085. 10.3732/apps.1300085.

Mariotti, R., Cultrera, N.G.M., Díez, C.M., Baldoni, L., Rubini, A., 2010. Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. BMC Plant Biol. 10, 211.

Martin, W., Deusch, O., Stawski, N., Grünheit, N., Goremykin, V., 2005. Chloroplast genome phylogenetics: why we need independent approaches to plant molecular evolution. Trends Plant Sci. 10, 203-209.

Miller, J.R., Koren, S., Sutton, G., 2010a. Assembly algorithms for next-generation sequencing data. Genomics 95, 315-327.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010b. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA.

Miller, W., Drautz, D.I., Ratan, A., Pusey, B., Qi, J., Lesk, A.M., Tomsho, L.P., Packard, M.D., Zhao, F., Sher, A., Tikhonov, A., Raney, B., Patterson, N., Lindblad-Toh, K., Lander, E.S., Knight, J.R., Irzyk, G.P., Fredrikson, K.M., Harkins, T.T., Sheridan, S., Pringle, T., Schuster, S.C., 2008. Sequencing the nuclear genome of the extinct woolly mammoth. Nature 456, 387-390.

Milne, I., Stephen, G., Bayer, M., Cock, P.J.A., Pritchard, L., Cardle, L., Shaw, P.D., Marshall, D., 2013. Using Tablet for visual exploration of second-generation sequencing data. Brief. Bioinformatics 14, 193-202.

Moore, M.J., Dhingra, A., Soltis, P.S., Shaw, R., Farmerie, W.G., Folta, K.M., Soltis, D.E., 2006. Rapid and accurate pyrosequencing of angiosperm plastid genomes. BMC Plant Biol. 6, 17. DOI: 10.1186/1471-2229-6-17.

Noonan, J.P., Hofreiter, M., Smith, D., Priest, J.R., Rohland, N., Rabeder, G., Krause, J., Detter, J.C., Paabo, S., Rubin, E.M., 2005. Genomic sequencing of Pleistocene cave bears. Science 309, 597-600.

Parks, M., Cronn, R., Liston, A., 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 7, 84. DOI: 10.1186/1741-7007-7-84.

Peakall, R., Smouse, P.E., 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic softward for reaching and research - an update. Bioinformatics 28, 2537-2539.

Pruvost, M., Schwarz, R., Correia, V.B., Champlot, S., Braguier, S., Morel, N., Fernandez-Jalvo, Y., Grange, T., Geigl, E.-M., 2007. Freshly excavated fossil bones are best for amplification of ancient DNA. Proc. Natl. Acad. Sci. 104, 739-744.

Qian, J., Song, J., Gao, H., Zhu, Y., Xu, J., Pang, X., Yao, H., Sun, C., Li, X.e., Li, C., Liu, J., Xu, H., Chen, S., 2013. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. PLoS One 8, e57607. DOI: 10.1371/journal.pone.0057607.

Rambaut, A., Ho, S.Y.W., Drummond, A.J., Shapiro, B., 2009. Accommodating the effect of ancient DNA damage on inferences of demographic histories. Mol. Biol. Evol. 26, 245-248.

Ratan, A., 2009. Assembly algorithms for next-generation sequence data. Ph.D. dissertation. Pennsylvania State University, University Park, PA.

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

Roy, T., Chang, T.-H., Lan, T., Lindqvist, C., 2013. Phylogeny and biogeography of New World Stachydeae (Lamiaceae) with emphasis on the origin and diversification on Hawaiian and South American taxa. Mol. Phylogen. Evol. 69, 218-238.

Roy, T., Cole, L.W., Chang, T.-H., Lindqvist, C., 2015. Untangling reticulate evolutionary relationships among New World and Hawaiian mints (Stachydeae, Lamiaceae). Mol. Phylogen. Evol. 89, 46-62.

Rozen, S., Skaletsky, H.J., 2000. Primer3 on the WWW for general users and biologist programmers. In: Krawetz, S., Misener, S. (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 365-386.

Ruffalo, M., LaFramboise, T., Koyutürk, M., 2011. Comparative analysis of algorithms for next-generation sequencing read alignment. Bioinformatics 27, 2790-2796.

Salmaki, Y., Zarre, S., Ryding, O., Lindqvist, C., Brauchler, C., Heubl, G., Barber, J., Bendiksby, M., 2013. Molecular phylogeny of tribe Stachydeae (Lamiaceae subfamily Lamioideae). Mol. Phylogen. Evol. 69, 535-551.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., Pääbo, S., 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS One 7, e34131. DOI: 10.1371/journal.pone.0034131.

Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33, W686-689.

Scheen, A.-C., Albert, V., 2009. Molecular phylogenetics of the *Leucas* group (Lamioideae; Lamiaceae). Syst. Bot. 34, 173-181.

Scheen, A.-C., Bendiksby, M., Ryding, O., Mathiesen, C., Albert, V., Lindqvist, C., 2010. Molecular phylogenetics, character evolution and suprageneric classification of Lamioideae (Lamiaceae). Ann. Mo. Bot. Gard. 97, 191-217.

Schuenemann, V.J., Bos, K., Dewitte, S., Schmedes, S., Jamieson, J., Mittnik, A., Forrest, S., Coombes, B.K., Wood, J.W., Earn, D.J.D., White, W., Krause, J., Poinar, H.N., 2011. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. Proc. Natl. Acad. Sci. 108, E746-E752.

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

Shaw, J., Lickey, E.B., Schilling, E.E., Small, R.L., 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. Am. J. Bot. 94, 275-288.

Shaw, J., Shafer, H.L., Leonard, O.R., Kovach, M.J., Schorr, M., Morris, A.B., 2014. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. Am. J. Bot. 101, 1987-2004.

Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., Wendel, J.F., 1998. The tortoise and the hare: Choosing between noncoding plastome and nuclear ADH sequences for phylogney reconstruction in a recently diverged plant group. Am. J. Bot. 85, 1301-1315.

Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C., Brumfield, R.T., 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Syst. Biol. 63, 83-95.

Soltis, D.E., Albert, V.A., Savolainen, V., Hilu, K., Qiu, Y.-L., Chase, M.W., Farris, J.S., Stefanović, S., Rice, D.W., Palmer, J.D., Soltis, P.S., 2004. Genome-scale data, angiosperm relationships, and 'ending incongruence': A cautionary tale in phylogenetics. Trends Plant Sci. 9, 477-483.

Staats, M., Cuenca, A., Richardson, J.E., Vrielink-van Ginkel, R., Petersen, G., Seberg, O., Bakker, F.T., 2011. DNA damage in plant herbarium tissue. PLoS One 6, e28448. DOI: 10.1371/journal.pone.0028448.

Staats, M., Erkens, R.H.J., van de Vossenberg, B., Wieringa, J.J., Kraaijeveld, K., Stielow, B., Geml, J., Richardson, J.E., Bakker, F.T., 2013. Genomic treasure troves: Complete genome sequencing of herbarium and insect museum specimens. PLoS One 8, e69189. DOI: 10.1371/journal.pone.0069189.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web-servers. Syst. Biol. 75, 758-771.

Stull, G.W., Moore, M.J., Mandala, V.S., Douglas, N.A., Kates, H.-R., Qi, X., Brockington, S.F., Soltis, P.S., Soltis, D.E., Gitzendanner, M.A., 2013. A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. Appl. Plant Sci. 1, 1200497. DOI: 10.3732/apps.1200497.

Wagner, W.L., 1999. Nomenclator and review of *Phyllostegia* (Lamiaceae). Novon 9, 265-279.

Wagner, W.L., Herbst, D.R., Sohmer, S.H., 1999. Manual of the flowering plants of Hawaii, revised edition. University of Hawaii Press, Honolulu, Hawaii, USA.

Warburton, P.E., Giordano, J., Cheung, F., Gelfand, Y., Benson, G., 2004. Inverted repeat structure of the human genome: The X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. Genome Res. 14, 1861-1869.

Weiss, C.L., Schuenemann, V.J., Devos, J., Shirsekar, G., Reiter, E., Gould, B.A., Stinchcombe, J.R., Krause, J., Burbano, H.A., 2015. Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. bioRxiv. DOI: 10.1101/023135.

Welch, A.J., Collins, K., Ratan, A., Drautz-Moses, D.I., Schuster, S.C., Lindqvist, C., 2016. Data characterizing the chloroplast genomes of extinct and endangered Hawaiian endemic mints (Lamiaceae) and their close relatives. Data in Brief In Press.

Wen, J., Xiong, Z., Nie, Z.L., Mao, L., Zhu, Y., Kan, X.-Z., Ickert-Bond, S.M., Gerrath, J., Zimmer, E.A., Fang, X.-D., 2013. Transcriptome sequences resolve deep relationships of the grape family. PLoS One 8, e74394. 10.1371/journal.pone.0074394.

Wicke, S., Müller, K.F., dePamphilis, C.W., Quandt, D., Wickett, N.J., Zhang, Y., Renner, S.S., Schneeweiss, G.M., 2013. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. Plant Cell 25, 3711-3725.

Wicke, S., Schäferhoff, B., dePamphilis, C.W., Müller, K.F., 2014. Disproportional plastome-wide increase of substitution rates and relaxed purifying selection in genes of carnivorous Lentibulariaceae. Mol. Biol. Evol. 31, 529-545.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., Ruhfel, B.R., Wafula, E., Der, J.P., Graham, S.W., Mathews, S., Melkonian, M., Soltis, D.E., Soltis, P.S., Miles, N.W., Rothfels, C.J., Pokorny, L., Shaw, A.J., DeGironimo, L., Stevenson, D.W., Surek, B., Villarreal, J.C., Roure, B., Philippe, H., dePamphilis, C.W., Chen, T., Deyholos, M.K., Baucom, R.S., Kutchan, T.M., Augustin, M.M., Wang, J., Zhang, Y., Tian, Z., Yan, Z., Wu, X., Sun, X., Wong, G.K.-S., Leebens-Mack, J., 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. 111, E4859-E4868. 10.1073/pnas.1323926111.

Witter, M.S., Carr, G.D., 1988. Adaptive radiation and genetic differentiation in the Hawaiian silversword alliance (Compositae: Madiinae). Evolution 42, 1278-1287.

Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20, 3252-5255.

Yi, D.-K., Lee, H.-L., Sun, B.-Y., Chung, M.Y., Kim, K.-J., 2012. The complete chloroplast DNA sequence of *Eleutherococcus senticosus* (Araliaceae); Comparative evolutionary analyses with other three asterids. Mol. Cells 33, 497-508.

You, F.M., Huo, N., Gu, Y.Q., Luo, M.-C., Ma, Y., Hane, D., Lazo, G.R., Dvorak, J., Anderson, O.D., 2008. BatchPrimer3: A high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9, 253. DOI: 10.1186/1471-2105-9-253.

Zedane, L., Hong-Wa, C., Murienne, J., Jeziorski, C., Baldwin, B.G., Besnard, G., 2015. Museomics illuminate the history of an extinct, paleoendemic plant lineage (Hesperelaea, Oleaceae) known from an 1875 collection from Guadalupe Island, Mexico. Biol. J. Linn. Soc. 117, 44-57.

Zhang, H., Li, C., Miao, H., Xiong, S., 2013. Insights from the complete chloroplast genome into the evolution of *Sesamum indicum* L. PLoS One 8, e80508. DOI: 10.1371/journal.pone.0080508.

Figures

Figure 1. Schematic phylogenetic trees demonstrating the position of the Hawaiian mints among close continental *Stachys* relatives in the Americas based on A) chloroplast DNA, B) nuclear ribosomal DNA, and C) low-copy nuclear DNA. SA=South America; NA=North America

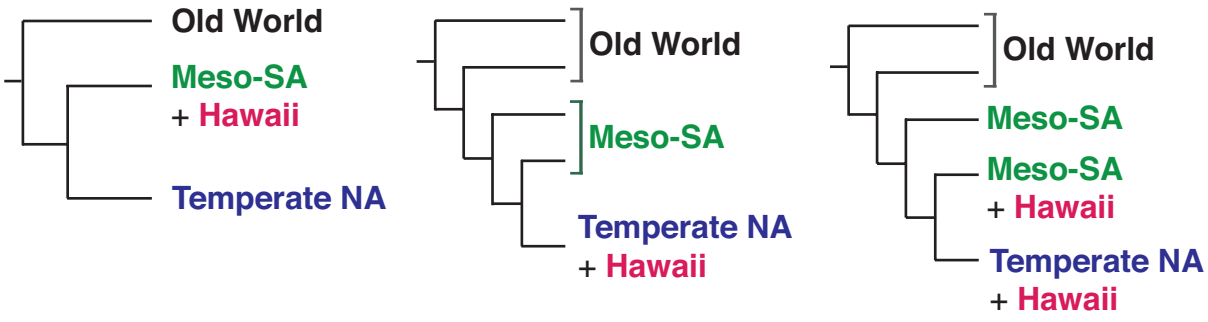

Figure 2. Diagram of the chloroplast genome of the endemic Hawaiian mint *Stenogyne haliakalae*. The outer black ring shows locations of genes, the next ring shows locations of regions sequenced in 96 additional accessions in blue, and the following rings (moving inward) show locations of SNPs for *Stenogyne bifida*, *Haplostachys haplostachya, Stachys chamissonis, Stachys coccinea, and Stachys sylvatica,* respectively. The inner black ring shows the locations of the large and small single copy (LSC and SSC) and inverted repeat (IRa and IRb) regions, as well as the GC content in gray.

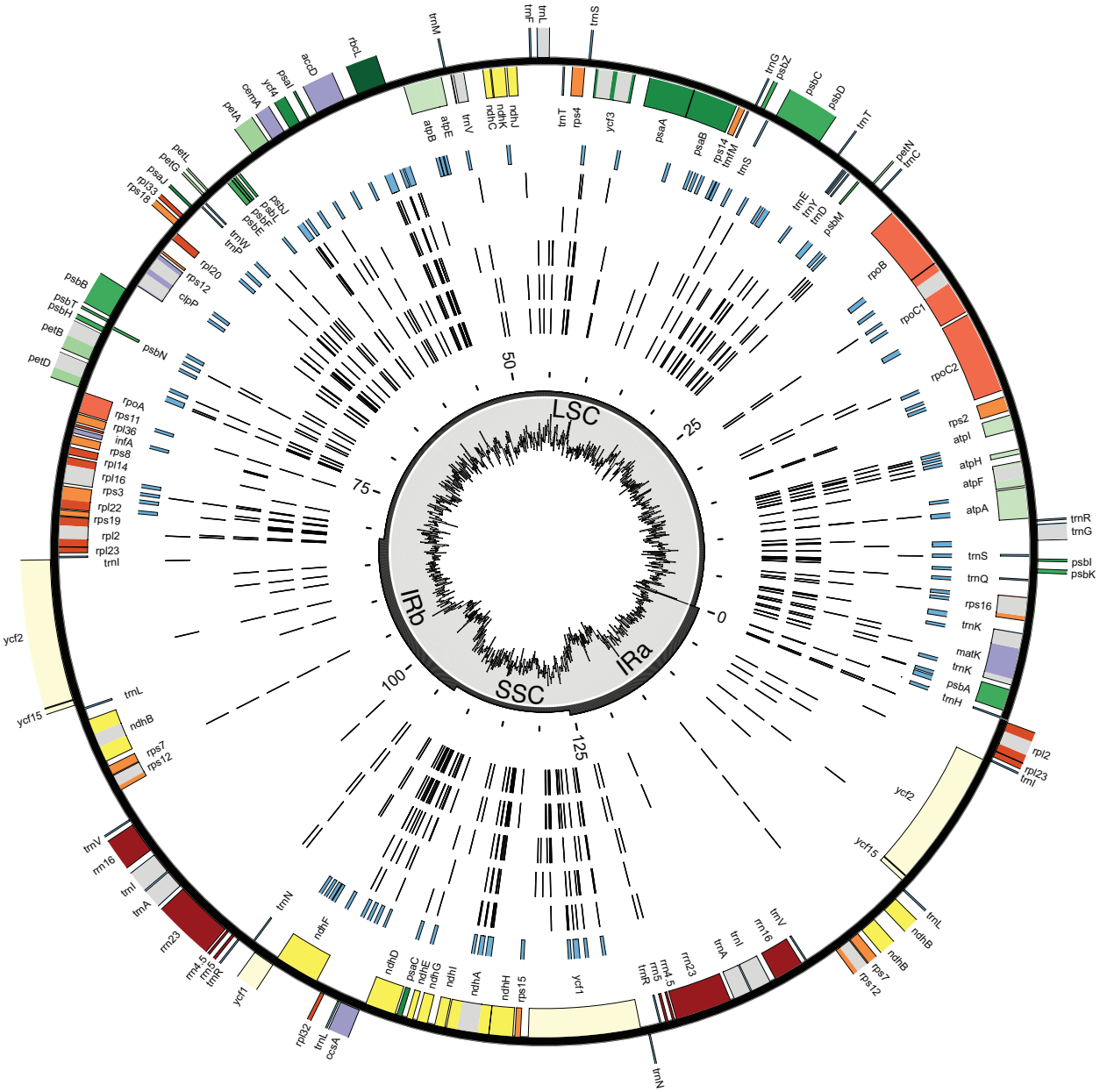

Figure 3. Conservation of chloroplast genomes from five Lamiales taxa. A physical map next to each genome shows content and organization, with regions colored according to function. Similarity plots are given next to the physical map. Regions that are conserved among all species are colored mauve, while those that are conserved among subsets of the taxa have different colors. Gaps in the plot indicate regions that could not be aligned and may be lineage specific. Taxa are arranged in approximate order of phylogenetic distance from the Hawaiian lineage (see Fig. 4).

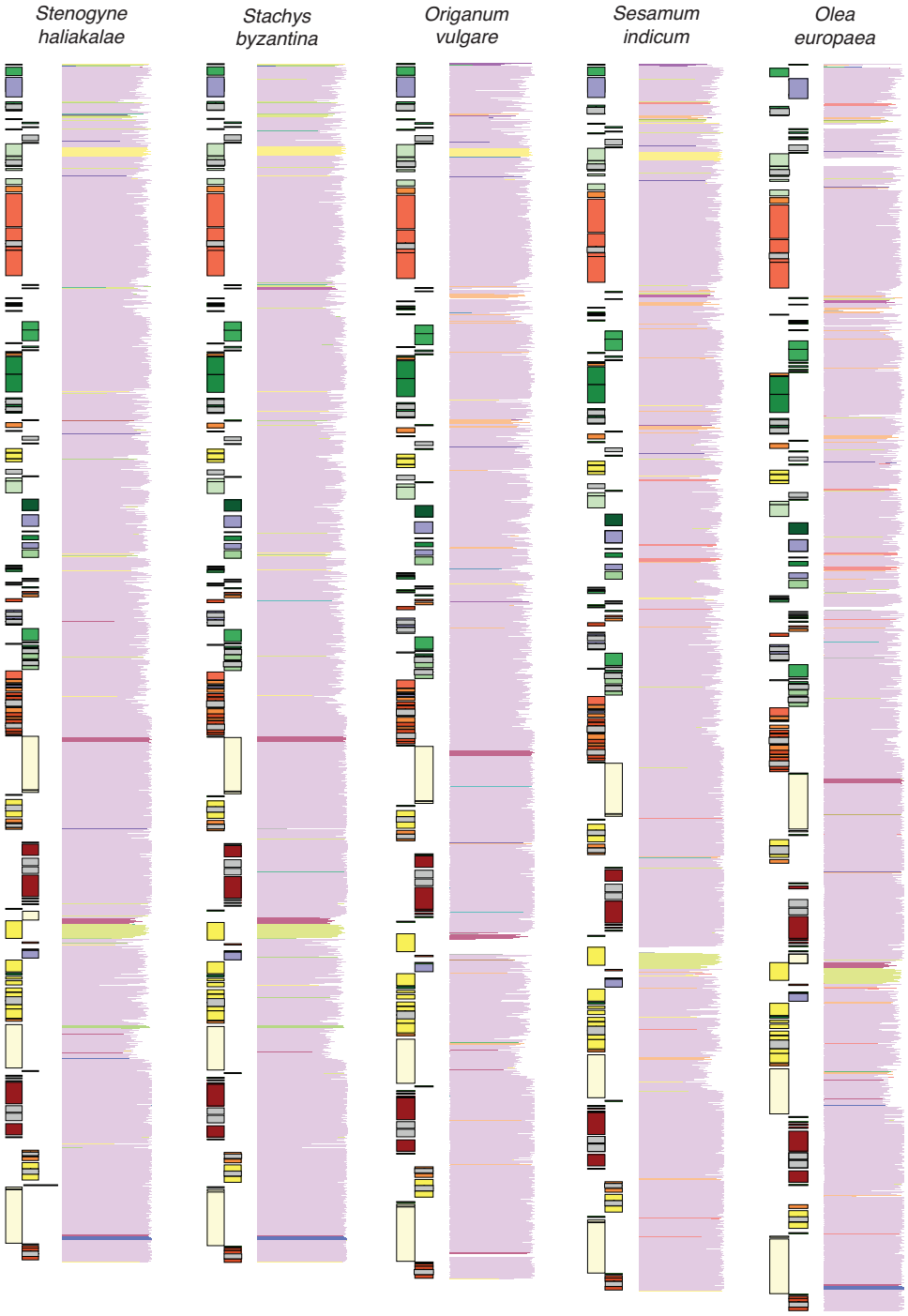

Figure 4. Phylogeny of complete Lamiales chloroplast genomes. Bootstrap support (circles) and posterior probabilities (squares) are given for all nodes outside of the Hawaiian lineage, with black squares and circles indicating support > 90 and 0.95, respectively. Both maximum likelihood and Bayesian inference support the monophyletic grouping of the Hawaiian mints (100 and 1.0, respectively) and their close relationship with *Stachys* taxa grouping within meso-South American (SA) clades (see Fig. 6). See supplementary Figure 4 for a closer view of the Hawaiian clade.

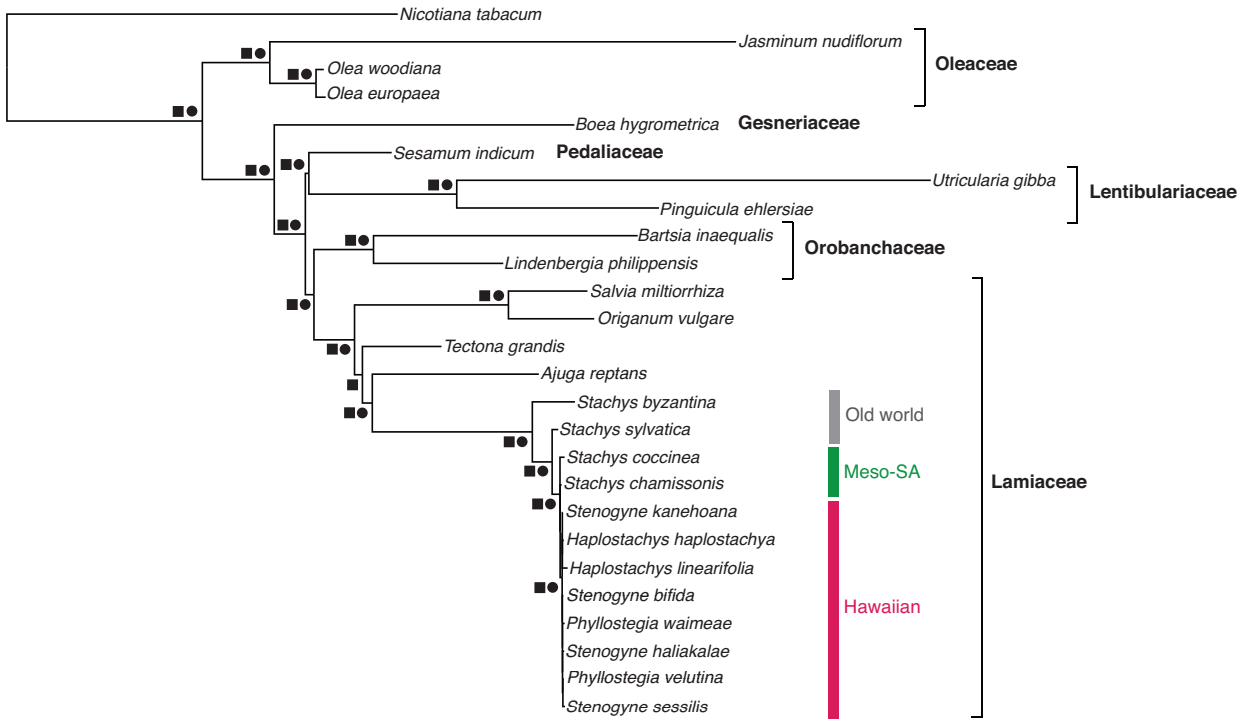

Figure 5. Comparison of variability in non-overlapping 1000 bp sliding windows for six high quality chloroplast genomes (*Stenogyne haliakalae, Stenogyne bifida, Haplostachys Haplostachya, Stachys chamissonis, Stachys coccinea,* and *Stachys sylvatica*). Pink = three Hawaiian taxa; Black = three *Stachys* taxa. PIC = Potentially informative characters. LSC = Large single copy, IR = Inverted repeat, SSC = Small single copy region.

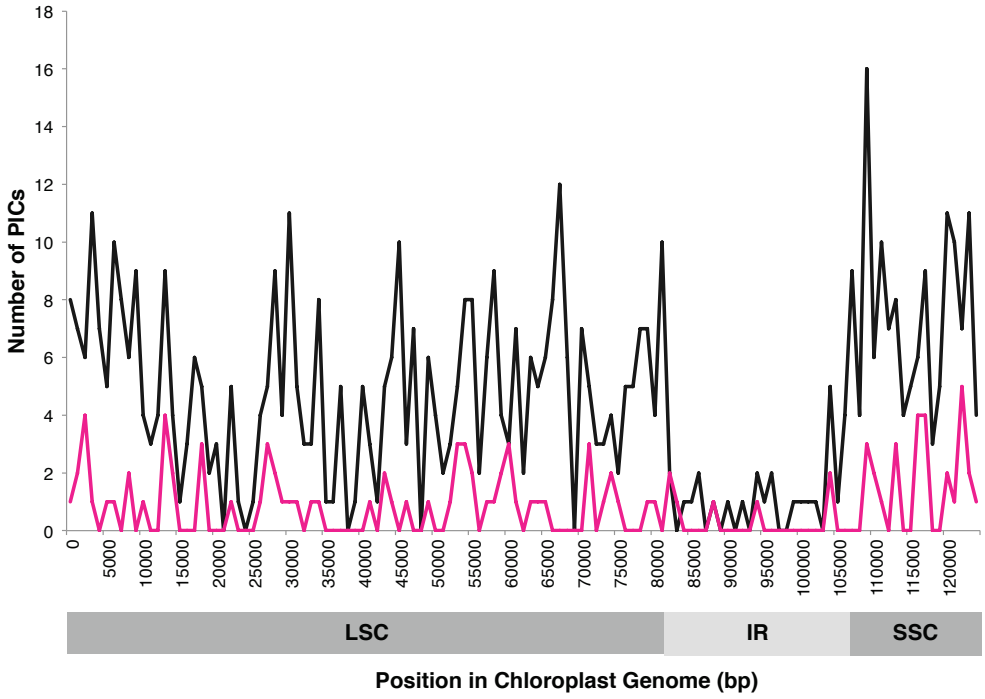

Figure 6. Chloroplast phylogeny of the Hawaiian mint radiation and *Stachys* relatives constructed from a partitioned analysis (coding vs non-coding) of ~20,000 bp targeting the most variable regions of the plastid genome (see Fig. 2). Support from 1000 maximum likelihood bootstrap replicates is indicated by circles with black = 90 – 100, gray = 80-89, white = 70 – 79. Posterior probability from Bayesian inference is indicated by squares with black = 0.95 – 1.0, gray = 0.90 – 0.94, white = 0.80 – 0.89. In the Hawaiian lineage, *Phyllostegia* species are indicated in purple, *Stenogyne* in orange, and *Haplostachys* in teal. For taxa with multiple accessions, the accession number is indicated in parentheses (see Table 2). *indicate taxa for which full chloroplast genomes were sequenced.

[Figure on next page]

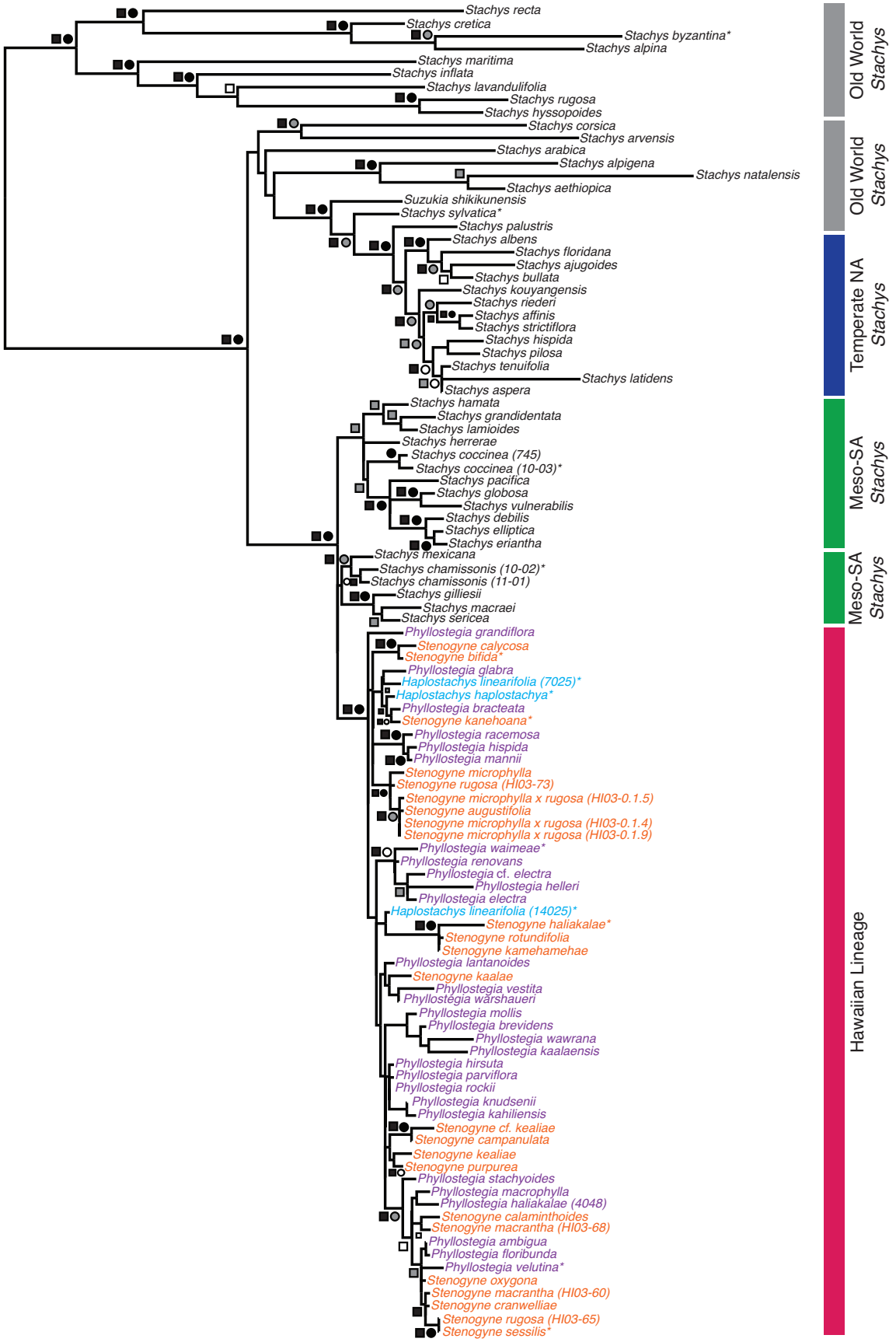
Stachys recta
Stachys cretica
Stachys byzantina*
Stachys alpina
Stachys maritima
Stachys inflata
Stachys lavandulifolia
Stachys rugosa
Stachys hyssopoides
Stachys corsica
Stachys arvensis
Stachys arabica
Stachys alpigena
Stachys natalensis
Stachys aethiopica
Suzukia shikikunensis
Stachys sylvatica*
Stachys palustris
Stachys albens
Stachys floridana
Stachys ajugoides
Stachys bullata
Stachys kouyangensis
Stachys riederi
Stachys affinis
Stachys strictiflora
Stachys hispida
Stachys pilosa
Stachys tenuifolia
Stachys latidens
Stachys aspera
Stachys hamata
Stachys grandidentata
Stachys lamioides
Stachys herrerae
Stachys coccinea (745)
Stachys coccinea (10-03)*
Stachys pacifica
Stachys globosa
Stachys vulnerabilis
Stachys debilis
Stachys elliptica
Stachys eriantha
Stachys mexicana
Stachys chamissonis (10-02)*
Stachys chamissonis (11-01)
Stachys gilliesii
Stachys macraei
Stachys sericea
Phyllostegia grandiflora
Stenogyne calycosa
Stenogyne bifida*
Phyllostegia glabra
Haplostachys linearifolia (7025)*
Haplostachys haplostachya*
Phyllostegia bracteata
Stenogyne kanehoana*
Phyllostegia racemosa
Phyllostegia hispida
Phyllostegia mannii
Stenogyne microphylla
Stenogyne rugosa (HI03-73)
Stenogyne microphylla x rugosa (HI03-0.1.5)
Stenogyne augustifolia
Stenogyne microphylla x rugosa (HI03-0.1.4)
Stenogyne microphylla x rugosa (HI03-0.1.9)
Phyllostegia waimeae*
Phyllostegia renovans
Phyllostegia cf. electra
Phyllostegia helleri
Phyllostegia electra
Haplostachys linearifolia (14025)*
Stenogyne haliakalae*
Stenogyne rotundifolia
Stenogyne kamehamehae
Phyllostegia lantanoides
Stenogyne kaalae
Phyllostegia vestita
Phyllostegia warshaueri
Phyllostegia mollis
Phyllostegia brevidens
Phyllostegia wawrana
Phyllostegia kaalaensis
Phyllostegia hirsuta
Phyllostegia parviflora
Phyllostegia rockii
Phyllostegia knudsenii
Phyllostegia kahiliensis
Stenogyne cf. kealiae
Stenogyne campanulata
Stenogyne kealiae
Stenogyne purpurea
Phyllostegia stachyoides
Phyllostegia macrophylla
Phyllostegia haliakalae (4048)
Stenogyne calaminthoides
Stenogyne macrantha (HI03-68)
Phyllostegia ambigua
Phyllostegia floribunda
Phyllostegia velutina*
Stenogyne oxygona
Stenogyne macrantha (HI03-60)
Stenogyne cranwelliae
Stenogyne rugosa (HI03-65)
Stenogyne sessilis*
Old World Stachys
Old World Stachys
Temperate NA Stachys
Meso-SA Stachys
Meso-SA Stachys
Hawaiian Lineage

Figure 7. Median-joining network demonstrating reticulate relationships among the Hawaiian mints. Observed haplotypes are colored by genus and represented as circles, with the size of the circle being proportional to the number of individuals with that haplotype. Length of edges is related to the number of changes between the taxa. Dashed lines indicate names of taxa for internal nodes. For taxa with multiple accessions, the accession number is indicated in parentheses (see Table 2).

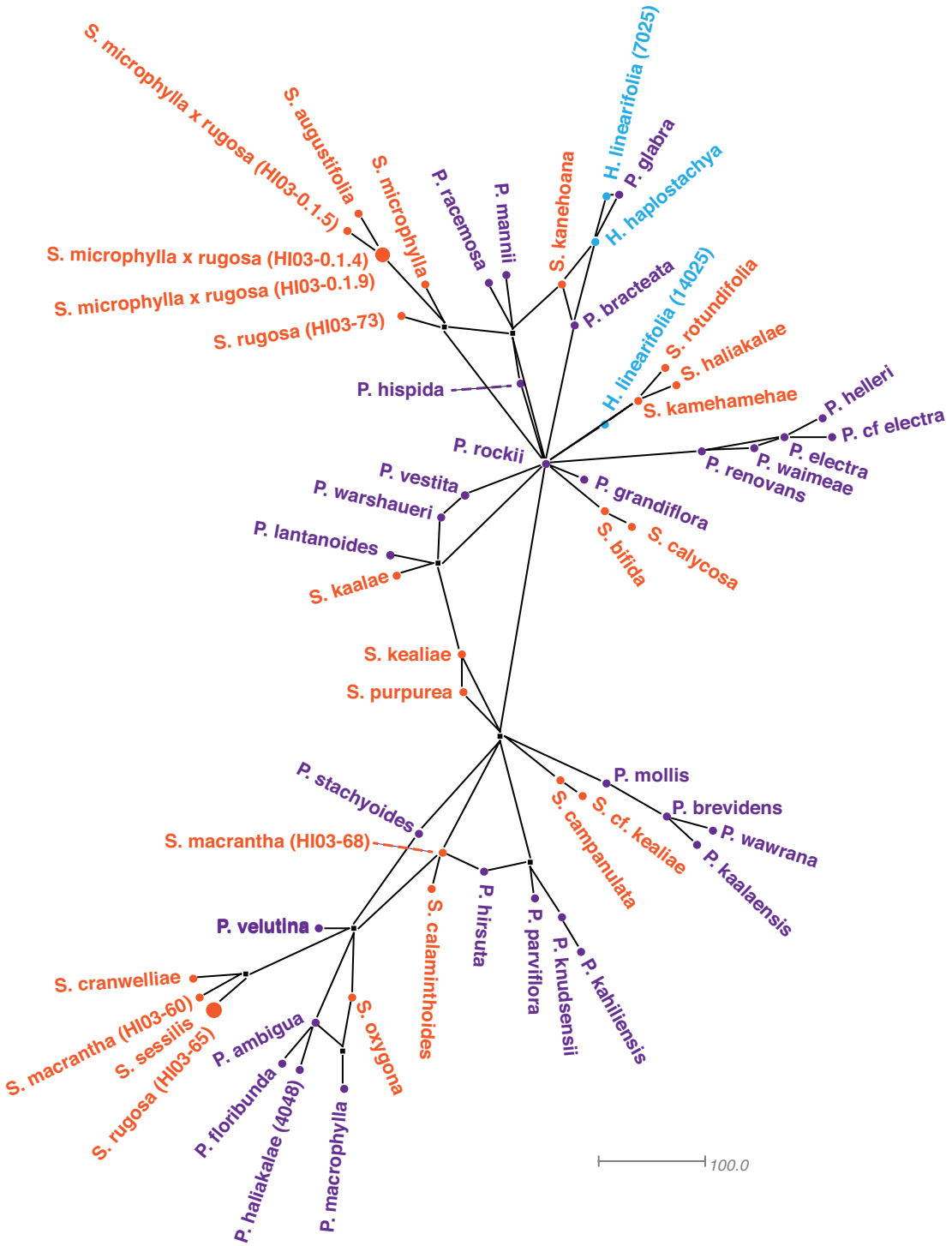

Table 1. 18 samples (representing 16 species) selected for shotgun next-generation sequencing. When multiple accessions of the same species were sequenced the data was pooled for assembly. N/A indicates that the data were not assembled into chloroplast genome sequences due to low coverage.

Species	Year Collected	Total Reads	% Reads Mapped[1]	Length (bp)	Depth	GC Content	Coverage of Reference
Haplostachys haplostachya	2003	503,672	2.05	150,132	10.4x	38.5%	99.9%, 10.3x
*Haplostachys linearifolia**	1918, 1910	4,881,804	0.29	150,129	3.1x	38.7%	89.1%, 4.2x
*Haplostachys munroi**	1916, 1927	1,192,670	0.45	N/A	N/A	N/A	36.3%, 1.5x
Phyllostegia haliakalae	1916	1,139,796	0.45	N/A	N/A	N/A	42.0%, 2.7x
Phyllostegia rockii	1910	936,946	0.28	N/A	N/A	N/A	43.4% 2.6x
Phyllostegia variabilis	1903	75,618	0.08	N/A	N/A	N/A	N/A
Phyllostegia velutina	2003	6,768,775	0.32	150,141	6.9x	38.2%	99.9%, 6.9x
Phyllostegia waimeae	2003	522,434	1.28	150,137	6.7x	38.2%	99.6%, 6.4x
Stenogyne bifida	2003	7,207,150	1.43	150,129	33.2x	38.5%	100%, 33.6x
Stenogyne haliakalae	1936	18,808,049	0.77	149,736	43.9x	38.5%	100%, 43.9x
Stenogyne kanehoana	1934	11,283,860	0.45	150,116	16.2x	38.5%	99.9%, 14.9x
Stenogyne sessilis	2003	599,670	1.00	150,120	6.0x	37.9%	98.7%, 6.0x
Stachys byzantina	2010	425,836	1.84	149,722	8.6x	38.5%	97.3%, 7.1x
Stachys chamissonis	2010	5,140,822	3.71	150,254	193x	38.5%	100%, 190x
Stachys coccinea	2010	585,188	3.03	150,674	17.9x	38.5%	99.9%, 17.7x
Stachys sylvatica	2010	1,185,322	3.03	150,195	36.9x	38.5%	100%, 34.5x

[1]PCR duplicates removed; *Species for which sequences were pooled from multiple accessions (see Table 2)

Table 2. Voucher information for the 105 mint accessions utilized in phylogenomic analyses, as well as four additional accessions that were shotgun sequenced but not included due to missing data (indicated by §).

Species	Voucher Information[1]	Year collected	Geography	Conservation status[2]
Haplostachys haplostachya	V.A. Albert et al. HI03-33 (NTBG; 950403)*	2003	Hawai'i	E
Haplostachys linearifolia	J.F. Rock 14025 (BISH)[3]*	1918	Maui, Moloka'i	Ex
	J.F. Rock 7025 (BISH)[3]*	1910		
Haplostachys munroi	C.N. Forbes 486 (BISH)§	1916	Lanai	Ex
	J.C. Munro 17 (BISH)§	1927		
Phyllostegia ambigua	V.A. Albert et al. HI03-59 (VRPF; 000606)	2003	Hawai'i, Maui, Moloka'i	NA
Phyllostegia bracteata	H.L. Oppenheimer H80116 (BISH)	2001	Hawai'i, Maui	R
Phyllostegia brevidens	W.L. Wagner 5897 (BISH)	1988	Hawai'i, Maui	E
Phyllostegia cf. *electra*	V.A. Albert et al. HI03-34 (NTBG; 12-8-01)	2003	Kaua'i	NA
Phyllostegia electra	K. Wood 2967 (BISH)	1994	Kaua'i	NA
Phyllostegia floribunda	V.A. Albert et al. HI03-63 (VRPF; 971002)	2003	Hawai'i	NA
Phyllostegia glabra	K. Wood 3962 (NY)	NA	Lana'i, Maui, Moloka'i, O'ahu	NA
Phyllostegia grandiflora	S.H. Sohmer s.n. (BISH)	1985	O'ahu	NA
Phyllostegia haliakalae	G.C. Munro 490 (BISH)§	1916	Lana'i, Maui, Moloka'i	Ex
	D. Herbst 4048 (BISH)	1974		
Phyllostegia helleri	S. Perlman 15690 (BISH)	1997	Kaua'i	R
Phyllostegia hirsuta	J. Obata s.n. (BISH)	1993	O'ahu	NA
Phyllostegia hispida	L. Stemmerman 3973 (BISH)	1979	Moloka'i	NA
Phyllostegia kaalaensis	S. Perlman 6117 (BISH)	1987	O'ahu	E
Phyllostegia kahiliensis	W.L. Wagner 5217 (BISH)	1983	Kaua'i	R
Phyllostegia knudsenii	K. Wood 2583 (BISH)	1993	Kaua'i	Ex
Phyllostegia lantanoides	J. Obata 86-624 (BISH)	NA	O'ahu	NA

Phyllostegia macrophylla	V.A. Albert et al. HI03-54 (VRPF; 010616)	2003	Hawai'i, Maui	NA
Phyllostegia mannii	F.R. Warshauer 2418 (BISH)	1979	Maui, Moloka'i	E
Phyllostegia mollis	W. Takeuchi 3497 (BISH)	1987	O'ahu	E
Phyllostegia parviflora	J. Obata s.n. (BISH)	1990	O'ahu	E
Phyllostegia racemosa	V.A. Albert et al. HI03-57 (VRPF; 950601)	2003	Hawai'i	E
Phyllostegia renovans	S. Perlman 16491 (BISH)	1999	Kaua'i	R
Phyllostegia rockii	C.N. Forbes 199 (BISH)[3*]	1910	Maui	Ex
Phyllostegia stachyoides	V.A. Albert et al. HI03-28 (NTBG; 010358)	2003	Hawai'i, Maui, Moloka'i	R
Phyllostegia variabilis	W.A. Bryan s.n. (BISH)[§]	1903	Laysan Island	Ex
Phyllostegia velutina	V.A. Albert et al. HI03-61 (VRPF; 960302)*	2003	Hawai'i	E
Phyllostegia vestita	V.A. Albert et al. HI03-62 (VRPF; 010206)	2003	Hawai'i	NA
Phyllostegia waimeae	V.A. Albert et al. H103-27 (NTBG; 010580)*	2003	Kaua'i	Ex?
Phyllostegia warshaueri	V.A. Albert et al. H103-30 (NTBG; 990810)	2003	Hawai'i	E
Phyllostegia wawrana	S. Perlman 13690 (NY)	NA	Kaua'i	E
Stenogyne angustifolia	F.R. Warshauer 2171 (BISH)	1979	Hawai'i	E
Stenogyne bifida	V.A. Albert et al. HI03-32 (NTBG; 8-7-98)*	2003	Moloka'i	E
Stenogyne calaminthoides	V.A. Albert et al. HI03-04 (O)	2003	Hawai'i	NA
Stenogyne calycosa	A.C. Medeiros 532 (BISH)	1983	Maui	NA
Stenogyne campanulata	V.A. Albert et al. HI03-36 (NTBG; 030081)	2003	Kaua'i	E
Stenogyne cranwelliae	V.A. Albert et al. HI03-64 (VRPF; 010509)	2003	Hawai'i	E
Stenogyne haliakalae	G.E. Olson 5 (BISH)*	1936	Maui	Ex
Stenogyne kaalae	W. Takeuchi 2035 (BISH)	1983	O'ahu	NA
Stenogyne kamehamehae	S. Perlman 6933 (BISH)	1987	Maui, Moloka'i	NA
Stenogyne kanehoana	H. St. John 14046 (BISH)*	1934	O'ahu	E
Stenogyne cf. *kealiae*	K.R. Wood 4017 (BISH)	1995	Kaua'i	E
Stenogyne kealiae	V.A. Albert et al. HI03-31 (NTBG; 990434)	2003	Kaua'i	E

Stenogyne macrantha	V.A. Albert et al. HI03-68 (VRPF; 020425)	2003	Hawai'i	NA
	V.A. Albert et al. HI03-60 (VRPF; 990512)	2003		
Stenogyne microphylla x rugosa	V.A. Albert et al. HI03-0.1.9 (O)	2003	Hawai'i	NA
	V.A. Albert et al. HI03-0.1.4 (O)	2003		
	V.A. Albert et al. HI03-0.1.5 (O)	2003		
Stenogyne microphylla	V.A. Albert et al. HI03-21 (O)	2003	Hawai'I, Maui	NA
Stenogyne oxygona	V.A. Albert et al. HI03-66 (O)	2003	Hawai'i	Ex?
Stenogyne purpurea	V.A. Albert et al. HI03-40 (O)	2003	Kaua'i	NA
Stenogyne rotundifolia	F.R. Warshauer 2545 (BISH)	1980	Maui	NA
Stenogyne rugosa	V.A. Albert et al. HI03-65 (VRPF; 021207)	2003	Hawai'i, Maui	NA
	V.A. Albert et al. HI03-73 (O)	2003		
Stenogyne sessilis	V.A. Albert et al. HI03-67 (VRPF; 021212)*	2003	Hawai'i, Maui	NA
Stachys aethiopica	B. Petterson 2146 (UPS)	1982	S. Tropical to S. Africa	NA
Stachys affinis	C. Lindqvist & V.A. Albert 359 (UNA)	2001	East Asia	NA
Stachys ajugoides var. *rigida*	C. Lindqvist 11-02 (UB)	2011	W. USA to Mexico	NA
Stachys albens	C. Lindqvist 11-06 (UB)	2011	California	NA
Stachys alpigena	O. Ryding 2133 (UPS)	1990	East Africa	NA
Stachys alpina	P. Frost-Olsen 4239 (AAU)	1982	Europe to N. Iran	NA
Stachys arabica	I. Gruenberg 685 (UPS)	1957	S. Turkey to Israel	NA
Stachys arvensis	N. Lundqvist 8157 (UPS)	1973	Macaronesia to Taiwan	NA
Stachys aspera	F.F. Forbes s.n. (UPS)	1911	N.C. & E. USA	NA
Stachys bullata	C. Lindqvist 11-04 (UB)	2011	California	NA
Stachys byzantina	C. Lindqvist 10-04 (UB)*	2010	N. Turkey to N. Iran	NA
Stachys chamissonis	C. Lindqvist 11-01 (UB)	2011	W. Canada to W. USA	NA
	C. Lindqvist 10-02 (UB)*	2010		NA
Stachys coccinea	K.I. Matthews 745 (AAU)	1982	Arizona to Texas and C.	NA

	C. Lindqvist 10-03 (UB)*	2010	America	NA
Stachys corsica	J. Lambinon 81/sa/191 (AAU)	1981	Corse, Sardegna	NA
Stachys cretica	M.H.G. Gustafsson 128 (UPS)	1991	Mediterranean to W. Asia	NA
Stachys debilis	C & E. Franquemont 106 (AAU)	1983	Ecuador	NA
Stachys elliptica	E. Asplund 6377 (US)	1939	Ecuador	NA
Stachys eriantha	J.A. Ewan 16392 (US)	1944	Mexico, Colombia to N.W. Venezuela and Ecuador	NA
Stachys floridana	J.B. Nelson s.n. (USCH)	2009	Florida	NA
Stachys gilliesii	S. Pfanzelt 241 (M)	2007	W. South America to Chile and Brazil	NA
Stachys globosa	S. Lemus 143 (AAU)	1983	N.E. Mexico	NA
Stachys grandidentata	W.J. Eyerdam 10081 (US)	1957	N. & C. Chile	NA
Stachys hamata	B. Lojtnant et al. 11835 (AAU)	1979	Colombia to Ecuador, NE. Venezuela	NA
Stachys herrerae	F.L. Herrera 3499 (US)	1932	Peru	NA
Stachys hispida	S.R. Zeigler & M.F. Leykern 1963 (AAU)	1975	C. & E. Canada, NC. & E. U.S.A.	NA
Stachys hyssopoides	E. Retief 1080 (US)	1983	S. Africa	NA
Stachys inflata	J.S. Andersen & I.C. Petersen 68 (AAU)	1969	NE. Turkey to Iran	NA
Stachys kouyangensis	B. Bartholomew et al. 1362 (US)	1984	Tibet, SC. China to Indo-China	NA
Stachys lamioides	E. Asplund 17092 (US)	1955	Colombia to Ecuador	NA
Stachys latidens	O. Ryding 2010 (UPS)	1989	N.C. & E. USA	NA
Stachys lavandulifolia	J.S. Andersen & I.C. Petersen 31 (AAU)	1969	Turkey to Iran	NA
Stachys macraei	F.C. Joseph 3628 (US)	1925	C. & S. Chile	NA
Stachys maritima	I. Gergely 3362 (US)	1971	Medit. to W. Caucasus	NA
Stachys mexicana	L. Holm BCU70 (UPS)	1959	Alaska to California	NA

Stachys natalensis	O. Ryding 2332 (UPS)	1992	Swasiland	NA
Stachys pacifica	B.L. Phillips 1043 (TEX)	1990	S. Tropical & S. Africa	NA
Stachys palustris	G. Kulikova 5886 (US)	1974	Europe to Himalaya	NA
Stachys pilosa	G.A. Wheeler 11047 (AAU)	1988	Alaska to USA	NA
Stachys recta	M.D. Tidestrom (US)	1936	Europe to N.W. Iran	NA
Stachys riederi	J.F. Rock 14181 (US)	1926	Siberia to Japan	NA
Stachys rugosa	W.J. Hanekom 2487 (US)	1966	S. Africa	NA
Stachys sericea	F.C. Joseph 4327 (US)	1926	C. Chile	NA
Stachys strictiflora	C. Lindqvist 10-07 (UB)	2011	China	NA
Stachys sylvatica	C. Lindqvist & V.A. Albert 358 (UNA)*	1998	Europe to Himalaya	NA
Stachys tenuifolia	J.B. Nelson s.n. (USCH)	2009	E. Canada, C. & E. U.S.A.	NA
Stachys vulnerabilis	G.B. Hinton et al. 24774 (TEX)	1995	N.E. Mexico	NA
Suzukia shikikunensis	Chii-Cheng Liao et al. 564 (A)	1992	Taiwan	NA

[1]A: Harvard University; AAU: Aarhus University; BISH: Bishop Museum; M: Botanische Staatssammlung München; NY: New York Botanical Garden; NTBG: National Tropical Botanic Garden; O: Oslo University; TEX: University of Texas at Austin; UB: University at Buffalo; UNA: University of Alabama; UPS: Uppsala University; US: Smithsonian Institution; USCH: University of South Carolina; VRPF: Volcano Rare Plant Facility. [2]Conservation status only reported for the Hawaiian taxa. E: Endangered, Ex: Extinct, R: Rare, NA: not applicable (Wagner et al. 1999, Wagner 1999). [3]Included using both shotgun and targeted sequencing data). *Included in phylogenies based on shotgun sequencing data.