Genomic histories of polyploidy, diversification, and admixture in a Hawaiian plant radiation

Charlotte Lindqvist

cl243@buffalo.edu

University at Buffalo https://orcid.org/0000-0002-4190-727X

Crystal Tomlin

University at Buffalo

Sitaram Rajaraman

Nanyang Technological University https://orcid.org/0000-0001-5171-3578

Jeanne Sebesta

University at Buffalo

Anne-Cathrine Scheen

Stavanger Botanic Garden

Mika Bendiksby

University of Oslo

Yee Wen Low

Singapore Botanic Gardens https://orcid.org/0000-0002-0939-9068

Jarkko Salojärvi

University of Helsinki

Todd Michael

The Salk Institute https://orcid.org/0000-0001-6272-2875

Victor Albert

University at Buffalo https://orcid.org/0000-0002-0262-826X

Article

Keywords:

Posted Date: July 27th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3167181/v1

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Communications on April 10th, 2024. See the published version at https://doi.org/10.1038/s41467-024-47247-y.

Genomic histories of polyploidy, diversification, and admixture in a Hawaiian plant radiation

Crystal M. Tomlin[1], Sitaram Rajaraman[2,3], Jeanne T. Sebesta[1], Anne-Cathrine Scheen[4], Mika Bendiksby[5], Yee Wen Low[6], Jarkko Salojärvi[2,3], Todd P. Michael[7], Victor A. Albert[1], Charlotte Lindqvist[1]

[1]Department of Biological Sciences, University at Buffalo, New York, USA.
[2]School of Biological Sciences, Nanyang Technological University, Singapore.
[3]Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland.
[4]Stavanger Botanic Garden, City of Stavanger, Norway.
[5]Natural History Museum, University of Oslo, Oslo, Norway.
[6]Singapore Botanic Gardens, National Parks Board, Singapore.
[7]The Plant Molecular and Cellular Biology Laboratory, Salk Institute for Biological Studies, La Jolla, California, USA.

Author Contributions

V.A.A. and C.L. designed research; C.M.T., S.R., J.T.S, Y.W.L., J.S., T.P.M., V.A.A., and C.L. performed research or analyzed data; A.-C.S., M.B., Y.W.L., J.S. and T.P.M. contributed reagents, materials, and/or tools; C.M.T., V.A.A., and C.L. wrote the paper.

Acknowledgements

The authors acknowledge financial support from the School of Biological Sciences of Nanyang Technological University and U.S. National Science Foundation awards 2139311 to C.L. and 2030871 to V.A.A. The authors thank Singapore Botanic Gardens for laboratory support, Lyman Perry for permission to collect on Hawaiian State lands, and Patty Moriasu at the Volcano Rare Plant Facility, the National Tropical Botanical Garden, Elizabeth Stacy and Jennifer Johansen for plant specimens.

Data Availability

The genome data generated in this study has been deposited in the NCBI database. For reference individual *Stenogyne calaminthoides*, Hi-C reads are under accession SRR23341345, RNA-seq data under accession SRR23341344, Illumina shotgun data under accession SRR23341343, and Nanopore raw reads under accession SRR23341342. The *Stenogyne calaminthoides* genome assembly and annotation is also available on CoGe XXX (https://genomevolution.org/coge/). Raw reads for resequenced samples can be found under accession numbers SAMN32767766-SAMN32767919. Processed data generated in this study and used for main text figures are provided in source data files.

Abstract

Island systems provide important contexts for studying processes underlying lineage migration, species diversification, and organismal extinction. The Hawaiian endemic mints (Lamiaceae family) are the second largest plant radiation on the isolated Hawaiian Islands. We generated a chromosome-scale reference genome for one Hawaiian species, *Stenogyne calaminthoides*, and resequenced 45 relatives, representing 34 species, to uncover the continental origins of this group and their subsequent diversification. We further resequenced 109 individuals of two *Stenogyne* species, and their purported hybrids, found high on the Mauna Kea volcano on the island of Hawai'i. The three distinct Hawaiian genera, *Haplostachys*, *Phyllostegia*, and *Stenogyne*, are nested inside a fourth genus, *Stachys*. We uncovered four independent polyploidy events within *Stachys*, including one allopolyploid hybridization event underlying the Hawaiian mints and their direct western North American ancestors. While the Hawaiian taxa may have principally diversified by parapatry, localized admixture may have played an important role early in lineage diversification. Our genomic analyses provide a view into how organisms have radiated on isolated island chains, a topic that provided one of the principal natural laboratories for Darwin's thinking about the evolutionary process.

Introduction

Organismal radiations are unique cases in which species exhibit great morphological and ecological diversity despite limited genetic differentiation. Many such diversifications have been termed adaptive radiations[1], even in the absence of clarity on whether adaptive forces primarily led to species differentiation. The Hawaiian Islands are an exemplar natural laboratory to study radiations and investigate their possible adaptive nature, due to their habitat diversity and isolated location over 3,700 km from any substantial landmass. The island chain comprises a series of volcanoes that formed sequentially via movement of the Pacific Plate over a mantle hot spot, and they have well-characterized formation and erosional age profiles, both among them and within given islands[2,3]. On the youngest and largest island, Hawai'i, the creation of new habitats and dissection of older ones by lava flows is observable in real time. This continuous eruption-based fragmentation over geological time of both pioneer and old-growth habitats, and the age and ecological/geological gradients within the high islands themselves, has no doubt promoted speciation in many lineages. Indeed, the Hawaiian Islands feature one of the most dramatic rainfall gradients in the world[4], providing incredibly diverse environments from alpine deserts to wet forests[5]. The question, however, remains: are bursts of speciation, e.g., in the Hawaiian Islands, due principally to geographic (neutral) speciation, to ecological (adaptive) speciation, or a combination of both? There is also the question of what genomic features might underpin a successful organismal radiation. Two commonly proposed phenomena are hybridization[1] and polyploidization[6,7], which can of course occur together. In the case of hybridization, incomplete lineage sorting (ILS) of genetic variation that defies the species phylogeny can leave a similar signature of disrupted monophyly, necessitating careful distinction between the two phenomena. This is especially important in the case of rapidly radiating lineages, in which time for ancestral polymorphisms to completely sort by lineages is often extremely short[8].

Despite great interest in the evolution of organismal radiations, only one tenth of (presumed) adaptive radiation studies focus on plants[9]. Of these relatively few plant studies, the vast majority have used limited DNA markers, or reduced genomic representation for phylogenomic reconstruction[10], which frequently do not provide the necessary resolution to understand evolutionary histories complicated by features such as whole genome duplication, paralogy, hybridization, and rapid genetic divergence[11]. With such rapid divergence and little genetic differentiation it can be challenging to distinguish ILS from hybridization using either maternally inherited (even whole plastome) or single-copy nuclear loci. Plant radiations studied to-date, including the Hawaiian silverswords[12], lobeliads[13,14], and Hawaiian endemic mints[15-19], have mostly fallen into this experimental realm, although recent research, including the o'hia tree, *Metrosideros polymorpha*[20,21], and the species-rich Southeast Asian *Syzygium*[22], has taken advantage of a reference genome and large-scale genome resequencing.

Among the most species-rich angiosperm radiations to occur on the Hawaiian Islands is that of the Hawaiian endemic mint lineage (Lamiaceae), consisting of ~60 species assigned to three genera, *Haplostachys*, *Phyllostegia,* and *Stenogyne*[23,24]. *Haplostachys*, which consists of five species with only a single extant taxon, *H. haplostachya*, is the only genus with dry fruits, and it mostly occurs in dry habitats at low-mid elevation. *Haplostachys* and *Phyllostegia* (34 species) both have mostly pink to white, fragrant flowers with a prominent lower corolla lip, typically associated with insect pollination. *Stenogyne* (21 species) and *Phyllostegia* primarily occur in wet/mesic forest environments[24]. *Stenogyne* is unique in that it has primarily pink to red

nectar-producing flowers with a reduced lower lip and a longer corolla tube, typical of bird pollination. Prior, marker-sparse phylogenetic analyses of the Hawaiian mints discovered that the Hawaiian endemic mints form a monophyletic group nested within the nearly global genus *Stachys*[15]. *Stachys* is the most species-rich genus within the tribe Stachydeae, a tribe within Lamiaceae subfamily Lamioideae that exhibits evolutionary complexities such as a broad range of ploidy levels, frequent natural hybridization events, and paraphyletic genera (of which *Stachys* itself is a prominent example)[25,26].

According to allelic data and chromosome counts, the Hawaiian mints are likely paleo-octoploids, bearing evidence for two polyploidy events[17] following the ancient *gamma* triplication at the base of all core eudicots[27]. Further studies indicated that the initial colonizer of the Hawaiian Islands may have been of allopolyploid hybrid origin, as suggested by phylogenetic discordance between genetic markers, where plastid sequences showed the lineage to be most closely related to meso-South American *Stachys*, while nuclear markers linked the clade to temperate North American *Stachys* species[15,18,19,28]. In addition to a hybrid origin for the lineage, there have been reports of ongoing interspecies admixture at Mauna Kea on the Big Island, Hawai'i[24]. Hence, the Hawaiian mints are a remarkable case not only of morphological radiation, but also polyploidization, past and present hybridization, and as a model system in which to study the role of these phenomena in evolutionary diversification on oceanic islands.

In this study, we assembled a high-quality reference genome of a Hawaiian mint species, *Stenogyne calaminthoides* (**Fig. 1a**), utilizing DNA reads generated with Oxford Nanopore and Illumina sequencing technologies and HiC-scaffolding. We resequenced 30 additional Hawaiian mint taxa and 15 of their relatives in the genus *Stachys*. We also resequenced 109 individuals from a putative hybrid swarm of *Stenogyne rugosa* and *S. microphylla* found on Mauna Kea[24] and recorded their morphological traits. We used these genomic resources to detail evolutionary paths within the Hawaiian mint lineage, from deep to present time, studying the genomic signatures and impacts of polyploidization, taxonomic radiation, and hybridization.

Results and Discussion

Reference genome characterization unveils a repeat-dense genome

The genome size of *Stenogyne calaminthoides* was first predicted to be ~1.2 – 1.6 billion bases (Gb) based on *k*-mer analyses of Illumina sequences. The initial assembly of *S. calaminthoides,* using over 75 Gb of Nanopore reads greater than 35 thousand bases (Kb) in length, resulted in an assembly of ~2.4 Gb. This genome size of approximate diploid value suggested that our primary assembly may have contained two divergent haplotypes. Visualizing the assembly graph, we identified a large "knot" of repetitive sequences, perhaps corresponding to a relatively recent transposable element (TE) burst (Supplementary Fig. 1). This initial genome assembly consisted of over 77% repetitive elements, 46% of which correspond to long terminal repeat (LTR) retrotransposons (Supplementary Table 1). After filtering diploid regions from the assembly, the resulting haploid assembly reduced to ~1.4 Gb in size, consistent with the *k*-mer-based haploid estimates. After scaffolding with Hi-C reads (**Fig. 1b**), the final assembly had an N50 over 37 Mb and contained 434 scaffolds, including 32 scaffolds larger than 25 Mb, closely matching the expected chromosome count of $n = 32$[29] (**Fig. 1c**, Supplementary Table 2). The unannotated assembly had a BUSCO (Benchmarking Universal Single-Copy Orthologs) completeness score of 94.9% (49.4% complete single copy and 45.5% duplicated genes) (Supplementary Table 3).

Annotation of the genome resulted in a total of 77,090 genes with a BUSCO score of 86% (consisting of 35.3% single copy and 50.7% duplicated genes) and an N50 of 4,162 bp (Supplementary Table 3). The large number of genes (and high percentage of BUSCO duplicates) may reflect an incompletely fractionated (diploidized) gene space following a recent whole genome duplication (WGD) event (discussed further below)[30], as has been noted for polyploid species, including the paleo-octoploid Lamiaceae species *Pogostemon cablin*, which had a reported 110,850 genes total[31].

Genome architecture and polyploid history within Lamiales is marked by a shared ancient allopolyploid event

The *Stenogyne calaminthoides* genome was compared with several other representative core eudicot genomes using syntenic depth FractBias plots[32]. Such plots are a useful means to query not only ploidy depths, but also potential subgenomic dominance biases at the time a WGD occurred[33]. Fractionation bias patterns of *S. calaminthoides* mapped against grapevine (*Vitis vinifera*) showed 4:1 synteny for the two species, respectively (**Fig. 1d**). This is strong evidence, along with the accompanying syntenic dotplot (SynMap) visualization (Supplementary Fig. 2), that the Hawaiian mint species underwent two sequential WGD events since common ancestry with *Vitis*. Of the four *Stenogyne* genomic blocks mapping to grapevine chromosomes, there are two pairs of blocks wherein each member of a pair follows the other closely in gene retention (fractionation) percentage, i.e., presenting little evidence for diploidization bias. Each of these pairs are in turn broadly separated in fractionation percentage, suggesting substantial subgenome dominance during the earlier WGD event. Distinct epigenetic masking in progenitor genomes of allopolyploid lineages can lead to considerable purifying selection differences between subgenomes, resulting in dominant/recessive (biased) patterns. Unbiased fractionation between subgenomes can in turn reflect little masking difference, or even autopolyploidy[34,35]. As such, the most recent, less biased WGD visible in *Stenogyne* may have involved a narrow cross, whereas the earlier polyploidy event likely involved divergent progenitor species.

Other sequenced Lamiales genomes also show signatures of ancient WGD since ancestry with grapevine. The 1KP Project (the One Thousand Plant Transcriptomics Project or OTPTP) reported a WGD event (denoted as ANMAα) shared by many core Lamiales families including Lamiaceae[36]. In contrast, other investigators reported that the WGD events of the Lamiaceae species *Callicarpa americana* and *Tectona grandis* were independent based on synonymous substitution (*K*s) distributions for paralogous gene pairs, and that *Tectona* experienced an additional ploidy event[37]. However, this determination based on *K*s alone may have been flawed without macrosyntenic analysis, and stood in contrast to results from the *Tectona* genome[38]. Our comparisons of the *Stenogyne* genome with other Lamiales species support the findings of the 1KP Project, suggesting that the early WGD in *Stenogyne* was likely a shared, single event shared with *Callicarpa* and *Tectona,* as well as many other core Lamiales, such as *Mimulus guttatus* (Phrymaceae) (**Fig. 1e**). For example, the fractionation patterns of *Callicarpa* and *Tectona* mapped to *Vitis* are remarkably similar, as these two Lamiaceae taxa both map 2:1 to *Vitis* and even share some chromosomal fusion events since common ancestry with that taxon (Supplementary Fig. 3a). Furthermore, fractionation bias between the two subgenomes of *Buddleja* (Scrophulariaceae) and *Mimulus* (the latter far outside Lamiaceae) are broad (Supplementary Fig. 3b), similar to the first WGD event in *Stenogyne*. Hence, we conclude the presence of one shared allopolyploidy event at the base of Lamiaceae and most other core Lamiales. Further, a second WGD in *S. calaminthoides* is apparent (syntenic depth plots showed

up to 8X depth, while *Callicarpa* and *Tectona* had 6X depth; Supplementary Fig. 4), verifying the additional, more recent polyploidy event in *Stenogyne* since its split with these other members of Lamiaceae. Finally, syntenic dotplots of *Stenogyne* vs. *Callicarpa* and *Stenogyne* vs. *Tectona* (Supplementary Fig. 5) similarly showed 4:2 ratios, confirming a doubling of the *Stenogyne* genome since its split with these two relatives.

Phylogenomic analyses reveal that Hawaiian mints are most closely related to western North American *Stachys*

To investigate relationships among the Hawaiian mints and their closest relatives in Stachydeae, we mapped Illumina reads for 45 mint samples to the *Stenogyne calaminthoides* reference genome and generated maximum likelihood phylogenies based on both nuclear genomic single nucleotide polymorphisms (SNPs) and assembled chloroplast genomes (Supplementary Table 4-7). Consistent with previous analyses[28], the chloroplast genome phylogeny demonstrated weak resolution of relationships among Hawaiian mints, as well as very short internal branch lengths, as expected of a rapid radiation (Supplementary Fig. 6). However, two main clades were strongly supported, one grouping Hawaiian mints with *Stachys coccinea* and one individual of *Stachys chamissonis*, and another clade containing eastern North American and Asian (ENAA) *Stachys*, including a second accession of *S. chamissonis*. Confirming previous findings[15,18,19], phylogenies based on nuclear data (**Fig. 2**, Supplementary Fig. 7) were discordant with the plastid tree in that (*i*) the closest relatives of Hawaiian mints were their western North American (WNA) *Stachys* relatives (*S. bullata,* two representatives of *S. chamissonis, S. ajugioides,* and two representatives of *S. albens*), (*ii*) the two *S. chamissonis* individuals appeared in the same clade, and (*iii*) *S. coccinea* was instead a distant relative. These relationships could be explained if a *S. coccinea*-like ancestor was the maternal progenitor of an allopolyploid hybrid lineage that colonized the Hawaiian Islands, and that the chloroplast genome was captured by *S. chamissonis* (at least the population bearing this plastome) in this hybridization event. The recentmost WGD visible in *Stenogyne*, if an allopolyploidy event, may have co-occurred with this plastome capture. However, the fact that all Hawaiian mints and a single WNA *Stachys chamissonis* individual share a plastid haplotype suggests a more complex scenario. It is possible that hybridization and polyploidization co-occurred multiple times independently, or that hybridization among allopolyploids followed WGD events, as has been suggested for *Dactylorhiza*[39]. This scenario may have been further confounded by subsequent ILS of the two plastid haplotypes.

In the SNP phylogeny, monophyly of the entire Hawaiian lineage was strongly supported (**Fig. 2**, Supplementary Fig. 7). Within the Hawaiian mints, two main clades of *Phyllostegia* were joined by a very short internal branch. Two *Phyllostegia warshaueri* samples and *P. racemosa* each had long external branches corresponding to unique variation of unknown origin, unsampled species diversity, or possibly extensive extinction along their lineages[40]. There was no strong pattern of phylogeographic structuring within *Phyllostegia* (Supplementary Fig. 8); however, the two species from Kaua'i (*P. electra* and *P. waimeae*) grouped together. Within *Stenogyne*, one first-branching clade consisted of samples found only on Kaua'i (these species are hereafter referred to as the Kaua'i clade). Sister to this clade, *Stenogyne* split into two additional clades, here referred to as the *S. rugosa* clade (also including *S. microphylla* and *S. bifida*) and the *S. macrantha* clade (also containing *S. calaminthoides*, *S. cranwelliae*, and *S. scrophularioides*). Interestingly, within the *S. macrantha* clade, not all samples of *S. macrantha* grouped together, possibly reflecting representation of two extreme morphological forms, as previously reported[24].

Genes from de novo assembled genomes show *Phyllostegia* as paraphyletic

Because SNP trees reflect a phylogenetic summary across the entire genome, we next used an independent method to generate a locus-by-locus coalescent species tree, both to avoid potential reference genome mapping bias[41] and to supplement the SNP-based results. We generated *de novo* Illumina-based genome assemblies from the 45 resequenced samples and extracted their conserved BUSCO genes (Supplementary Fig. 9). The average BUSCO completeness score was ~87% (Supplementary Table 8). The multilocus coalescent phylogeny, based on 1,336 BUSCO genes, agreed with the SNP tree in that WNA *Stachys* were the closest relatives to the Hawaiian mints, but *Phyllostegia* was resolved as paraphyletic (**Fig. 2**, Supplementary Fig. 10). An additional discordance between the SNP- and BUSCO-based trees was that the position of *S. sylvatica* and *S. coccinea* interchanged. In cases of discordance within Hawaiian mints, differences corresponded to very short internal branches in both trees, possibly caused by rapid radiation bursts that are difficult to tease apart phylogenetically. There was also discordance within specific lineages, such as the *Stenogyne* Kaua'i clade, wherein *S. kealiae* (in the SNP tree), instead of *S. campanulata* (in the BUSCO tree), occupied a position as sister to the rest of this clade. There was one instance of discordance within WNA *Stachy*s, where one *S. chamissonis* individual was sister to all other WNA *Stachys* samples plus Hawaiian mints in the BUSCO tree only; this sample instead grouped within WNA *Stachys* in the SNP tree. Such differences based on coalescent analyses of single loci versus a whole-genome variant average may reflect ILS being particularly confounding in the heritage of this lineage.

Whole genome duplication histories reveal four independent polyploidy events

To evaluate polyploid depths among species, we compared percent duplicated BUSCOs (D) in the *de novo* Illumina assemblies (**Fig. 3a**, Supplementary Fig. 9). D scores >35% marked likely polyploid lineages, whereas scores of ~4% suggested diploid status. There were three clades of putative polyploids, the Hawaiian mints and their closest WNA *Stachys* relatives, the *S. coccinea* lineage, and the *S. sylvatica* lineage, whereas the outgroup taxon *S. byzantina* and most ENAA *Stachys* taxa (*S. floridana*, *S. tenuifolia*, and *S. strictiflora*) appeared to be diploid. These determinations were also reflected by chromosome counts. *Stachys byzantina* and *S. floridana* (and diploid populations of *S. tenuifolia*) are $2n = 30$ and 34, respectively[42,43]. This number is roughly doubled in Hawaiian mints (e.g., *Stenogyne purpurea,* $2n = 66$), WNA *Stachys* relatives (e.g., *S. bullata,* $2n = 66$[18]), and *S. sylvatica*, which is $2n = 66$[42]. *Stachys coccinea* differs with $2n = 84$[42], suggesting that this species has a more complicated polyploid history. One ENAA *Stachys* sample, *S. affinis*, had an excess of duplicated BUSCOs, with about three times the number in *S. byzantina*, but not to the same level as the other paleo-octoploids. However, chromosome counts of this sample report $2n = 66$[18], suggesting that it has also experienced a WGD event. Because this WGD is not shared with other members of its clade (other ENAA taxa) and these other apparent high-polyploid taxa are interspersed within *Stachys sensu lato* phylogeny, it is probable that each WGD observed represents an independent event (Supplementary Fig. 9).

Next, to further evaluate WGD independence and investigate subgenome sharing and potential progenitor lineages of the WGD among Hawaiian mints and their closest relatives, we used the GRAMPA[44] application to resolve a subgenome phylogeny from multi-copy BUSCO genes (Supplementary Fig. 11). This analysis showed that the Hawaiian endemic mints and WNA *Stachys* likely shared an allopolyploidy event, with a diploid ancestor related to *S.*

coccinea (here after refered to as UC – unsampled *S. coccinea*-like) contributing one progenitor lineage, and a diploid ENAA *Stachys*-like ancestor contributing the other. Based on a previous study of five low-copy, independently inherited nuclear loci[19], this UC ancestor could represent a relative of unsampled Mesoamerican or South American species, such as Mesoamerican *S. agraria* or South American *S. eriantha* (both $2n = 32$[42,45]), which belong to the "Meso-SA I" clade. *Stschys coccinea* is found in their "Meso-SA II" clade[19]. GRAMPA also resolved *Stachys sylvatica* to be an allopolyploid of distinct origin, involving a relative of *S. coccinea* and a *S. byzantina*-like ancestor, despite its similar chromosome number to Hawaiian mints. Finally, *Stachys coccinea* itself was resolved as a third, independent allopolyploid (and likely allohexaploid) lineage also descendant from the ENAA *Stachys* lineage and an unsampled, closely related ancestor, which may be a similar but not necessarily the same progenitor inferred for the WNA *Stachys*/Hawaiian mint hybridization event (Supplementary Fig. 11). Lastly, GRAMPA placed both *Stachys affinis* subgenomes in the same clade with its sister taxon, *S. strictiflora*, indicating a rather close and recent allopolyploidy, or perhaps an autopolyploid event (Supplementary Fig. 11). These predicted and independent allopolyploidy events are summarized in **Fig. 3a**.

Overall, polyploidy events appear to be recurrent in the *Stachys s.l.* lineage[25], but they are not necessarily paired with phylogenetic radiations. Likewise, numerous plant radiations do not seem to be associated with polyploidy events, such as Hawaiian *Melicope*[46]. However, it is still possible that the polyploid event underlying Hawaiian mints and North American *Stachys* contributed some genomic substrate for evolutionary radiation of the lineage in the dynamic and fast-evolving Hawaiian landscape[47].

Genome phasing of *S. calaminthoides* supports the ENAA and UC diploid lineages as progenitors for the Hawaiian mints and WNA relatives

We uncovered numerous differentiating features unique to either a homeologous chromosome of ENAA origin, versus one of UC origin. First, when calculating depth of coverage across chromosomes, eastern North America *Stachys* species (*S. floridana* and *S. tenuifolia*) had a set of 17 chromosomes that mapped with high coverage and 15 chromosomes that mapped with much lower coverage (Supplementary Fig. 12, Supplementary Table 9). Conversely, the 15 chromosomes that mapped with lower coverage in ENAA samples mapped with high coverage in *S. coccinea*, while the 17 chromosomes with high coverage in ENAA instead had low coverage in *S. coccinea*. These differences are suggestive of ENAA and UC being the progenitor lineages for *Stenogyne calaminthoides* (and by extension, all Hawaiian mints and WNA *Stachys*; **Fig. 3a**). The 17 chromosomes belonging to the ENAA subgenome likely have higher depth of coverage with ENAA samples due to close phylogenetic relationship, mapping poorly to the UC chromosomes due to their more distant phylogenetic relationship. We found a few unique cases in which the pattern was violated, i.e., where we saw roughly twice the expected coverage for a specific chromosome. For example, sister taxa *Phyllostegia floribunda* and *P. stachyoides* shared doubling of an UC chromosome, and the two *P. warshaueri* specimens shared doubling of a different UC chromosome (Supplementary Fig. 12). All of these instances reflect different known chromosome numbers, and hence, independent formation events. Interestingly, *Stenogyne* species had no such chromosomal variation detected, consistent with known chromosome counts $2n = 64$, while *Phyllostegia* chromosome counts vary, with $2n = 64$ or 66[29]. The latter appear to be cases of aneuploidy, specifically gain of a single chromosome, which has been thoroughly studied in the case of the allopolyploid composite *Tragopogon*[48].

We were further able to identify homeologous pairs of chromosomes, each consisting of one ENAA chromosome and one UC chromosome, using both SubPhaser[49] and DAGchainer[50] in CoGe (Supplementary Table 10). SubPhaser clearly split the ENAA and UC subgenomes based on *k*-mers, the former into a monomorphic cluster suggesting highly similar repeat content, while the latter was much more polymorphic, possibly because UC represents a ghost lineage that was defined on the basis of the allopolyploid *Stachys coccinea*, or that the UC lineage has a deeper coalescence, based on the phylogeny of chromosomes (**Fig. 3b,** Supplementary Fig. 13). Subgenome-wise dating of LTR blooms using two different mutation rate estimates (6.7E-09 vs. 1.75E-09 mutations per site per year, emulating *Arabidopsis*[51]) yielded medians of ~4.5 Million years ago (Mya) and 17 Mya, respectively. The younger date closely corresponds with a previous estimate for the Hawaiian mints/North American *Stachys* clade origin based on nuclear ribosomal and plastid genome markers[18].

Using CoGe, we output syntenic gene pair results to a Circos plot, which also demonstrated relatively clear pairings between homeologs (**Fig. 1b**), with the exception of two ENAA chromosomes that did not have a clear syntelogs among chromosomes greater than 25 Mb. Indeed, also confirmed using CoGe SynMap (Supplementary Fig. 14), chromosome 47 seemingly shared synteny with blocks smaller than 25 Mb, possibly representing scaffolding errors or potentially chromosomal fission events. The other singleton, chromosome 42, was partially syntenic to smaller scaffolds and partially syntenic to another chromosome (chromosome 41), possibly also representing scaffolding errors or chromosomal fusions. Hence, for downstream analyses based on chromosome and subgenome comparisons, we omitted these chromosomes of unclear homeology focusing instead on the 15 well-defined syntenic pairs.

Incongruent phylogenetic signal among the Hawaiian mints and relatives

Since the Hawaiian mint radiation likely involved divergence over a relatively short time period, extensive ILS may be a confounding factor for phylogenetic reconstruction. We aimed to search for incongruent phylogenetic signals along the genome using *Twisst*[52] (topology weighting by iterative sampling of sub-trees). First, to investigate competing topologies, each Hawaiian mint genus was assigned a group and *Stachys* was separated into five monophyletic groups (Supplementary Fig. 15a), producing 25,656 SNP trees based on windows of 50 SNPs each. We found that the dominant topology (**Fig. 3c**) followed the BUSCO tree topology, followed by the genome-wide average SNP tree topology (**Fig. 2**). In the third most common topology, *Stachys sylvatica* and *S. coccinea* were resolved as sister taxa. Indeed, most of the conflicting phylogenetic signals among the 50 SNP windows could be attributed to different positioning of *S. sylvatica* and *S. coccinea*. This was expected given our model of subgenome sharing, where *S. sylvatica* and *S. coccinea* both possess at least one subgenome from the lineage leading to UC *Stachys* (**Fig. 3a**).

Next, to tease apart phylogenetic signals by subgenome, we plotted *Twisst* results according to chromosome assigned to each of the ENAA and UC subgenomes. We found a clear trend in phylogenetic signal for the UC chromosomes compared to the ENAA chromosomes (Supplementary Fig. 16). The ENAA chromosome topologies were fairly consistent with the pattern seen using genomewide SNPs, with the two top topologies mainly reflecting *Stachys sylvatica* and *S. coccinea* phylogenetic discordance (Supplementary Fig. 16a). For four of the fifteen ENAA chromosomes, beginning with the third best topology, phylogenetic placements within Hawaiian mints shuffled, with *Stenogyne* and *Haplostachys* switching places. Subsequent best trees had *Haplostachys* and WNA *Stachys* interchanging positions. Both of these cases are

consistent with the fourth- and fifth-best trees resulting from all genomic windows (**Fig. 3c**). For the UC chromosomes (Supplementary Fig. 16b), the topology weights were roughly between one-half to one-third that of the ENAA chromosomes, matching the SubPhaser repeats discordance pattern (Supplementary Fig. 13).

As there was some shuffling observed among the Hawaiian mints, as described above, we next performed *Twisst* based on only Hawaiian mints and their WNA *Stachys* relatives, all expected to be descendants of the same allopolyploidy even, and hence, contain the same subgenomes (see Supplementary Fig. 15b for the groupings used). We used *Twisst* with 16,652 SNP trees, each based on non-overlapping windows of 50 SNPs. Again, we found that the dominant, most represented tree was consistent with the BUSCO and SNP trees (**Fig. 3d**). In the second most represented tree, monophyly of each Hawaiian mint genus was retained, although the *Stenogyne* Kaua'i clade (indicated by *S. purpurea*) grouped with the *S. rugosa* clade instead of being sister to all other *Stenogyne*, similar to the BUSCO tree (**Fig. 2**). However, in the third-best topology, monophyly was violated, in that *Phyllostegia* became paraphyletic, consistent with the BUSCO tree.

To further explore patterns of phylogenetic conflict, we also generated a NeighborNet[53]. We uncovered a five-pointed "star" of ambiguity among *Stachys byzantina*, *S. sylvatica*, *S. coccinea*, ENAA *Stachys,* and WNA *Stachys* plus Hawaiian mints (**Fig. 3e**, Supplementary Fig. 17), edges within which are nonetheless consistent with the patterns of the hypothesized subgenome sharing. For example, *S. sylvatica* lies between *S. coccinea* and *S. byzantina* in the star, supporting it as an allopolyploid hybrid containing both *S. byzantina*-like and UC subgenomes. We also generated a NeighborNet using only Hawaiian mints to better identify conflicting signal within that lineage alone (Supplementary Fig. 17b). Supporting some of the *Twisst* topologies, *Phyllostegia* is parapyletic in the NeighborNet, unlike *Stenogyne*, which despite the phylogenetic conflict within the genus, is resolved as a monophyletic group. Finally, we used DensiTree[54] to further explore relationships among the Hawaiian genera and closest North American outgroups using subsampling of SNPs by linkage group, in one case via whole-chromosome trees, and in the second by genomic windows of 25Kb. These analyses similarly visualized significant phylogenetic incongruence among taxa (Supplementary Fig. 18). Interestingly, in some trees *Haplostachys* grouped with a subclade of *Phyllostegia*, which in some cases showed incongruent relationships to a second *Phyllostegia* subclade. Within *Stenogyne*, incongruent relationships between the Kaua'i and *S. rugosa* clades were revealed.

***Stenogyne*, *Haplostachys*, and *Phyllostegia* show contrasting genetic structure and diversity**

To further explore patterns of relationship among Hawaiian mints and close *Stachys* relatives, we used principal component analysis (PCA)[55] based on SNP data. As expected, most of the variation in the overall dataset was explained by differences among the distant *Stachys* outgroups (Supplementary Fig. 19). The first two principal components (PC1 and PC2) primarily distinguished *S. byzantina* and *S. coccinea*, respectively (Supplementary Fig. 19a). The next PC (PC3) mainly separated ENAA species from the rest. PC2 and PC3 could represent SNP diversity unique to subgenomes or ploidy levels, with PC2 representing UC and PC3 representing ENAA; indeed, *S. sylvatica,* which shares some of each subgenome, placed between the ENAA and *S. coccinea* and the "ingroup" taxa, respectively, in these two plots. To gain resolution for ingroup taxa, we subsequently removed *S. byzantina* and found that a rough cline appears shared by the rest of *Stachys* and the Hawaiian mints, one that is especially linear for Hawaiian mints and their close WNA relatives (**Fig. 4a**, Supplementary Fig. 19b). Such clinal

variation may correspond to progressive cladogenesis via geographic speciation[22]. Among the first four PCs identified, *Stachys chamissonis* and *S. bullata* were the taxa most proximate to the Hawaiian mints. Next, to more deeply analyze diversity among the Hawaiian mints, we removed *Stachys* from the PCA analysis (**Fig 4a,** Supplementary Fig. 19c). Here, *Haplostachys* was separated from the remaining Hawaiian mints in the first PC, as expected given its long branch in phylogenies and its large number of singleton/doubleton SNPs (see next paragraph). Further, in PC2, *Phyllostegia* remains a tighter cluster, whereas *Stenogyne* segregated into clusters by clade identity, including the *S. macrantha* clade, the *S. rugosa* clade, and the Kaua'i clade, the latter grouping closest to *Phyllostegia*. *Stenogyne* formed a cline along PC3, and in PC4 *Phyllostegia* largely separated into three clusters. When plotting PC2 and PC3 together and connecting the samples to the phylogenetic tree based on SNP data, the *Stenogyne* cline show a nearly perfect phylogenetic order (**Fig. 4a**). Further, the *P. warshaueri* samples shared a long branch in the SNP phylogeny, and these samples are distinct in the PCA as well. Although most samples followed a pattern consistent with phylogenetic progression, some samples appeared to be unique, such as *P. racemosa*, which tends to cluster within or near *Stenogyne*, pointing to an admixed origin (Supplementary Fig. 19c; also see below discussion). The clinal variation observed among the Hawaiian mints may correspond to progressive cladogenesis via geographic speciation[22]. For example, allelic variation may have become fixed along an ongoing cladogenetic process caused by serial founder events in an island-hopping model of geographic speciation[56]. Such relatively simple isolation-by-distance processes may be facilitated in an extremely young and rapidly expanding and dissecting volcanic environment, such as that of the Hawaiian Islands.

We further investigated the diversity among the mint genomes using various population genetic statistics. Of note, we found that *Haplostachys haplostachya* was highly unique in that it had significantly higher number of private alleles than any of the other Hawaiian mints (**Fig. 4b)** and in fact had the most doubletons (homozygous sites for a unique SNP) and the second highest level of singletons (*Stachys sylvatica* had the most singletons) among all taxa (Supplementary Fig. 20a). Assuming a single colonization event of the Hawaiian lineage, the *Haplostachys* doubletons are likely rather recent, having evolved since the split of *Haplostachys* from the rest of the Hawaiian mints, given that no other Hawaiian mint shares these alleles. An increase of rare alleles could also be associated with an ancient range expansion[57], which is consistent with the former wide distribution of *Haplostachys* across three islands (Hawai'i, Kaua'i, and Maui), although *Haplostachys* is today only found on Hawai'i[24]. An alternative hypothesis is that the Hawaiian Islands were colonized in two events, involving slightly different hybrid-polyploid lineages (as discussed above), as the number of private alleles can increase with age of the allopolyploid[39], suggesting that *Haplostachys* may represent an older sibling allopolyploidy event. Interestingly, the number of heterozygous sites for *Haplostachys* is comparable to other Hawaiian mints that in general have lower levels of heterozygosity than their mainland relatives, although *Stenogyne* species tend to have more heterozygous sites than their Hawaiian relatives (**Fig. 4c**, Supplementary Fig. 20b). These results suggest that closely related Hawaiian taxa, including *Phyllostegia* and *Stenogyne*, which both experienced extensive morphological and rapid diversification, may exhibit contrasting genetic structure and diversity levels, likely caused by different paths of radiation and subsequent isolation.

The majority of present-day Hawaiian mints appear unadmixed

To explictly test for possible signatures of admixture among our Hawaiian mint samples, we employed a multifaceted approach using the f_3-statistic[58] to compare all possible 3-way combinations of samples, as well as TreeMix[59] and ADMIXTURE[60]. Our results show that significantly negative f_3 values reflected close phylogenetic relationships among samples, as anticipated based on previous studies[16]. Crosses of negative values in our f_3 heat maps (wherein all source combinations showed negative f_3) likely reflected strong identity by descent (IBD), as also seen in other organisms, such as *Syzygium*[22] and *Ursus*[61]. For example, the two representatives of *Phyllostegia warshaueri* was marked by the cross pattern of negative values when one was configured as a target and the other as a source (Supplementary Fig. 21). We then employed TreeMix, for which a migration edge was observed between *Stenogyne sessilis* and a *S. rugosa* relative (Supplementary Fig. 22). *Stenogyne sessilis* also appeared to be admixed in the ADMIXTURE analysis (best-fitting $K = 9$; **Fig. 2**), and it appears consistently admixed across most K values (Supplementary Fig. 23). This taxon also showed extensive webbing involving *S. sessilis* and *S. rugosa* in the NeighborNet analysis (Supplementary Fig. 17), which further suggesting the former taxon may be admixed. Another possibly admixed species is *P. racemosa*, which in ADMIXTURE results was resolved as a mixture of *Phyllostegia* clades and the *Stenogyne rugosa* clade. Although this admixture signal was not identified by TreeMix or f_3-statistics, this taxon clustered closely with *Stenogyne* in the PCA (Supplementary Fig. 19c). It is also possible this could represent shared, incompletely sorted ancestral alleles. As such, it is not clear if *P. racemosa* represents a case of ILS or gene flow. Similar to *P. racemosa*, one *S. macrantha* individual appeared admixed with ADMIXTURE analysis (**Fig. 2**), but there was no strong support for this in other analyses. Intriguingly, ADMIXTURE showed *Haplostachys haplostachya* as sharing ancestral groups with the *Stenogyne* Kaua'i clade (~74%) and WNA *Stachys* (~25%), and only a small fraction with *Phyllostegia* (~1%). Such patterns, however, were not identified by f_3 or TreeMix as stemming from inter-lineage admixture, and could instead reflect ILS among WNA *Stachys* and early- and rapidly-diverging Hawaiian mint clades. It is important to note that ADMIXTURE results are often overinterpreted as indicative of cross-lineage admixture[62], and the K components from ADMIXTURE simply represent subsets of inherited SNP variation that could reflect any underlying mixtures, of which ILS may be another underlying causal factor[22].

Hybridization has been identified as one key contributor to radiations[63], and the Hawaiian mints appear to be no exception. However, the principal hybridization detected, an allopolyploidy event, occurred prior to colonization of Hawaii and subsequent diversification, with only limited interspecific gene flow occurring. Ancient hybridization, in addition to polyploidization, may instead have contributed to a rich genomic diversity among Hawaiian mint ancestors, possibly facilitating radiation in the context of ecological opportunity and founder effects.

Admixture and demographic history in a putative hybrid swarm

Because signals of ancient interlineage admixture may be confounded by ILS, we finally sought to explore admixture in recent times as a potential source of diversification in the Hawaiian lineage. We investigated a putative hybrid swarm of *Stenogyne* individuals found on the Mauna Kea volcano on the island of Hawai'i. This population has been predicted to comprise largely F1 hybrids with frequent backcrossing into *S. microphylla,* featuring a spectrum of morphological traits intermediate between *S. microphylla* and *S. rugosa*[24]. Generally, *S. microphylla* is known for its small leaves and tendency to grow as vines in *Sophora chrysophylla* trees, while *S. rugosa*

usually grows in the shade below these trees, making the latter susceptible to feral ungulate grazing on the slopes of Mauna Kea[16]. Ten *Stenogyne* individuals each were collected in 11 transect groups along the Kaaliali trail on Mauna Kea (**Fig. 5a**), and the 109 individuals were resequenced, mapped to the *S. calaminthoides* reference genome and SNPs were called as we did for other species in this study (Supplementary Table 11). We included four additional samples in our analyses: three presumed unadmixed representatives (on the basis of morphology) of *S. rugosa*, and one ostensibly unadmixed *S. microphylla*. We first generated a phylogenetic network to visualize potential conflicting phylogenetic signal in the data (Supplementary Fig. 24). The majority of the samples group with the putatively unadmixed *S. microphylla* in a tight cluster, suggesting common ancestry. However, the branches leading from this cluster to the putatively unadmixed *S. rugosa* feature extensive webbing and represent individuals primarily from group 0.1 and group 1.0, and to a lesser extent, group 0.2, at the two ends along the trail. Most of group 0.1 samples form a webbed lineage leading directly to *S. rugosa*, while group 1.0 samples appear distinct, suggesting that individuals at this location share alleles with an unsampled species. To test if a different *Stenogyne* species could fall within this group, we added all our *Stenogyne* species to the analysis and found that none of our sampled species cluster within or near group 1.0 samples (Supplementary Fig. 25). These results were consistent with phylogenetic analysis of the Mauna Kea samples, in which a clade consisting of samples from group 1.0 was distinct from *S. rugosa* (Supplementary Fig. 26). A previous study based on amplified fragment length polymorphism (AFLP) fragment data found that *S. rugosa* samples from Mauna Kea grouped with a sympatric species, *S. angustifolia*[16], which could represent the "ghost" taxon in this case.

To further explore these intial findings, we applied analyses of ADMIXTURE and PCA. ADMIXTURE indicated that most samples (group 0.4 through 0.7) represented unadmixed *Stenogyne microphylla*, while samples with admixed ancestry were found primarily in groups 0.1 and 1.0 (**Fig. 5b**, Supplementary Fig. 27). Transect groups occurring next to these latter locations (0.2, 0.9, and 1.1 groups) also showed some *S. rugosa* ancestry. In PCA (Supplementary Fig. 28), PC1 mainly separated two *S. rugosa* accessions from South Kona Forest Reserve that are recognized by having uniquely fuzzy leaves; otherwise, most principal components exhibited a tight cluster comprised of *S. microphylla* and individuals from most of the collected groups. In addition, distinct clines involving mainly two groups emerged (**Fig. 5c**): one cline consisting of group 0.1 and 0.2 samples leading to *S. rugosa* (sample #8, collected from Mauna Kea) and the second cline consisting of only 1.0 samples. Next, we employed the f_3-statistics of selected samples as target and all *Stenogyne* samples as potential sources to compare individuals that appeared admixed to those that appeared relatively unadmixed with ADMIXTURE (see Supplementary Fig. 27). Two hypothetically admixed samples (from group 0.1 and 1.0, respectively) exhibited particularly strong patterns of identity by descent (IBD) with other presumed admixed individuals from the same groups, suggesting localized allele sharing, apparently also with *S. rugosa* (Supplementary Fig. 29a and 29b). One presumably unadmixed individual from group 0.6 showed strong IBD with other individuals of suggested *S. microphylla* ancestry (Supplementary Fig. 29c), while the apparent unadmixed individual from group 0.1 showed strong IBD with the three presumably unadmixed *S. rugosa* samples but otherwise exhibited largely positive f_3 values (Supplementary Fig. 29d). Increased heterozygosity and private alleles was found among individuals from the 0.1 and 1.0 groups (Supplementary Fig. 30), which likely reflects increased allelic diversity due to introgressive hybridization.

Finally, to test for an association between the observed patterns of genetic structure with morphology, we recorded the following traits for all samples of *Stenogyne microphylla* and *S. rugosa* and their purported hybrids: trichome length, stem width, average leaf area, and average width and length ratios across three leaves for each individual (Supplementary Table 12). We tested for potential correlation among these morphological traits and principal components based on SNP data and admixture proportion from ADMIXTURE analysis (Supplementary Fig. 31). Overall, we found only weak correlation between genetic data and morphological traits. Leaf area is one of the main distinguishing features between *S. microphylla* and *S. rugosa*, with *S. microphylla* having far smaller leaves[24]. The strongest correlation between morphology and the genetic data was indeed between PC1 and average leaf area (Supplementary Fig. 32). Accordingly, the majority of samples, particularly from groups 0.3 through 0.9, had small leaves, except for the groups 0.1-0.2 and 1.0, where samples generally had larger leaves, with some even larger than the unadmixed *S. rugosa* individual (**Fig. 5c**), supporting their hybrid background.

In summary, the genetic and morphological data point to recent, albeit localized and limited, admixture on Mauna Kea between *Stenogyne microphylla* and *S. rugosa*, and possibly also a ghost taxon related to *S. angustifolia*. However, these events may well be historical and anthropogenic in origin, related to introduced ungulate herbivory, and therefore not reflective of any ancient admixture processes. It is also possible that such introgression events between closely related taxa, via limited breakdown of recent parapatry, have commonly occurred during early diversification of different Hawaiian mint lineages.

Hawaiian mints provide key insights on island plant radiations

In this study we have uncovered the most detailed evolutionary history to-date of a major Hawaiian plant radiation, the endemic mints, generating a near chromosome-level reference genome and whole-genome analyses of 23 Hawaiian mint species and 11 of their Old and New World relatives. We confirmed that the Hawaiian mints are monophyletic and most closely related to western North American *Stachys*. However, contrary to our initial expectations[16], given their hypothesized hybrid origin, our analyses demonstrate that Hawaiian mints do not appear to be highly admixed at present, except for localized introgressive hybridization on Mauna Kea. Instead, the phylogenomic incongruence we observed more likely reflects a combination of distinct subgenome evolutionary histories and/or ILS. Hawaiian mint genomes have clearly been duplicated twice since common ancestry with grapevine, with one ancient WGD shared among most Lamiales, and a more recent WGD shared between Hawaiian mints and their closest western North American *Stachys* relatives. We found strong support that this latter, most recent WGD was an allopolyploidy event, likely involving hybridization of a relative of the eastern North American and Asian *Stachys* and an unsampled diploid *Stachys* lineage related to *S. coccinea*. Additionally, we discovered that independent WGD events may be prevalent among *Stachys*, with four independent polyploidy events having occurred within our taxon sample. Despite the analytical depth of this study and apparent monophyly of the Hawaiian lineage, the number of times its members may have colonized the Hawaiian Islands remains unclear, with the possibility that the unique genus *Haplostachys* may have had an independent origin (and migration to the Hawaiian Islands) from within the same allopolyploid clade. Indeed, such an additional layer of complexity is hinted at from the excess of *Haplostachys* doubletons. Also supporting sibling allopolyploid events prior to colonization is the interchanging phylogenetic patterns for the two *Stachys chamissonis* individuals between the plastid and nuclear SNP tree, in which only one sample of *S. chamissonis* shared the plastid haplotype with

S. coccinea and Hawaiian mints, while other WNA *Stachys* did not. Future work should investigate additional presently unsampled *Stachys* species, particularly those ranging from southern North America into South America, to discover and detail the identity of the other progenitor lineage of the Hawaiian mints, a diploid relative of *S. coccinea*, possibly represented today by *S. agraria* of the "Meso-SA I" clade.

In summary, our work is consistent with a model of parapatric speciation associated with founder events, concomitant with rapid environmental changes in a dynamic volcanic landscape. The allopolyploid ancestry of the Hawaiian radiation may have set up an extensive underlying genomic diversity that could have fueled morphologically distinctions driven by drift alone. However, such rapid evolutionary radiations can set up a "nightmare scenario" for disentangling phylogenetic discordances caused by ILS, wherein allelic inheritance may not follow the cladogenetic sequence of events.

Methods

Sample collection, DNA and RNA extraction, and genome sequencing

Young leaf tissue was removed from a cultivated individual of *Stenogyne calaminthoides* and immediately weighed, flash frozen with liquid nitrogen, and placed in a -80C freezer. Roughly five grams of flash frozen tissue was used for high-molecular-weight (HMW) genomic DNA isolation. In order to enrich our extraction for nuclei, we followed the BioNano NIBuffer nuclei isolation protocol[22] in which tissue was ground in liquid nitrogen into a fine powder, then added to 10 mL of 0.2 micron filtered IBTB and incubated for 10 minutes on ice. This mixture was then strained using 100 µM followed by 40 µM filters to remove undissolved plant debris. Triton X-100 (1%) was added to lyse organelles before centrifugation at 2000 x g for 10 minutes to pellet the nuclei. Next, we followed the PacBio shared protocol "Preparing *Arabidopsis* Genomic DNA for Size-Selected ~20 Kb SMRTbell™ Libraries", beginning with an addition of 10 mL Carlson Lysis buffer and 25 µL of β-Mercaptoethanol (BME) to the pellet and incubating for two hours at 74°C, swirling every 30 minutes. Next, we performed a double extraction with chloroform/isoamyl 24:1 using equal volumes, retained the aqueous layer after centrifugation at 3200 rcf for 10 minutes and precipitated the DNA overnight at 4°C. The sample was then centrifuged for 90 minutes at 3200 rcf, washed with 70% cold ethanol, followed by another centrifugation, and the DNA left to dry for 20 minutes. Genomic DNA was further purified with a Qiagen® Genomic-Tip 500/G as follows: the DNA was resuspended in 20 mL buffer G2 and 20 µL RNase A for a 5 minute room temperature incubation. Next, 100 µL proteinase K was added and incubated for one hour at 50°C. Finally, the DNA was purified using the QIAGEN Genomic-tip 500/G and the manufacturer's instructions, precipitated overnight at 4°C with 0.7 volume of isopropanol, and washed twice with 70% cold ethanol. Once the DNA was dried, Tris-EDTA was added and the DNA was placed at 37°C for two days. The resulting HMW DNA was quantified and quality checked using a Thermo Scientific™ NanoDrop™ Spectrophotometer, a Qubit fluorometer, and agarose gel electrophoresis prior to sequencing. The reference individual was Illumina sequenced to ~235 Gb. Oxford Nanopore sequencing was performed on multiple DNA extracts using both a GridION and a PromethION at the J. Craig Venter Institute (La Jolla, CA). All Nanopore sequencing runs were then combined into a single fastq file, resulting in a total ~225 Gb sequence reads.

RNA was extracted using a Qiagen® RNeasy PlantMini Kit for three plant tissues from *Stenogyne calaminthoides*: root (R2), stem (S2) and young leaves (Y3). Library preparation and Illumina-based RNAseq was performed by NovoGeneAIT Singapore.

For Illumina resequencing of all other samples, DNA extraction from 20 milligrams of leaf tissue dried in silic gel was performed using the Qiagen® DNeasy Plant Mini Kit. All procedures followed manufacturer's instructions, except for the final step, in which two elutions of 50 µL each were used in order to increase DNA concentration. Following extraction, DNA was quantitated using a Thermo Scientific™ NanoDrop™ Spectrophotometer. NovaSeq Illumina sequencing was performed by NovoGene AIT in Singapore to ~30 Gb per sample for 45 taxa of Hawaiian mints and relatives (Supplementary Table 4) and to ~15 Gb for 110 samples of *Stenogyne rugosa* and *S. microphylla* and their purported hybrids (Supplementary Table 11).

Genome size estimation

Illumina reads and *k*-mer based estimation were used to predict the genome size of *Stenogyne calaminthoides*. Jellyfish[64] and KmerGenie[65] predicted a genome size of ~1.2 billion bases (Gb) and ~1.6 Gb, respectively. KmerGenie was used under both haploid and diploid modes. The haploid report produced a normal concave plot with a clear optimum for a *k*-value at $k = 117$. We also ran KmerGenie under diploid mode, as this mode can distinguish homozygous and heterozygous peaks.

Initial reference genome assembly and filtering

Given the large amount of Nanopore raw reads and computer memory limitations, the reads were first filtered using NanoFilt[66] such that only reads 35 Kb or longer were retained. This resulted in 77.6 Gb of reads retained, with a mean read length of 57 Kb and a mean quality score of 8.1, as calculated by NanoStat v. 1.1.2[66]. These reads were then used as input for minimap2[67] v. 2.16-r922 with flag -r 10000 and subsequently miniasm[68] v. 0.3-r179. The resulting gfa file was visualized using Bandage[69] v. 0.8.1, which showed a large "knot" of repetitive sequences, perhaps corresponding to a relatively recent LTR (Long Terminal Repeat) retrotransposon burst that had not yet diverged enough for the assembler to tease apart (Supplementary Fig. 1). We visualized transposable elements on this graph by BLAST of a library of the genome against consensus TEs reported by RepeatModeler2[70]. Furthermore, using the EDTA repeat annotation pipeline, we found that this initial genome assembly consisted of over 77% repeats, 46% of which correspond to LTRs, primarily *Copia* at 30% of the total genome, followed by unknown LTRs at 9.5%, and *Gypsy* at 7.4% (Supplementary Table 1). There was also a substantial percentage of DNA transposons, with *Helitrons* comprising 8.4% of the assembly (Supplementary Table 1).

The resulting raw assembly was polished using Racon[71] v. 1.3.3. The length-filtered Nanopore reads were mapped to the assembly using minimap2 as input for Racon and this process was repeated for a total of three rounds, each time using the most polished Racon assembly. Next, Illumina reads were mapped to the 3x Racon polished assembly using bwa mem and the resulting bam was used as input for Pilon[72] v. 1.2.3. Like Racon polishing, this process was also repeated three times.

Next, because the resulting genome size was nearly two times greater than bioinformatic estimates for genome size, purge_haplotigs[73] v. 1.1.0was used to check for and remove duplicated haplotigs. We used minimap2 to map Nanopore reads to the polished reference, and used SAMtools[74] view, sort, and index to process the sam file. The resulting plot from the

purge_haplotigs hist function showed a clear diploid peak, hence the purged version was selected for further processing. The final BUSCO score was 96.1%.

Scaffolding with Hi-C

Despite using an abundance of Nanopore reads, the assembly was still relatively discontiguous with 2874 contigs and an N50 around 650 Kb. Hence, to create a more contiguous assembly, Hi-C reads for scaffolding were generated by Arima Genomics (Arima-HiC Kit, #A410030). The Hi-C reads were mapped for scaffolding using Juicer[75] v. 1.5.7 under default conditions, to attain the merged_nodups.txt file for input to 3D-DNA[76] v. 5.0.2. 3D-DNA parameters were adjusted to 100 Kb for editing coarse resolution and 150 Kb input fragments for polisher and splitter, with an editor stringency of 45. The final touches were applied by hand using Juicebox[77] v. 1.11.08. Quality was assessed using QUAST[78] v. 5.0.1, BUSCO, and by comparing assemblies using CoGe SynMap.

Repeat annotation and gene model prediction

The EDTA (Extensive de-novo TE Annotator) pipeline[79] v. 1.8.3 was used to mask repeats in the *Stenogyne calaminthoides* genome. Using the EDTA repeat annotation pipeline, we found that this initial genome assembly consisted of over 77% repeats, 46% of which correspond to LTRs, primarily *Copia* at 30% of the total genome, followed by unknown LTRs at 9.5%, and *Gypsy* at 7.4% (Supplementary Table 1). There was also a substantial percentage of DNA transposons, with *Helitrons* comprising 8.4% of the assembly (Supplementary Table 1).

For transcriptome assembly, the raw reads from each tissue were separately assembled *de novo* using Trinity[80] v. 2.6.6 with default parameters (*k*-mer *k* = 25). Additionally, *de novo* assemblies were produced using Trans-ABySS[81] v. 2.0.1 for *k*-mers 51-111 for every alternate increment resulting in 31 assemblies for every plant tissue. All these assemblies were passed to EvidentialGene[82] v. 2017.12.21 to produce a single high-confidence transcriptome assembly for each tissue. These three final transcriptome assemblies were combined and passed once again to evigene to produce the final transcriptome assembly containing 181,789 transcripts for a BUSCO completeness score of 90.4%.

The annotation of the genome was carried out using a modular approach. The transcriptome assembly was splice-aligned against the genome assembly using PASA[83] v. 2.3.3t o produce reference ORFs. This was followed by the gene prediction step which involved a collection of hmm-based (genemark-es[84] v4.38, BRAKER[85] v. 2.1.2 with STAR aligner[86] v. 2.7.2b and AUGUSTUS[87] v3.3.2) and homology-based gene predictors (GeMoMa[88] v. 1.6.1 using model species *Arabidopsis thaliana* and *Populus trichocarpa*). The predictions from all these tools and repeats information were combined to produce a single high-confidence final gene prediction using EVidenceModeler[89] v. 1.1.1 containing 77,090 gene models with a BUSCO completeness score of 86%.

Reference genome ploidy analysis and characterization

We used the CoGe platform to analyze fractionation bias, synteny (using SynMap) with *Stenogyne calaminthoides* and its relatives. For example, the *Stenogyne calaminthoides* genome was compared with several other representative core eudicot genomes using syntenic depth FractBias[32] plots. This approach maps percentages of genes retained post-polyploidy for each subgenome of a polyploid species relative to chromosomes of a reference species. We extracted the syntenic pair results in a text table, and modified this file to be used as connections for a

Circos[90] plot. With both Circos, and CoGe SynMap, we identified which chromosomes were syntenic to each other. We further detailed the Circos plot by adding tracks for main repeat categories (*Gypsy*, *Copia*, and Helitron), as well as gene space. For all tracks, a genomic region density was added using the genomicDensity function from the R package circlize[91] v. 4.2.1 over windows of 1 Mb. Only scaffolds above 25 Mb in length were included in the Circos plot.

Chloroplast mapping and *de novo* assembly of Illumina resequenced samples

Raw Illumina reads were trimmed using Trimmomatic[92] v. 0.38, in order to remove adapter contamination. The trimmed reads for each sample were individually mapped to the *Stenogyne calaminthoides* reference genome using bwa[93] mem v. 0.7.17, and each resulting bam file was filtered for a quality score of 20 using SAMtools[74] view, and sorted using SAMtools sort v. 0.1.19. Picard MarkDuplicates v. 2.7.1 (http://broadinstitute.github.io/picard/) was used to remove PCR duplicates from the mapped reads and report mapping statistics. Depth and width of mapping coverage was calculated using the BEDTools[94] v. 2.23.0 function genomeCoverageBed. Consensus sequences were called using samtools mpileup to attain plastid sequences for phylogenetics.

To check for any potential reference bias, chloroplasts were also assembled *de novo* using trimmed reads and NOVOplasty[95] v. 3.0. A seed of the *rbcL* gene of *Stenogyne microphylla* was used (Accession AF502024.1). Phylogenies were built using RAxML[96] provided by the CIPREs Science Gateway[97] with 1000 bootstrap replicates and visualized using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

***De Novo* Nuclear Assembly of Illumina resequenced samples**

MaSuRCA[98] v. 3.2.7 was used to generate *de novo* assemblies based on Illumina data from each of the 45 mint samples. Each assembly was quality checked using QUAST[78] to determine statistics, such as the N50. Each *de novo* assembly was used as a reference for its own reads and mapped using the same pipeline used for mapping to the *S. calaminthoides* reference assembly.

Reference mapping and SNP calling of Illumina resequenced samples

The trimmed Illumina reads for each sample were individually mapped to the *Stenogyne calaminthoides* reference genome using the same pipeline used for the plastid mapping. Additionally, we calculated depth and width on a per chromosome basis, to detect potential differences in mapping to each subgenome. SNP calling was performed using GATK[99] v. 3.8 HaplotypeCaller in ERC mode for each sample to produce a g.vcf file for each sample. GenotypeGVCFs was then used to call joint genotypes and produce a combined file.

For filtration of called SNPs, GATK VariantFiltration was used with the following filter expression based on GATK recommendations: 'QD < 2.0 || FS > 60.0 || MQ < 50.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0 || SOR > 4.0'. This filtration was applied prior to all downstream analyses. Additionally, to avoid organellar contaminants and any spurious small fragments, only scaffolds greater than one megabase were used for all downstream analyses. Using VCFtools[100] a depth constraint of at least 5X coverage and no more than 500X coverage was applied to all datasets. VCFtools flags –het and –singletons were used to calculate heterozygosity and singleton/doubleton content respectively. Heterozygosity and singleton statistics were calculated independently as well for each subgenome and each of the 30 syntenic chromosomes.

BUSCO assessment and phylogenetics of genome assemblies

BUSCO[101] v. 4 was employed for each *de novo* assembly with the flag --limit 8 to retain up to 8 copies per BUSCO. For each BUSCO gene, all samples and their copies were combined into one fasta file using an in-house script. Next, these fasta files were aligned using TranslatorX[102]. To remove any low quality sections of the alignment, amino acid alignments were trimmed using trimAl[103], using the automated1 option to optimize trimming for downstream maximum likelihood analysis and the –backtrans option to convert back into nucleotides. Additionally, gaps accounting for greater than 20% were removed. Next, for each alignment a distance matrix was computer for each BUSCO gene using the DistanceMatrix function from the DECIPHER[104] package. Samples that deviated more than one standard deviation from the average were removed. This was important to filter out errors, such as bacterial genes found in the alignments. We then removed trees with less than 20 taxa. RAxML was used to generate phylogenies for each BUSCO gene. The resulting 1290 RAxML trees were combined into one file as input for ASTRAL-Pro[105]. The 1290 cleaned BUSCO trees were also used as input for GRAMPA[44], with the number of groups set to ten and all other parameters default. The best scoring tree was then opened in FigTree and nodes h1 and h2 were labelled. Next, we found the best scoring trees for the other polyploidy events.

Confirming subgenome identity

To further support the GRAMPA results and identify homeologous pairs of chromosomes, we calculated depth of coverage on a per-chromosome basis for each sample, explored DAGchainer[50] results produced by CoGe, and in parellel we ran SubPhaser[49] with the -intact_ltr option.

SNP phylogenetic analyses

Phylogenetic analyses were performed on datasets DS4, DS4a, DS4b, and DS4c (Supplementary Table 6). A maximum likelihood tree was generated using RAxML[96] v. 8.0.0 including adjustments for ascertainment bias (--asc-corr lewis) as reccomended for SNP data[106] and 500 bootstrap replicates. Trees were visualized and edited using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Additionally, SplitsTree4[107] v. 4.16.2 was used to generate a NeighborNet[53,108] network using LogDet[109] distances.

Using the R package ape[110], newick trees for the BUSCO and SNP phylogeny were converted into ultrametric trees using the function chronos. The dendextend R package[111] was used to create tanglegrams. The untangle function was used to find the best untangling method as reported with the entanglement function score. The tanglegram function was used to generate the final figures.

Furthermore, using a subset of Hawaiian mints and their closest relatives, phylogenies were generated based upon chromosomes and summarized using the DensiTree[54] function implemented in R package phangorn[112]. Additionally we generated non-overlapping windows of 25,000 SNPs, resulting in 287 trees, also summarized with DensiTree.

Twisst

Twisst[52] (topology weighting by iterative sampling of subtrees) is a method that estimates phylogenies based on sliding windows of SNPs along the genome and quantifies the contribution of a given topology to a full tree via weighting. SNP trees were prepared as recommended by the *Twisst* software and included the phasing step using beagle[113] v. 4.1. In the *Twisst* program, the

maximum number of monophyletic groups is eight so that the number of potential trees is computationally feasable. Hence, we created two datasets to address different questions. First, we used all samples, assigning a monophyletic group to *S. byzantina, S. coccinea, S. sylvatica*, ENAA *Stachys*, WNA *Stachys*, and one group for each Hawaiian mint genus. Second, we limited our analysis to the Hawaiian mints and their WNA relatives, which all have the same ploidy level. Here, we made subsets of the data into seven monophyletic clades based on the SNP tree. WNA *Stachys* made up one group, *Phyllostegia* was separated into two monophyletic groups and *Stenogyne* into three monophyletic groups based on the SNP phylogeny (partitions shown in Supplementary Fig. 15). We used 50 SNP non-overlapping windows as recommended[52]. We used the R script provided with *Twisst* to plot the results by chromosome and printed the weight below each tree.

TreeMix

We ran TreeMix[59] on DS4, DS4a, DS4b, and DS4c (Supplementary Table 6). TreeMix was run for m values 0 to 10, in each case with the --noss option included as this was run on an individual level and not a population level. In order to find the optimal m values for DS4a and DS4b, replicates of each TreeMix run were necessary, for m 1-10, five replicates of each m value were used. These replicates, if all parameters were kept the same, resulted in identical results suggesting robust results. To acquire differences in various replicates, the random seed was altered to produce differences in the results, from which standard deviation could now be calculated.

ADMIXTURE analyses

ADMIXTURE[60] v. 1.3 for *K* values 2-10 were used along with the –cv option to find the optimal (lowest cross validation) *K* value for the number of ancestral groups for datasets DS4, DS4a, DS4b, and DS4c, as well as data set HM1 (Supplementary Table 6). The ADMIXTURE results were plotted using the barplot function in R, and the cv values were plotted in ggplot2 (https://github.com/tidyverse/ggplot2).

Principal component analysis (PCA)

PCA was run under both an individual and genus level for datasets DS4, DS4a, DS4b, and DS4c, as well as dataset HM1. The eigensoft package[55] v. 6.1.3 convertf was used to convert plink .map and .ped files into .ind, .geno, and .snp files. Then, the smartpca.perl script was used to run PCA for PC1 to PC10 under default parameters. To plot the results, the ggplot2 function geom_point() was used along with the R package ggrepel[114] v. 3.5.1.

***f*$_3$ statistics**

The *f*$_3$ statistics were calculated using the admixtools[58] package with dataset DS7 as input with missing data allowed. First, using the vcftools --plink option, the vcf file was converted into ped and map files as input for convertf. Every possible combination of three samples was tested. The *f*$_3$ Z-scores were converted into p-values for correction using the p.adjust function and an FDR method[51,115] and converted back into Z-scores. We used geom_tile function in the R package ggplot2 (https://github.com/tidyverse/ggplot2) to plot the corrected Z-scores.

Morphometric analyses

The following morphological traits for all samples of *S. microphylla* and *S. rugosa* and their purported hybrids on Mauna Kea: trichome length, stem width, average leaf area, and average width and length ratios across three leaves for each sample (Supplementary Table 12). Measurements were recorded using Nikon SMZ25, a motorized multi-focus stereo microscope with a ring fiber illumination set (C-FLED2 LED light source) attached, at the Singapore Botanic Gardens. Leaf area, length, and width were calculated. Per each sample, three leaves were used at random for measurements. Additional measurements were taken, such as stem width. To test for potential correlation among morphological traits and principal components based on SNP data, principal components from SNPs were added to the measurements file and used to search for correlations using the pairs.panels function from the R package psych. Furthermore, principal components from SNPs and morphological data were input into PAST[116] v. 4.11 to create scatterplots. Correlation was tested using the Pearson correlation coefficient (r).

References

1. Gillespie, R.G. *et al.* Comparing adaptive radiations across space, time, and taxa. *Journal of Heredity* **111**, 1-20 (2020).
2. Stearns, H.T. Geology of the Hawaiian islands. (Honolulu Advertiser, 1946).
3. Price, J.P. & Clague, D.A. How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**, 2429-2435 (2002).
4. Stone, C.P. & Stone, D.B. *Conservation biology in Hawai'i*, (University of Hawaii Cooperative National Park Resources Studies Unit, Honolulu, Hawai'i, 1989).
5. Simon, C. Hawaiian evolutionary biology: an introduction. *Trends in ecology & evolution* **2**, 175-178 (1987).
6. Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nature Reviews Genetics* **18**, 411-424 (2017).
7. Meudt, H.M. *et al.* Polyploidy on islands: its emergence and importance for diversification. *Frontiers in plant science* **12**, 637214 (2021).
8. Glor, R.E. Phylogenetic insights on adaptive radiation. *Annual Review of Ecology, Evolution, and Systematics*, 251-270 (2010).
9. Schenk, J.J. The next generation of adaptive radiation studies in plants. *International Journal of Plant Sciences* **182**, 245-262 (2021).
10. Guo, C. *et al.* Phylogenomics and the flowering plant tree of life. *Journal of Integrative Plant Biology* **65**, 299-323 (2023).
11. Stallman, J.K., Funk, V.A., Price, J.P. & Knope, M.L. DNA barcodes fail to accurately differentiate species in Hawaiian plant lineages. *Botanical Journal of the Linnean Society* **190**, 374-388 (2019).
12. Lawton-Rauh, A., Robichaux, R. & Purugganan, M. Diversity and divergence patterns in regulatory genes suggest differential gene flow in recently derived species of the Hawaiian silversword alliance adaptive radiation (Asteraceae). *Molecular Ecology* **16**, 3995-4013 (2007).
13. Givnish, T.J. *et al.* Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proceedings of the Royal Society B: Biological Sciences* **276**, 407-416 (2009).
14. Givnish, T.J., Bean, G.J., Ames, M., Lyon, S.P. & Sytsma, K.J. Phylogeny, floral evolution, and inter-island dispersal in Hawaiian Clermontia (Campanulaceae) based on ISSR variation and plastid spacer sequences. *PLoS One* **8**, e62566 (2013).
15. Lindqvist, C. & Albert, V.A. Origin of the Hawaiian endemic mints within north American Stachys (Lamiaceae). *American Journal of Botany* **89**, 1709-1724 (2002).
16. Lindqvist, C., Motley, T.J., Jeffrey, J.J. & Albert, V.A. Cladogenesis and reticulation in the Hawaiian endemic mints (Lamiaceae). *Cladistics* **19**, 480-495 (2003).
17. Lindqvist, C. *et al.* An expressed sequence tag (EST) library from developing fruits of an Hawaiian endemic mint (Stenogyne rugosa, Lamiaceae): characterization and microsatellite markers. *BMC Plant Biology* **6**, 1-15 (2006).
18. Roy, T., Chang, T.-H., Lan, T. & Lindqvist, C. Phylogeny and biogeography of New World Stachydeae (Lamiaceae) with emphasis on the origin and diversification of Hawaiian and South American taxa. *Molecular Phylogenetics and Evolution* **69**, 218-238 (2013).
19. Roy, T., Cole, L.W., Chang, T.-H. & Lindqvist, C. Untangling reticulate evolutionary relationships among New World and Hawaiian mints (Stachydeae, Lamiaceae). *Molecular Phylogenetics and Evolution* **89**, 46-62 (2015).
20. Izuno, A. *et al.* Demography and selection analysis of the incipient adaptive radiation of a Hawaiian woody species. *PLoS Genetics* **18**, e1009987 (2022).
21. Choi, J.Y. *et al.* Ancestral polymorphisms shape the adaptive radiation of Metrosideros across the Hawaiian Islands. *Proceedings of the National Academy of Sciences* **118**, e2023801118 (2021).

22. Low, Y.W. *et al.* Genomic insights into rapid speciation within the world's largest tree genus Syzygium. *Nature communications* **13**, 1-15 (2022).
23. POWO, R.B.G. Plants of the world online. *Facilitated by the Royal Botanic Gardens, Kew* (2019).
24. Wagner, W.L., Herbst, D.R. & Sohmer, S.H. Manual of the flowering plants of Hawai'i. in *Manual of the Flowering Plants of Hawai'i* (University of Hawaii Press, 1990).
25. Salmaki, Y. *et al.* Molecular phylogeny of tribe Stachydeae (Lamiaceae subfamily Lamioideae). *Molecular Phylogenetics and Evolution* **69**, 535-551 (2013).
26. Scheen, A.-C. *et al.* Molecular phylogenetics, character evolution, and suprageneric classification of Lamioideae (Lamiaceae) 1. *Annals of the Missouri Botanical Garden* **97**, 191-217 (2010).
27. Jiao, Y. *et al.* A genome triplication associated with early diversification of the core eudicots. *Genome biology* **13**, 1-14 (2012).
28. Welch, A.J. *et al.* The quest to resolve recent radiations: plastid phylogenomics of extinct and endangered Hawaiian endemic mints (Lamiaceae). *Molecular Phylogenetics and Evolution* **99**, 16-33 (2016).
29. Carr, G.D. Chromosome evolution and speciation in Hawaiian flowering plants. *Evolution and speciation of island plants*, 5-47 (1998).
30. Qiao, X. *et al.* Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome biology* **20**, 1-23 (2019).
31. He, Y. *et al.* Building an octaploid genome and transcriptome of the medicinal plant Pogostemon cablin from Lamiales. *Scientific data* **5**, 1-11 (2018).
32. Joyce, B.L. *et al.* FractBias: a graphical tool for assessing fractionation bias following polyploidy. *Bioinformatics* **33**, 552-554 (2017).
33. Bird, K.A., VanBuren, R., Puzey, J.R. & Edger, P.P. The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytologist* **220**, 87-93 (2018).
34. Wendel, J.F., Lisch, D., Hu, G. & Mason, A.S. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Current opinion in genetics & development* **49**, 1-7 (2018).
35. Freeling, M. *et al.* Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Current opinion in plant biology* **15**, 131-139 (2012).
36. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679-685 (2019).
37. Hamilton, J.P. *et al.* Generation of a chromosome-scale genome assembly of the insect-repellent terpenoid-producing Lamiaceae species, Callicarpa americana. *Gigascience* **9**, giaa093 (2020).
38. Zhao, D. *et al.* A chromosomal-scale genome assembly of Tectona grandis reveals the importance of tandem gene duplication and enables discovery of genes in natural product biosynthetic pathways. *Gigascience* **8**, giz005 (2019).
39. Brandrud, M.K. *et al.* Phylogenomic relationships of diploids and the origins of allotetraploids in Dactylorhiza (Orchidaceae). *Systematic biology* **69**, 91-109 (2020).
40. Nixon, K.C. & Wheeler, Q.D. Extinction and the origin of species. *Extinction and phylogeny*, 119-143 (1992).
41. Günther, T. & Nettelblad, C. The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS genetics* **15**, e1008302 (2019).
42. Mulligan, G.A. & Munro, D.B. Taxonomy of species of North American Stachys (Labiatae) found north of Mexico. *Naturaliste canadien* **116**, 35-51 (1989).
43. Baltisberger, M. Cytological investigations of some plants from Turkey. *Willdenowia*, 225-232 (1991).
44. Thomas, G.W., Ather, S.H. & Hahn, M.W. Gene-tree reconciliation with MUL-trees to resolve polyploidy events. *Systematic biology* **66**, 1007-1018 (2017).

45. Cave, M.S. Index to Plant Chromosome Numbers for 1962. (California Botanical Society, Berkeley, 1963).
46. Paetzold, C., Kiehn, M., Wood, K.R., Wagner, W.L. & Appelhans, M.S. The odd one out or a hidden generalist: Hawaiian Melicope (Rutaceae) do not share traits associated with successful island colonization. *Journal of Systematics and Evolution* **56**, 621-636 (2018).
47. Soltis, D.E. *et al.* Polyploidy and angiosperm diversification. *American Journal of Botany* **96**, 336-348 (2009).
48. Chester, M., Riley, R., Soltis, P. & Soltis, D. Patterns of chromosomal variation in natural populations of the neoallotetraploid Tragopogon mirus (Asteraceae). *Heredity* **114**, 309-317 (2015).
49. Jia, K.-H. *et al.* SubPhaser: a robust allopolyploid subgenome phasing method based on subgenome-specific k-mers. *New Phytologist* **235**, 801-809 (2022).
50. Haas, B.J., Delcher, A.L., Wortman, J.R. & Salzberg, S.L. DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics* **20**, 3643-3646 (2004).
51. Salojärvi, J. *et al.* Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. *Nature genetics* **49**, 904-912 (2017).
52. Martin, S.H. & Van Belleghem, S.M. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429-438 (2017).
53. Bryant, D. & Moulton, V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular biology and evolution* **21**, 255-265 (2004).
54. Bouckaert, R.R. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372-1373 (2010).
55. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904-909 (2006).
56. Rundell, R.J. & Price, T.D. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends Ecol. Evol.* **24**, 394–399 (2009).
57. Cubry, P., Vigouroux, Y. & François, O. The empirical distribution of singletons for geographic samples of DNA sequences. *Frontiers in genetics* **8**, 139 (2017).
58. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093 (2012).
59. Pickrell, J. & Pritchard, J. Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings*, 1-1 (2012).
60. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* **19**, 1655-1664 (2009).
61. Lan, T. *et al.* Insights into bear evolution from a Pleistocene polar bear genome. *Proceedings of the National Academy of Sciences* **119**, e2200016119 (2022).
62. Lawson, D.J., Van Dorp, L. & Falush, D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature communications* **9**, 1-11 (2018).
63. Grant, P.R. & Grant, B.R. Hybridization increases population variation during adaptive radiation. *Proceedings of the National Academy of Sciences* **116**, 23216-23224 (2019).
64. Marcais, G. & Kingsford, C. Jellyfish: A fast k-mer counter. *Tutorialis e Manuais* **1**, 1-8 (2012).
65. Chikhi, R. & Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **30**, 31-37 (2014).
66. De Coster, W., D'hert, S., Schultz, D.T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666-2669 (2018).
67. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).
68. Li, H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103-2110 (2016).
69. Wick, R.R., Schultz, M.B., Zobel, J. & Holt, K.E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350-3352 (2015).

70. Flynn, J.M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**, 9451-9457 (2020).
71. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome research* **27**, 737-746 (2017).
72. Walker, B.J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one* **9**, e112963 (2014).
73. Roach, M.J., Schmidt, S.A. & Borneman, A.R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC bioinformatics* **19**, 1-10 (2018).
74. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
75. Durand, N.C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell systems* **3**, 95-98 (2016).
76. Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92-95 (2017).
77. Durand, N.C. *et al.* Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell systems* **3**, 99-101 (2016).
78. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072-1075 (2013).
79. Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome biology* **20**, 1-18 (2019).
80. Haas, B.J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* **8**, 1494-1512 (2013).
81. Robertson, G. *et al.* De novo assembly and analysis of RNA-seq data. *Nature methods* **7**, 909-912 (2010).
82. Gilbert, D.G. Genes of the pig, Sus scrofa, reconstructed with EvidentialGene. *PeerJ* **7**, e6374 (2019).
83. Haas, B.J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic acids research* **31**, 5654-5666 (2003).
84. Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O. & Borodovsky, M. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic acids research* **33**, 6494-6506 (2005).
85. Hoff, K.J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-genome annotation with BRAKER. in *Gene prediction* 65-95 (Springer, 2019).
86. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
87. Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic acids research* **34**, W435-W439 (2006).
88. Keilwagen, J., Hartung, F. & Grau, J. GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. in *Gene Prediction* 161-177 (Springer, 2019).
89. Haas, B.J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome biology* **9**, 1-22 (2008).
90. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome research* **19**, 1639-1645 (2009).
91. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. Circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811-2812 (2014).
92. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
93. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**, 1754-1760 (2009).
94. Quinlan, A.R. BEDTools: the Swiss-army tool for genome feature analysis. *Current protocols in bioinformatics* **47**, 11.12. 1-11.12. 34 (2014).
95. Dierckxsens, N., Mardulyn, P. & Smits, G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic acids research* **45**, e18-e18 (2017).

96. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
97. Miller, M., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop, 2010, 1–8. (2010).
98. Zimin, A.V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669-2677 (2013).
99. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303 (2010).
100. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).
101. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. & Zdobnov, E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
102. Abascal, F., Zardoya, R. & Telford, M.J. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic acids research* **38**, W7-W13 (2010).
103. Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).
104. Wright, E.S. Using DECIPHER v2. 0 to analyze big biological sequence data in R. *R J.* **8**, 352 (2016).
105. Zhang, C., Scornavacca, C., Molloy, E.K. & Mirarab, S. ASTRAL-Pro: quartet-based species-tree inference despite paralogy. *Molecular biology and evolution* **37**, 3292-3307 (2020).
106. Leaché, A.D., Banbury, B.L., Felsenstein, J., De Oca, A.N.-M. & Stamatakis, A. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic biology* **64**, 1032-1047 (2015).
107. Huson, D.H. & Bryant, D. Estimating phylogenetic trees and networks using SplitsTree 4. *Manuscript in preparation, software available from www. splitstree. org* (2005).
108. Levy, D. & Pachter, L. The neighbor-net algorithm. *Advances in Applied Mathematics* **47**, 240-258 (2011).
109. Lockhart, P.J., Steel, M.A., Hendy, M.D. & Penny, D. Recovering evolutionary trees under a more realistic model of sequence evolution. *Molecular biology and evolution* **11**, 605-612 (1994).
110. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526-528 (2019).
111. Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718-3720 (2015).
112. Schliep, K.P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592-593 (2011).
113. Browning, S.R. & Browning, B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics* **81**, 1084-1097 (2007).
114. Slowikowski, K. ggrepel: Automatically position non-overlapping text labels with "ggplot2.". *R package version 0.8. 0 ed* (2018).
115. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289-300 (1995).
116. Hammer, Ø., Harper, D.A.T. & Ryan, P.D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4**(2001).

Figures

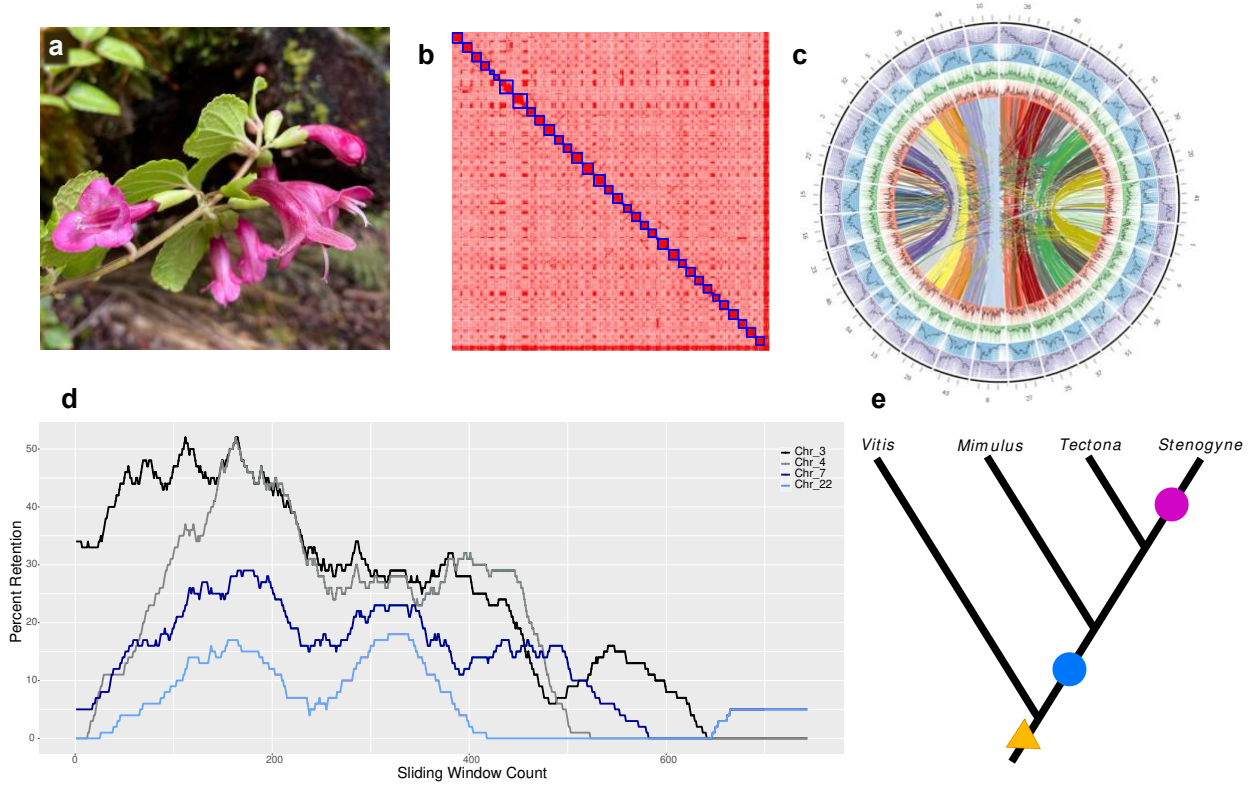

Fig. 1: Assembly and structural evolution of the *Stenogyne calaminthoides* genome. **a** *Stenogyne calaminthoides* showing flowers and leaves (©Jacey Savage, iNaturalist 214210157, CC BY-NC). **b** Hi-C contact map with inferred chromosomes in blue boxes. **c** Gene and repeat landscape of *S. calaminthoides* genome with outer to inner tracks showing genes (purple), *Copia* (blue), Helitron (green), and *Gypsy* (red) repeats, respectively. Density over 1 Mb intervals is shown by black lines. Regions of synteny are shown in the center and each pair shares a color. Chromosome number is shown on the exterior. **d** Fractionation bias plot of *S. calaminthoides* mapped onto *Vitis vinifera* chromosome 10. Each coloured line represents a scaffold. **e** Schematic phylogenetic tree showing inferred polyploid events in *Stenogyne*, with the triangle representing the gamma paleohexaploidy found in all core eudicots, the blue circle representing a whole genome duplication found in most core Lamiales, and the pink circle representing the WGD event exclusive to *Stenogyne*.

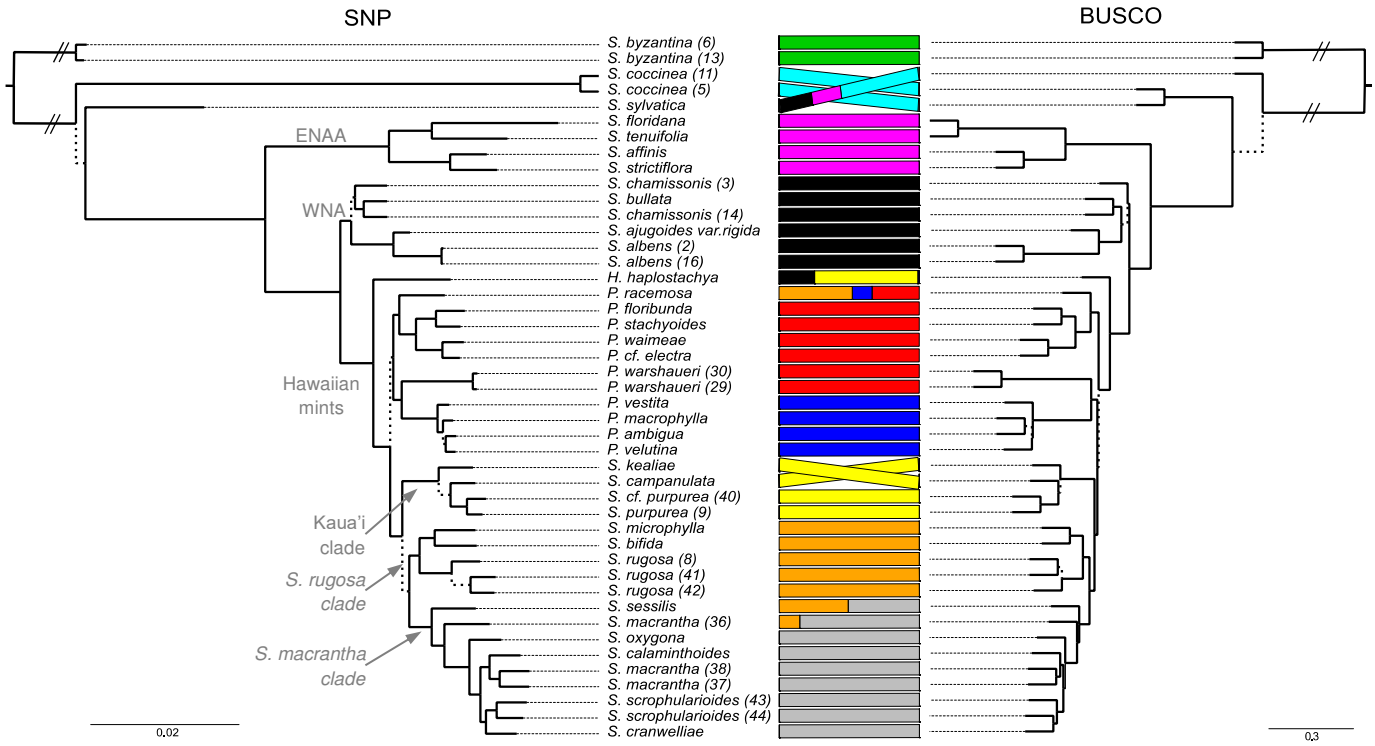

Fig 2: Phylogenetic relationships and admixture among Hawaiian mints and relatives. The phylogenetic trees are based on single nucleotide polymorphism (SNP) data (left) and BUSCO single copy nuclear genes (right). Clade names used in the text are indicated on the SNP tree: ENAA (eastern North American and Asian *Stachys*), WNA (western North American *Stachys*). Discordance between the two trees is marked with a tanglegram and dotted branches. ADMIXTURE results between the trees with each color representing a separate ancestral population, *K* (best-fitting $K = 9$ is shown).

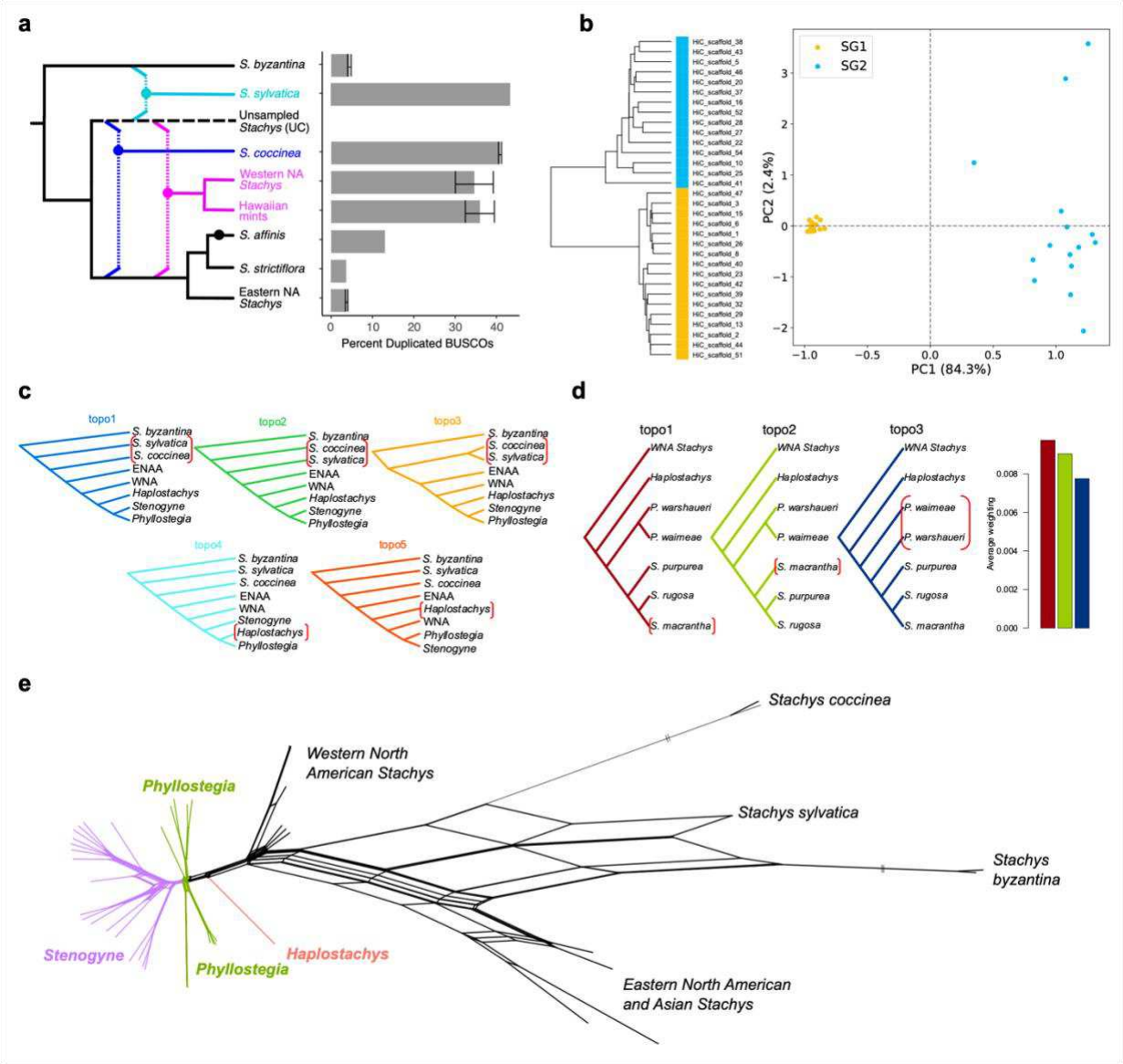

Fig. 3: Allopolyploid history and phylogenetic conflict among Hawaiian mints and relatives. **a** Schematic tree showing hypothesized allopolyploid events within this dataset (the circles represent polyploid events). The percent of duplicated BUSCO genes is shown to the right of the tree. Timing and order of ploidy events are not to scale. **b** PCA of kmers from reference *Stenogyne calaminthoides* inferred from SubPhaser. SG1 represents the ENAA subgenome and SG2 represents the UC subgenome. **c** The top five topologies found with *Twisst* analyses of dataset DS4 (all taxa included). Taxa that move among trees are highlighted with red brackets. See Supplementary Fig. 15a for details of the groupings used. **d** The top three trees found with *Twisst* analyses based only on Hawaiian mints and WNA *Stachys* relatives. Taxa that move among trees are highlighted with red brackets. The bar plot shows the average weighting values for each tree. See Supplementary Fig. 15b for details of the groupings used. **e** Phylogenetic network (NeighborNet) of all samples. See Supplementary Fig. 17 for complete network with branches labelled.

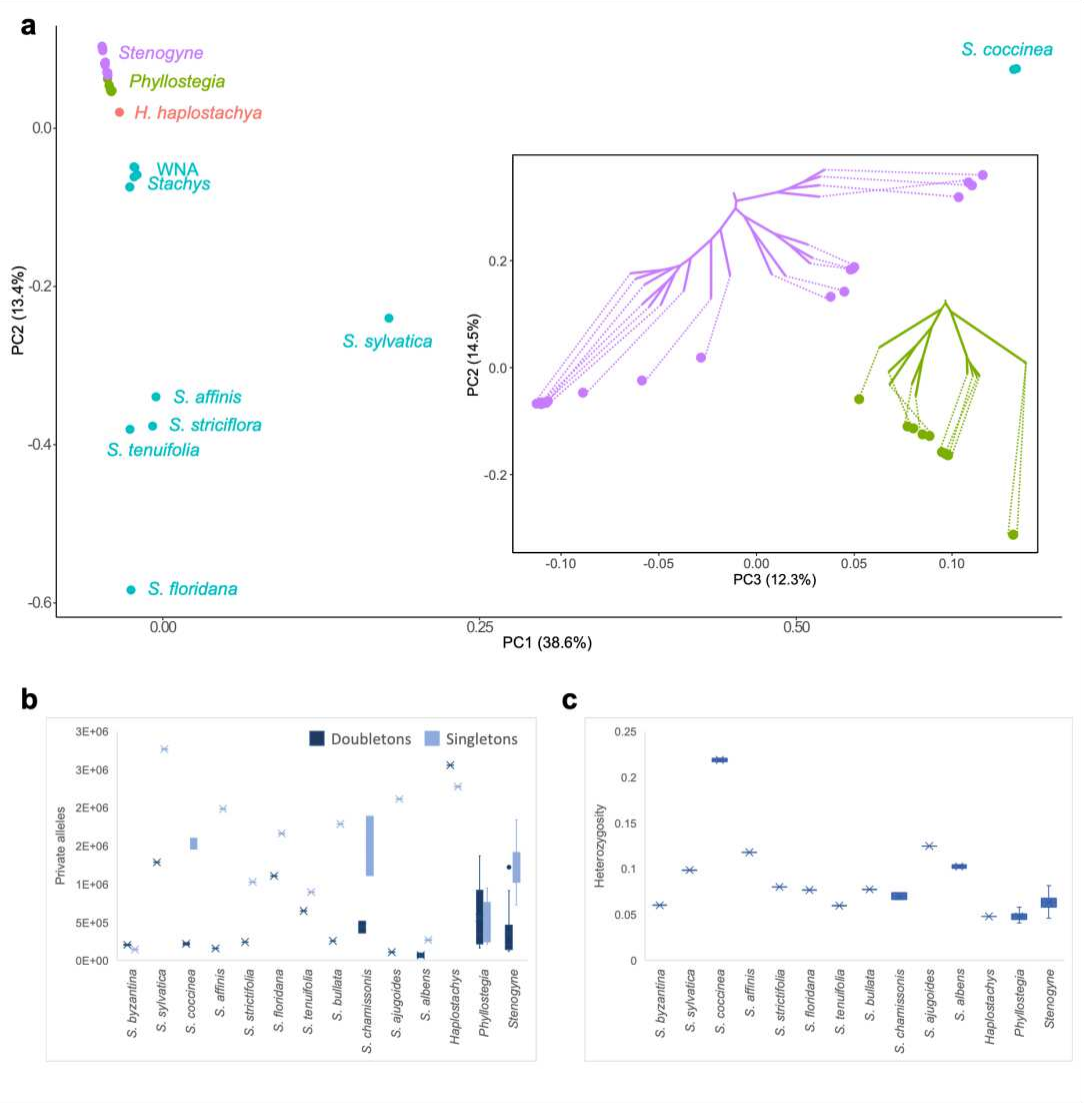

Fig. 4: Genetic structure and diversity among Hawaiian mints and relatives.
a Principal component analysis (PCA) of dataset DS4c (outgroup *S. byzantina* excluded). The insert shows a PCA plot of *Stenogyne* (purple) and *Phyllostegia* (green) with dotted lines connecting them to the phylogenetic tree based on SNP data. **b** Singletons and doubletons (homozygous sites for a unique SNP). Individuals for each Hawaiian genus, respectively, and representatives of the same *Stachys* species were combined. **c** Heterozygosity based on SNP dataset DS7 (Supplementary Table 6). Individuals for each Hawaiian genus, respectively, and representatives of the same *Stachys* species were combined.

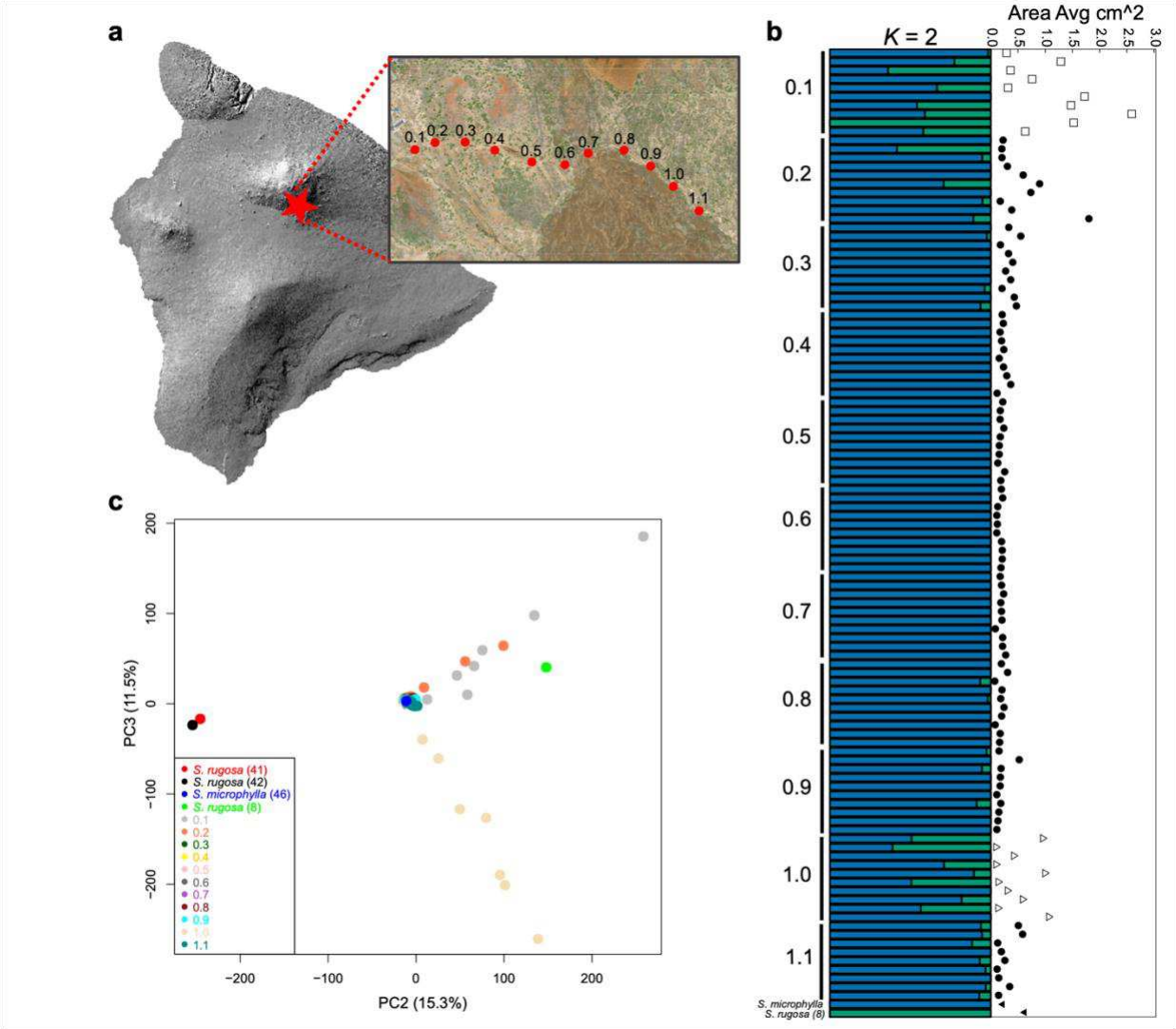

Fig. 5: Admixture and demographic history in a putative hybrid swarm of *Stenogyne microphylla* and *S. rugosa*.

a Map of Hawai'i showing the location of the hybrid swarm on Mauna Kea (red star), with the Kaaliali trail zoomed in. The dots represent the 11 sites sampled along the Kaaliali trail, each with ten *Stenogyne microphylla* and purported hybrid individuals collected. **b** ADMIXTURE results, with best-fitting $K = 2$, is shown to the left, and the average leaf area measured from each individual is displayed to the right. **C** Principal component analysis showing PC2 and PC3.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MintGenomePaperSupplementaryTables130723Submission.xlsx
- MintGenomePaperSupplementaryFigsALL130723Submission.pdf