Title

The ecological and genetic basis of floral scent differentiation in the orchid genus Gongora

Permalink

https://escholarship.org/uc/item/9qf5c33q

Author

Guizar Amador, Maria Fernanda

Publication Date

2022

Peer reviewed|Thesis/dissertation

The ecological and genetic basis of floral scent differentiation in the orchid genus *Gongora*

By

MARIA FERNANDA GUIZAR AMADOR
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Population Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Santiago Ramirez, Chair

_____

Graham Coop

_____

Philipp Zerbe

Committee in Charge

2022

Contents

List of Figures

List of Tables

Abstract of the Dissertation

The ecological and genetic basis of floral scent differentiation in the orchid genus *Gongora*

Pollinator-mediated selection on floral traits is thought to have influenced the evolution and diversification of angiosperms. Orchids exhibit highly specialized pollinator associations that are thought to promote and maintain reproductive isolation between sympatric lineages. However, the mechanisms by which angiosperms adapt to and shift among different pollinators remain poorly understood. Around 10% of neotropical orchid species are pollinated by scent-collecting male euglossine bees. In this system, floral scents simultaneously attract and reward bee pollinators. Here I used male euglossine bee pollinated orchid species from genus *Gongora* to determine the underlying mechanisms of pollinator attraction and reproductive isolation between sympatric lineages.

In Chapter 1 I examined the variation in volatile organic compounds (VOCs) emitted by the inflorescences of *Gongora* plants found in two natural populations from Costa Rica. Floral scent varied discretely in both populations with each chemically distinct group (chemotype) attracting a different subset of euglossine species. To test whether these differences in pollinator attraction contribute to reproductive isolation, we genotyped pollen masses recovered directly from male bees caught in the field to reconstruct pollinator networks and perform population genetic analyses. The results of this chapter revealed one population to be structured by chemotype despite varying levels of pollinator overlap. The second population showed little evidence of genetic differentiation between chemotypes despite high degrees of floral trait and pollinator attraction divergence. These findings suggest that natural selection may be driving the observed patterns of divergence in floral phenotypes despite gene flow.

To further explore the genetic basis of floral scent biosynthesis and differentiation, in Chapter 2 I assembled and annotated the reference genome of one of the characterized *Gongora* chemotypes. Terpenoids are the most common class of compounds found in the scent of male euglossine bee pollinated orchids and their biosynthesis and diversity is mainly mediated by terpene synthases (TPS). I manually annotated a total of 21 TPS genes in

Gongora. This number is comparable to the number of TPSs found in other sequenced orchid genomes. Finally, I sequenced individuals from two sympatric chemotypes with low levels of genomic differentiation to explore patterns of genomic differentiation across the genome. Genome-wide differentiation was low with several small and scattered regions showing high levels of differentiation. These results suggest that the genetic architecture might be complex involving many loci of small effect.

In conclusion, our results demonstrate the capacity of floral scent to finely control specific pollinator attraction, pollination network architecture, and their role in mediating reproductive isolation. The genome assembly will serve as the foundation for future research aimed at understanding the evolution of floral scent.

ACKNOWLEDGMENTS

being excellent role models. Finally, to my partner, Jorge Eder, although it has not always been easy and it has required a lot of sacrifice, I want to thank you for being my biggest supporter and for sharing this journey with me.

Chapter 1

Floral scent chemistry controls pollinator specificity and reproductive isolation in sympatric cryptic lineages of *Gongora* orchids

Abstract

Insect pollinators have profoundly influenced the diversification of angiosperms by exerting selective pressures on floral traits. Pollinator specialization, for example, can promote and maintain reproductive isolation among plant populations. However, the mechanisms by which angiosperms adapt to and shift among different pollinators remain poorly understood. Because pollinator attraction is often mediated by multiple sensory modalities (visual, chemical, tactile) few studies have successfully revealed links between variation in individual floral traits and the evolution of pollinator-mediated reproductive isolation. Here, we take advantage of a powerful chemical signaling system to determine the underlying mechanisms of pollinator attraction and reproductive isolation in a specialized plant-pollinator mutualism. *Gongora* orchids emit floral scents to attract and reward male euglossine bees in exchange for pollination services. Male euglossine bees collect and store these scents in hind-leg pockets to concoct perfumes used subsequently during courtship display. Thus, floral scents simultaneously attract and reward bee pollinators. To test whether floral scent modulates reproductive isolation among two populations of *Gongora* orchids, we reconstructed pollina-

tor networks by genotyping pollen masses recovered directly from male bees caught in the field. We identified seven cryptic scent chemotypes that are tightly associated with non-overlapping assemblages of bee pollinators. Our results demonstrate the capacity of scents to finely control specific pollinator attraction, pollination network architecture, and their role in mediating reproductive isolation.

Introduction

Plant-pollinator interactions have persisted for nearly 120 million years. Approximately 90% of angiosperms ($\sim$300,000 spp.) rely on insect pollinators (mostly bees) for sexual reproduction [96], and virtually all species of bees ($\sim$16,000 spp.) obtain nectar, pollen or other resources from flowering plants [89].

It has long been hypothesized that pollinators have shaped the evolution of numerous angiosperm lineages. Grant and Grant [53] and Stebbins [121] developed a model that provides a conceptual framework for the study of pollinator-driven ecological plant speciation. Under this model, geographic variation in pollinator communities facilitates the action of divergent natural selection on floral traits. Reproductive isolation among plant populations emerges when plant lineages adapt to the most effective local pollinator, leading to the formation of pollination ecotypes. According to this model, pollination ecotypes will show variation in the floral traits that mediate pollinator attraction and pollination efficiency; whereas pollinators will exhibit different levels of preference for these variable floral traits [53, 121]. Shifts in pollinator association (*e.g.*, a shift from bee to hummingbird pollination) often occur in parallel with modifications of floral traits (*e.g.*, a shift from short to long corolla) [14], and both factors may contribute to pollinator-mediated speciation. However, little is known about the mechanisms underlying the origin and maintenance of these pollination ecotypes, partly due to the complexity of most species interactions and the difficulty of isolating relevant traits [67, 71, 127, 128].

Speciation requires the establishment of gene-flow barriers among formerly interbreeding populations. Reproductive barriers in plants may be classified according to when they act along their life cycle. Pre-pollination (early-acting) barriers prevent or reduce pollen transfer between plant populations and are often mediated by shifts in flowering phenology

[113], geography or pollinators [52, 68, 71]. Post-pollination (late-acting) barriers evolve via diverse genetic mechanisms including hybrid unviability, hybrid sterility or reduced hybrid fitness (hybrid breakdown)[19, 109]. Multiple lines of evidence indicate that early-acting barriers arise more frequently and evolve faster than late-acting mechanisms and therefore contribute more to the total reproductive isolation among plant lineages in the early stages of divergence [12, 85, 109]. In particular, shifts in pollinator association, a type of early-acting barrier, occur frequently across the phylogeny of angiosperms and account for as much as ∼25% of speciation events [127].

Angiosperms have evolved numerous adaptations to regulate pollinator attraction and pollinator specificity. Floral traits—including morphology, color, and scent—collectively function as advertisement signals to lure, manipulate and reward animal pollinators in exchange for pollen transfer. Floral scent is a major component of both pollinator attraction and pollinator specificity, and it is particularly important in specialized plant-pollinator mutualisms [115]. More than 1,700 volatile organic compounds (VOCs) have been identified in floral scents emitted by ∼100 plant families [76, 104]. Most known VOCs belong to the terpenoid (5-carbon monoterpenes, 10-carbon sesquiterpenes) and aliphatic compound classes, all of which typically exhibit low molecular weight and high vapor pressure [76]. The chemistry of floral scent often correlates with the attraction of specific types of pollinators. For example, the oxygenated aromatic compounds (*e.g.*, methyl benzoate, benzyl alcohol) that are emitted in floral scents of numerous night-blooming plant species often elicit the innate attraction of nocturnal hawkmoth pollinators [110]. Thus, scent chemistry may function as a private communication channel that regulates pollinator attraction and pollinator specificity [104].

Among land plants, orchids possess some of the most remarkable adaptations for animal pollination [40]. In tropical America, ∼700 species of orchids exhibit specialized mutualistic associations with euglossine bees [36, 40]. Euglossine-pollinated orchids produce floral scents that attract and reward male euglossine bees. In exchange, bees visit these orchid flowers to collect scent compounds, which they store and accumulate in specialized hind-leg pockets. Male bees use these scents to concoct species-specific perfume mixtures, which they subsequently expose to females during an elaborate courtship display [44]. Orchid species from

3

different genera usually attach their pollinaria (pollen masses) on different parts of a bee's body (mechanical isolation); whereas, closely related orchid species are pollinated by different species of euglossines (ethological isolation)[105]. Differential attraction of euglossine bee species could be mediated by the floral scent emitted by each orchid species [36, 44].

The genus *Gongora* is one of the most diverse lineages of euglossine-pollinated orchids, with approximately 60 species distributed across neotropical lowland rainforests [60, 66]. All species of *Gongora* share the same pollination mechanism [36]. While collecting scents from a *Gongora* flower, a male bee clings precariously upside-down on the underside (adaxial surface) of the labellum (a specialized enlarged petal) while brushing and transferring scent compounds to the hind-leg pockets. The waxy surface and bristles of the labellum eventually force the male bee to slip and slide down the column similar to "a child on a toboggan" [6, 126]. As the bee slides down, the pollinarium becomes attached to the bee's dorsal surface (mesoscutellum) by means of a sticky viscidium. Subsequently, while repeating this behavior at a receptive flower— one in which the pollinarium has been previously removed—the pollinarium is deposited in the stigma and the pollination cycle is completed [60]. All species of *Gongora* emit strong scents and lack additional floral rewards (*e.g.*, nectar, pollen) that would typically attract other types of pollinators [36, 137]. Thus, a single floral trait, scent, appears to regulate pollinator attraction and pollinator specificity in this specialized mutualism.

We hypothesize that variation in floral scent chemistry regulates both pollinator specialization and reproductive isolation among closely related and sympatric *Gongora* orchids. To study this, we tested whether orchids from two populations in Costa Rica exhibit discrete phenotypic variation in scent chemistry, potentially corresponding to cryptic lineages. Then we tested whether this variation sufficiently mediates the attraction of unique sets of bee pollinator species. Finally, we determined whether shifts in pollinator associations correlate with genetic differentiation among *Gongora* lineages. Our study therefore combines multiple approaches including chemical analysis of floral scents, behavioral tests with bee pollinators, and population genetic analyses of co-occurring *Gongora* populations.

Materials and Methods

Plant Collection

Plants were collected from the surroundings of the La Gamba and La Selva research stations. In both sites we potted the plants and grew them in the stations until the inflorescences became available for scent and bee visitor observations. Plants were fertilized with Osmocote once a year.

Chemical Analysis of Floral Scents

We analyzed the chemical composition of *Gongora* floral scents using Gas Chromatography-Mass Spectrometry (GC-MS). Our approach aimed to elucidate the diversity of volatile compounds emitted by flowers in order to identify potential cryptic species and populations. To extract floral scent from an inflorescence, we implemented a static headspace method in both the greenhouse and in the field. We used scent traps made with clear glass tubing (2.4mm ID, 3.5cm length) and filled with 20mg of bulk carbide and 20mg of Tenax GC (Supelco, Bellefonte, PA, USA; mesh size 60/80). We eluted scent compounds by injecting 200μL of hexane into the scent traps. Extracts were stored at -20$^{\text{o}}$C before GC-MS analyses. We analyzed scent extracts as described in [61] using an Agilent 7890B GC fitted with a 30m x 0.25mm x 0.25mm HP-5 Ultra inert column, coupled to an Agilent 5977A mass spectrometer (Agilent Technologies). Relative proportions (%) were calculated by dividing the absolute amounts of individual compounds by the sum of all compounds.

We used multivariate statistical methods to investigate the variation of chemical profiles between individual plants and chemotypes. We normalized raw peak areas by estimating their relative area and calculated pairwise distances among individual orchid samples using the Bray-Curtis dissimilarity metric in R from the package *ecodist* v1.2.2 [50]. We used the dissimilarity matrices to perform a non-metric multidimensional scaling (nMDS) analysis using the function metaMDS from package *vegan* [95].

Bee Visitors Observations

We reconstructed the bee-visitor network of each *Gongora* chemotype present in our two sites using the R package *bipartite* v2.05 [37]. Between 2013 and 2019 and between 2016 and 2019 we documented visitation rates and the diversity of euglossine bee species visiting *Gongora*

inflorescences at La Gamba and La Selva, respectively. Immediately after extracting the floral headspace, we relocated each plant to a nearby forest site to observe pollinator behavior during the morning hours (9:00am to noon). Flowers from the same *Gongora* inflorescence begin anthesis simultaneously at dawn and remain open for three days before wilting. Thus, we conducted both headspace sampling and bee-visitor observations on the first and second day of anthesis. Whenever possible, multiple observations were conducted on the same individual plant if multiple inflorescences were produced. Male bees were allowed to land at least once on the inflorescence to perform scent-collecting behavior before being captured. All collected bees were preserved in ethanol and pinned for species identification.

Sample Genotyping

We collected and genotyped pollinaria from 130 plants, 85 from La Gamba and 45 from La Selva. Additionally, we used synthetic attractants as baits to collect bees carrying *Gongora* pollinaria. The collected bees were also preserved in ethanol and pinned for species identification. DNA was extracted from the pollinaria using DNeasy Plant Mini kits (Qiagen) and we implemented Genotype by Sequencing (GBS) following previously established protocols [43]. We pooled 95 samples that we genotyped using a single Illumina HiSeq lane. The resulting sequencing reads were screened for Single Nucleotide Polymorphism (SNPs) using the UNEAK pipeline [16] with error tolerance rates set to 0.03. We used *vcftools* v.0.1.15 [33] to filter out low quality SNPs and individuals with not enough data from our data set before proceeding with downstream analyses.

We used ADMIXTURE v.1.3.0 [4] to estimate ancestry in each population set of individuals. The results were combined with chemotype information and bee species identifications to reconstruct high resolution pollinator networks in R using the package *bipartite* v2.05 [37]. We calculated mean heterozigosity at segregating sites using the R package *dartR* v1.1.11 [56]. We also used the filtered SNPs to perform principal component analysis using the R package *adegenet* v2.0.1 [69]. To identify potential admixture between populations we calculated f3, f4 and D-statistics using ADMIXTOOLS [99] together with the R package *admixr* [100]. Fst between pairs of chemotypes was estimated using the R package *hierfstat* [51]. To estimate the evolutionary relationships between samples from all 7 chemotypes from both

populations we used RAxML [119].

Results

Study Populations

We collected a total of 174 adult *Gongora* plants from La Gamba and 110 plants from La Selva. Of these, 63 and 36 individuals respectively bloomed and were used for floral scent extraction, 48 and 39 plants were used for bee visitor observations and 84 and 48 plants were used for population genetic analyses. Flower coloration, in both populations, was consistently uniform within and among the inflorescences produced by any single plant. At La Gamba, morphological characters did not match previous species descriptions [66] and the absence of discrete diagnostic vegetative or floral traits made species delimitation challenging. At La Selva, the plants we collected matched the descriptions from two *Gongora* species previously reported in the region: *G. quinquenervis* and *G. unicolor* [11, 36].

Floral Scent Variation Reveals Cryptic Orchid Lineages

To assess the level of variation in floral scent chemistry, we extracted and analyzed floral blends emitted by 54 different plants from La Gamba and 43 from La Selva using Gas Chromatography-Mass Spectrometry (GC-MS). We detected a total of 33 distinct compounds belonging to the diverse phenylpropanoid, monoterpene and sesquiterpene compound classes. Analyzing both data sets using non-Metric Multidimensional Scaling (nMDS) revealed the presence of four distinct chemical groups at La Gamba, hereafter referred to as chemotypes A, M, S and X (Figure 1.1 A); and three distinct chemical groups at La Selva, referred to as P, R and T (Figure 1.1 B). We examined the amount of within-individual variation among replicate samples and found that repeated measurements from the same plant consistently produced nearly identical results. Thus, instrument error and individual-level variation were negligible compared to the variation observed between chemotypes.

La Gamba Chemotypes

We found that the composition of each chemotype is unique (ANOSIM R=0.928,p=0.001, minimum stress value=0.07) despite the overlap of some compounds between chemotypes (cineole, linalool, α-pinene, β-pinene, terpinen-4-ol and β-elemene). Chemotype A emits
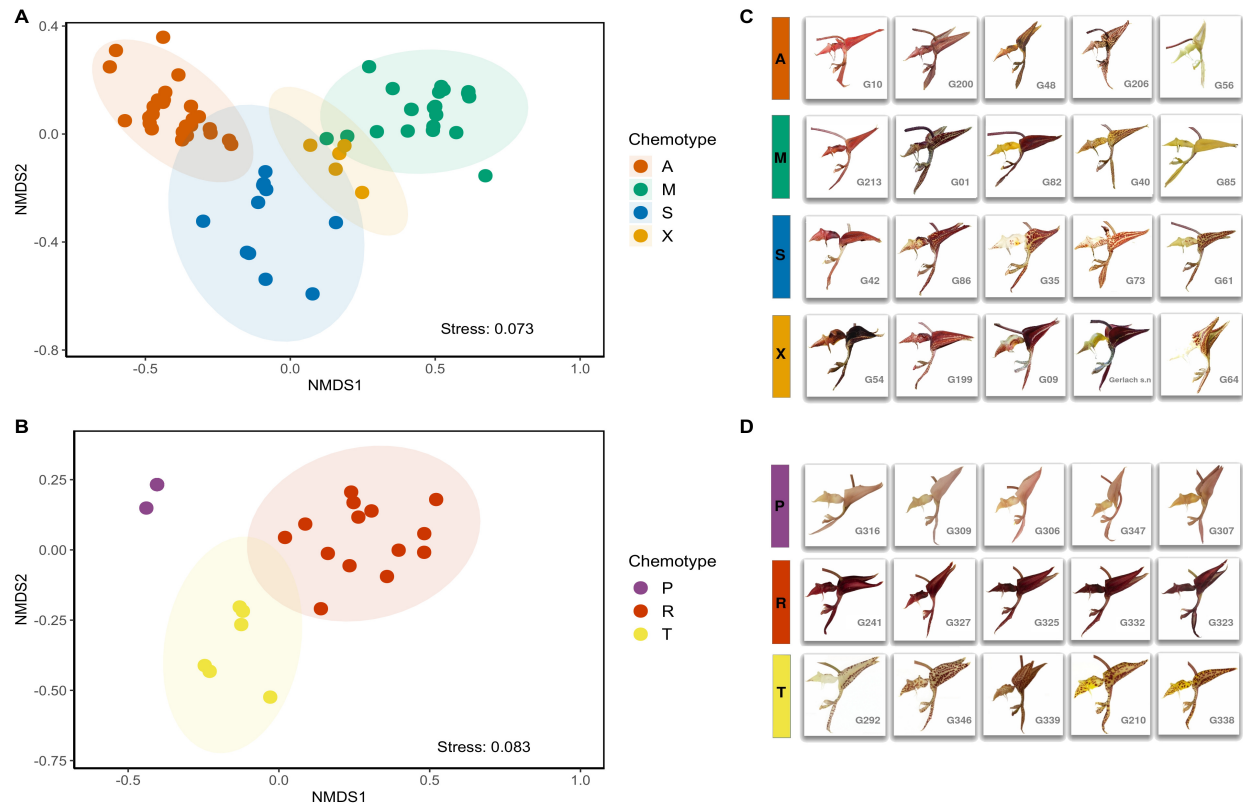
Figure 1.1: Cryptic *Gongora* chemotypes identified via floral scent chemistry. non-Metric Multi-dimensional Scaling (nMDS) plots based on multivariate GC-MS data revealed discrete clustering at La Gamba (**A**) and La Selva (**B**). Color plates depict the highly variable color patterns exhibited by each chemotype at La Gamba (**C**) and the discrete floral color variation between chemotypes at La Selva (**D**).

mostly phenylpropanoid derivatives, chemotype S emits both monoterpene and sesquiterpene compounds, and chemotypes M and X emit mostly monoterpene compounds (Table A.1).

Flower coloration varied continuously within chemotypes and ranged from dark solid mauve to solid pale yellow with numerous intermediate color forms displaying speckles of varying sizes (Figure 1.1 C). Three chemotypes (A, M, X) exhibited similar amounts of color variation. Chemotype X flowers ranged from a white background with large mauve speckles to an entirely white-yellow labellum. Careful observation, however, revealed subtle morphological differences of floral traits that allow differentiation between chemotypes S and X. Chemotype S exhibited the least amount of color variation.

La Selva Chemotypes

We also found that the composition of each chemotype is unique despite chemotypes R and T showing substantial overlap in their main scent compounds: estragole and eugenol (ANOSIM R=0.968, p=0.001, minimum stress value=0.08)(Figure 1.1 B). Chemotype T emits trans- and cis-methyl-methoxy-cinnamate, which are completely absent from chemotype R floral scents; while chemotype R emits methyleugenol and p-methyl-anisole which are absent from T. Chemotype P plants emitted a strong floral scent in the field, but we failed to characterize it. We were only able to detect the presence of two unknown compounds in very small amounts (Table A.2).

In contrast to the La Gamba chemotypes, the La Selva chemotypes are easily identified by their floral color and there is little floral color variation within each chemotype (Figure 1.1 D). Chemotype P is characterized by solid pale pink flowers, chemotype T shows bright yellow flowers with dark mauve speckles and chemotype R flowers are solid dark mauve.

Floral Scent Mediates the Attraction of Bees

To investigate whether each chemotype attracts unique sets of bee pollinators, we conducted field observations to estimate the diversity of bee visitors and their visitation rates. Between 2013 and 2019, we observed 48 *Gongora* plants from La Gamba, and 39 plants from La Selva between 2016 and 2019. Due to its rarity, no direct observations were possible for Chemotype X. We documented a total of 403 bee visitors (269 at La Gamba and 134 at La Selva) belonging to 23 bee species and three genera, namely *Euglossa* (Eg.), *Eulaema* (El.), and *Exaerete* (Ex.) (Figure A.1). A bipartite network analysis revealed that in both populations, each chemotype attracts mostly unique assemblages of bees (Figure 1.2 A,B). The resulting bee-orchid visitation networks exhibited a highly compartmentalized architecture, with each network cluster corresponding to a specific orchid chemotype (Figures A.3 and A.4). In order to estimate the amount of bee visitor overlap between chemotypes we calculated the proportional similarity (PS)[116] of visitor specificity for each pair of chemotypes as:

$$PS = 1 - \frac{1}{2} \sum_{i=1}^{n} |V_{ai} - V_{bi}|$$

Where n is the total number of bee species visiting *Gongora* in a population, and V_{ai} and V_{bi} are the percentage of individual bees from species i that visit chemotypes a and b,
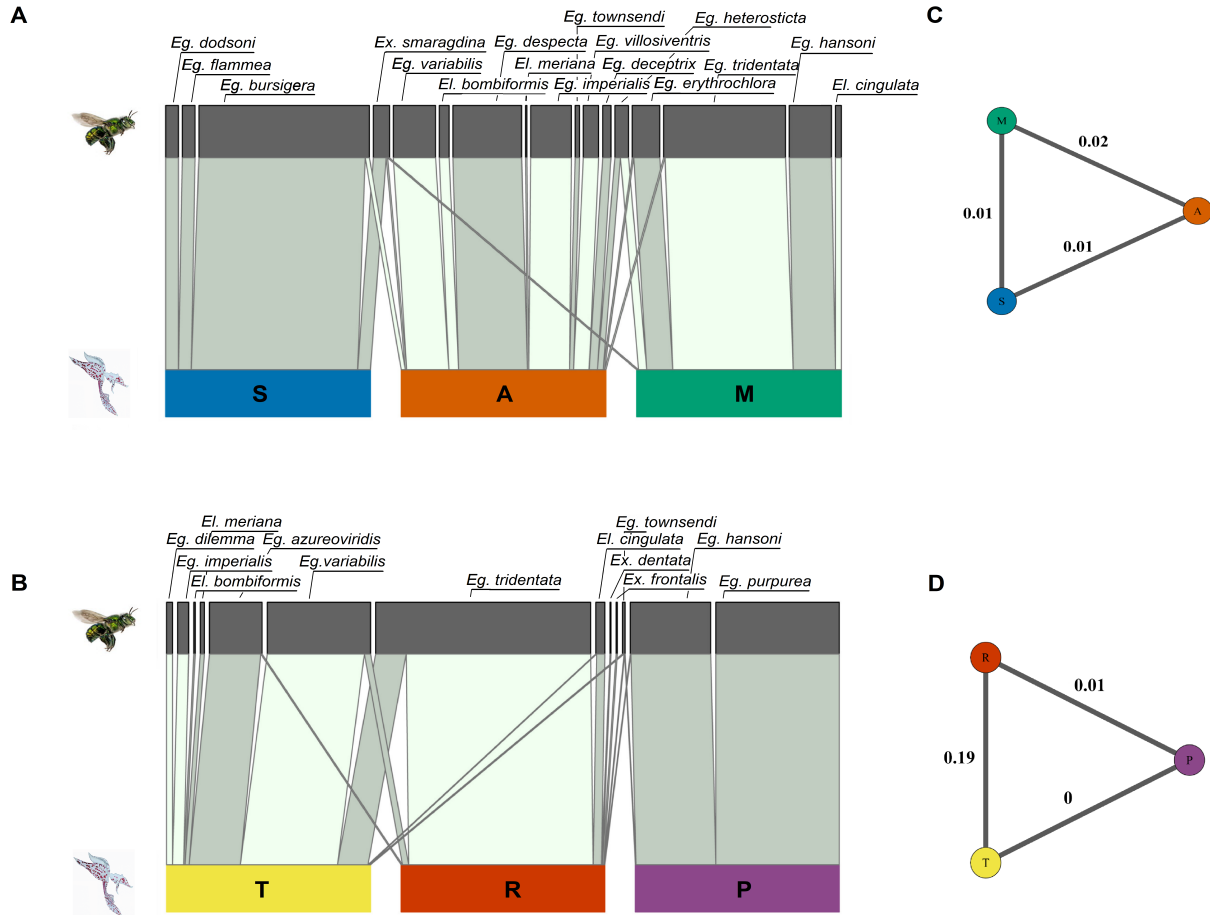
Figure 1.2: Bipartite bee-orchid visitation networks based on direct observation of bee visitors of plants from chemotypes A, M and S from La Gamba (**A**) and from chemotypes P, R and T from La Selva (**B**). Bee visitor overlap between pairs of chemotypes from La Gamba estimated by the Proportional Similiarity (PS) of visitor specificity for La Gamba (**C**) and La Selva (**D**). Direct observation of chemotype X from La Gamba was not possible due to its rarity.

respectively. This index goes from 0 to 1, with 0 being no bee visitation overlap and 1 being complete bee visitation overlap. All chemotype pairs in both populations exhibited low PS ranging from 0 between chemotypes P and T, to 0.19 between chemotypes R and T (Figure 1.2 C,D).

Scent Chemotypes Correspond to Genetically Differentiated Lineages

To test whether scent chemotypes correspond to genetically differentiated lineages, we conducted population genetic analyses based on two sampling strategies. First, we genotyped

individual plants for which we also acquired scent chemical data. Second, we genotyped pollinaria samples that we recovered directly from male bees caught in the field. Because in all orchids the male staminal filaments and the female style are fused into a single structure known as the column (or gynostemium), the anatomical match required for a pollinator to remove the pollinarium is identical to that required for its subsequent delivery at a receptive flower [40]. Therefore, a bee species capable of removing and carrying a pollinarium from an orchid species can be considered its true pollinator [106]. We reconstructed high-resolution bee-orchid pollination networks by genotyping pollinaria recovered directly from male bees by applying Genotype-by-Sequencing (GBS) [43]. In total, we genotyped 84 adult plants from La Gamba and 48 from La Selva, and 386 and 95 *Gongora* pollinaria recovered directly from male bees caught in the field (Figure A.2).

At La Gamba, our population genetic analysis revealed four genetically distinct lineages that precisely correspond to the four chemotypes that we identified based on floral scent chemistry and bee visitation patterns (Figure 1.3 A). Consistent with the relative placement in the PCA analysis (Figure A.6) and with our phylogenetic analysis (Figure A.8), overall genetic differentiation revealed that chemotypes A and M are more closely related to each other (FST=0.053) than they are to chemotypes S and X, which also appear to be closely related to each other (FST=0.074) (Table A.3). The mean heterozygosity at segregating sites was highest for the M chemotype (Ho=0.095), followed by A (Ho= 0.084), S (Ho=0.077), and X (Ho=0.06)(Table A.4).

In contrast to La Gamba, our population genetic analyses did not identify the La Selva chemotypes as being genetically distinct from each other. Chemotypes R and T did not form distinct genetic clusters in the ADMIXTURE plot (Figure 1.3 B), they were indistinguishable from each other in the PCA analysis (FigureA.7) and their overall genetic differentiation was low (FST= 0.007)(Table A.6). However, chemotype P formed its own genetic group in the ADMIXTURE plot and it formed a monophyletic group in the phylogenetic analysis (Figure A.11), despite showing low overall genetic differentiation from the other chemotypes (FST=0.07 and 0.05 for R and T, respectively). Chemotype P also showed the lowest mean heterozygosity at segregating sites (Ho=0.085) compared to chemotypes R (Ho=0.1) and T (Ho=0.13) (Table A.7).
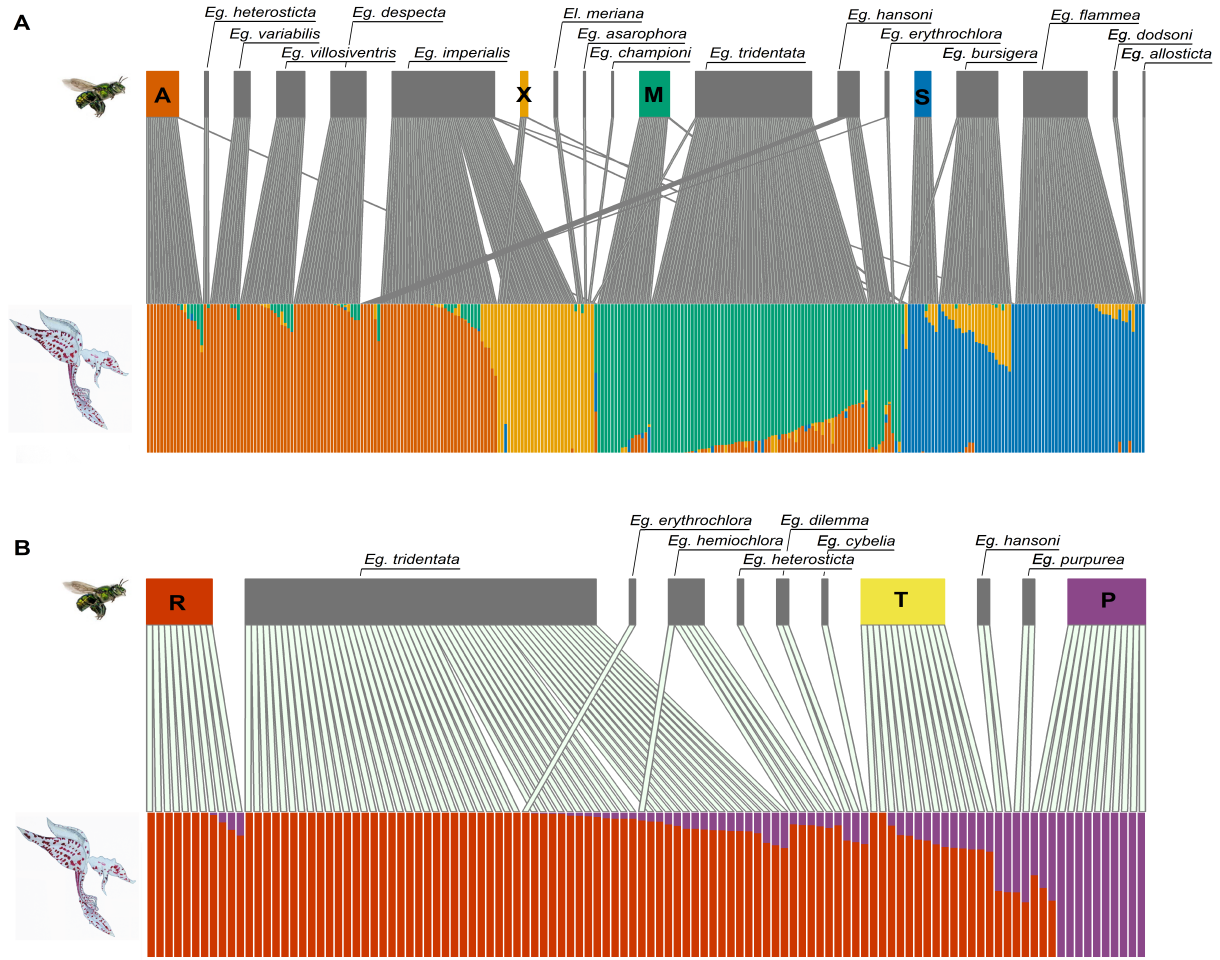
Figure 1.3: Population genetic analyses based on *Gongora* pollinaria recovered directly from plants and from male bees in the field. (**A**) Admixture analysis (k=4, based on 71,948 loci) showing distinct populations based on pollinaria samples that match the four floral scent chemotypes from La Gamba. (**B**) Admixture analysis (k=2, based on 79,674 loci) based on pollinaria samples from La Selva. Both Admixture plots are coupled with bipartite pollinator networks with each line representing a bee-orchid pollinator observation. Samples obtained directly from plants are indicated based on the chemotype they belong to.

To estimate patterns of introgression among populations, we used a 3-population test of admixture. We found that chemotype A exhibits strong signatures of admixture from both chemotype S and M (f3=-0.006, Z=-9.632), but no other chemotypes at La Gamba (Table A.3) or La Selva (Table A.8) showed evidence of admixture.

Scent Chemotypes and their Association with Bee Pollinators

By combining genotype data from plants with known floral scent chemistry and pollinaria recovered directly from male bees, we were able to reliably assign individual pollinaria sam-

ples to each chemotype. This approach allowed us to reconstruct true pollination networks of unprecedented sample size and resolution for both populations (Figure 1.3).

At La Gamba, our survey of bees carrying pollinaria identified a total of 17 species of true pollinators of *Gongora* belonging to genera *Euglossa* and *Eulaema* (Figure 1.3 A). Similar to the pattern we observed in the visitation network, we found that the true pollination network exhibits a highly compartmentalized architecture where each chemotype is tightly associated with a unique assemblage of bee pollinators. The pollination network closely resembled the visitation network (Figure 1.2).

To estimate the pollinator overlap between sympatric chemotypes, we computed the proportional similarity (PS) of pollinator assemblages [116] for each pair of chemotypes as:

$$PS = 1 - \frac{1}{2} \sum_{i=1}^{n} |P_{ai} - P_{bi}|$$

Where n is the total number of bee species pollinating *Gongora* at La Gamba, and P_{ai} and P_{bi} are the percentage of individual bees from species i that carried pollinaria from chemotypes a and b, respectively. This index goes from 0 to 1 with 0 being no pollinator overlap and 1 being complete pollinator overlap. Similar to the values for the PS visitation overlap, pollinator overlap between the La Gamba chemotypes was close to 0; however, with the addition of chemotype X data we see that the pollinator overlap between A and X is the highest between any two chemotypes at 0.4 (Figure A.5).

Additionally, we genotyped 12 pollinaria from La Gamba recovered from male bees that were captured while visiting orchid flowers for which we also analyzed scent chemistry. We ensured that these pollinaria were not removed from the observation plants but instead were removed and carried by bees from plants in the field. In all 12 cases, the chemotype assignment of the pollinaria matched the chemotype assignment of the plant, therefore revealing no cross-visitation between chemotypes.

At La Selva we identified a total of 12 species of true pollinators of *Gongora* belonging to genus *Euglossa* (Figure 1.3 B). We were unable to detect modularity in the pollinator network. Known visitors of chemotype P (*Eg. purpurea*, *Eg. hansoni*), represent only 5 out of the total of 95 bees collected carrying pollinaria at La Selva. However, four of them carried pollinaria that clustered with the P chemotype samples in our ADMIXTURE analysis. Due

13

to the lack of modularity in the pollinator network from La Selva we were unable to estimate pollinator overlap between chemotypes P, R and T.

Discussion

In the present study we took advantage of a powerful chemical signaling system to test some of the predictions laid out by the Grant-Stebbins model [52, 53, 121]. Species of the neotropical orchid genus *Gongora* exhibit specialized pollinator associations with male euglossine bees which are attracted to the volatile organic compounds emitted by the flowers [36]. We identified discrete variation in floral scent composition between sympatric *Gongora* lineages, referred to as chemotypes, and found that bee species display different levels of attraction between them. Our field observations and population genetic analysis suggest that each chemotype is pollinated by a distinct set of euglossine species, corresponding to pollination ecotypes with varying levels of pollinator overlap.

From our field observations and pollinaria genotyping, we were able to reconstruct visitation and pollination networks of unprecedented size and resolution. The modularity of the networks suggests that each chemotype is occupying a distinct pollination niche by attracting a unique set of bee species (Figure 1.2), which could result in speciation if the reduction of pollinator sharing between chemotypes is strong enough to mediate reproductive isolation [29, 52, 59, 127]. At La Gamba, the species of euglossines found carrying pollinaria from any given chemotype corresponded to the species observed visiting plants with that scent profile (Figure 1.3 A). This result confirms that most bee species visiting the flowers for scent collection are also pollinators, even though we did not observe pollination happening in the field. The visitation network for La Selva consists of three modules corresponding to the three floral scent chemotypes (Figure A.4), but according to our population genetic analysis, there is little to no population structure and genetic differentiation between chemotypes was low (Table A.6). Despite this result, floral scent variation was discrete, and we did not observe any plants with intermediate phenotypes in either population, suggesting that selection is maintaining the phenotypic differentiation.

We identified a total of 28 different volatile organic compounds in the floral scent, with chemotypes producing on average 6 different compounds (highest 8 and lowest 2)(Tables

A.1 and A.2). These compounds are not taxonomically specialized, and they have all been previously reported in other euglossine-pollinated plants [18, 36, 48, 61]. Compounds within a chemotype tend to belong to one biosynthetic pathway (terpenoids or phenylpropanoids), suggesting that differences in floral scent are mainly due to variation in gene expression rather than gene coding mutations.

Euglossine bees exhibit species-specific preferences for volatile compounds and several species can co-exist within a population [1, 18, 36]. General attractants, such as 1,8-cineole, attract the largest amount of bee species, while other compounds may attract only one or a few species. When these general attractants are present in a mix, the presence of additional compounds usually reduces the number of bee species attracted to a chemotype [36, 136], which could increase pollinator specificity [90]. This hypothesis is supported by our observations of chemotypes R and T from La Selva, where the two main components in both floral scents are eugenol and estragole. It is the presence of additional compounds that makes the floral scent of each chemotype distinct (Table A.2). This difference in floral scent appears to result in the differential attraction of orchid bee species. *Euglossa variabilis* and *Eg. tridentata* were frequently observed visiting both chemotypes; however, 90% of *Eg. tridentata* visitors were caught on chemotype R plants, and only 10% were caught visiting chemotype T. This proportion is completely reversed for *Eg. variabilis*, where 90% of visitors were caught on chemotype T and only 10% on chemotype R. These differences in visitation frequency could play a role in maintaining assortative mating between both chemotypes, especially if they work in concert with other reproductive isolation barriers. In other sympatric interfertile plant lineages with pollinator overlap, reproductive isolation has been shown to be maintained through a combination of ethological isolation and postmating reproductive barriers [19, 26, 52, 73, 141].

In contrast to chemotypes R and T, chemotypes A and X from La Gamba exhibit completely distinct floral scent profiles -A emits mostly phenylpropanoids and X emits mostly terpenoids- but they share *Eg. imperialis* as their main pollinator. Our population genetic analysis suggests that there is ongoing gene flow between A and X; however, genetic differentiation between them remains high (Fst=0.26). Chemotype A appears to be more common at La Gamba than chemotype X (22 chemotype A plants flowered and were observed for bee

15

visitors, compared to only 1 chemotype X plant). This result suggests that, in androeuglossophilous plants, pollinator convergence is possible even in the case of complete floral scent divergence.

The results from La Selva and La Gamba lead us to believe that floral isolation is not the only mechanism involved in maintaining reproductive isolation in *Gongora*. Floral isolation could be acting together with other isolating factors, such as micro-habitat differences [26, 72], hybrid inviability, or reduced hybrid fitness [19, 109], to reduce gene flow between chemotypes. However, because reproductive isolating mechanisms operate sequentially, and pre-pollination barriers act early during the reproductive cycle (i.e., pollen transfer)[12, 85], the pollinator specialization that we detected is likely to make a significant contribution to the total reproductive isolation among sympatric lineages. Hence, we hypothesize that pollinator specialization, regulated by scent chemistry, has played a central role in the speciation of *Gongora*.

The mechanisms by which floral scent differentiates and results in pollinator transitions are not understood yet. Since reproductive success in *Gongora* orchids relies on the attraction of euglossine bees through floral scent, we expect selection to favor the floral scent profile that promotes visitation by the most frequent and effective pollinators [121]. Previous work has shown that the effectiveness of volatile compounds in attracting orchid bees varies geographically and temporally [1, 36]. These spatial differences in orchid bee preferences could be shaped by factors such as the community of co-flowering plants, competitors, and predators [1, 55, 92, 114], and have the potential to impose divergent selection on floral scent, resulting in pollination ecotypes [49, 52, 53, 55, 68]. However, even in highly specialized pollination systems, other agents, such as herbivores, pathogens, and abiotic factors can also exert selection on floral traits [3, 113, 122]. Geographic and temporal variation in non-pollinator agents of selection could influence the evolution of novel chemotypes even when pollinator assemblages do not differ [10].

In both study sites, we observed the presence of florivorous weevils (Coleoptera, Curculionoidea) on *Gongora* inflorescences. The insects feed and mate on floral tissue, while females oviposit in the floral buds. As the larvae develop, they feed on the labellum and can sometimes cause the abortion of an entire inflorescence (personal observation). If florivory

results in a high fitness cost for the plant, the weevils could potentially exert strong selection on floral traits including scent, but more research is needed to understand the role of these insects in floral scent evolution. Finally, we cannot exclude the role of other non-adaptive evolutionary processes that could influence floral scent, including genetic drift [125], gene flow [93], and hybrid introgression [108].

Since scent alone can attract pollinators, we do not believe that floral color is playing a major role in the orchid-pollinator interaction. Color appears to be correlated with floral scent at La Selva but not at La Gamba. Color variation at La Gamba could be due to a shared ancestral polymorphism that has not yet been fixed in any chemotype due to gene flow, or it could also be maintained by diversifying selection. At La Selva, floral color could have been fixed within chemotypes due to genetic drift, selection, or it could be in linkage disequilibrium with floral scent. It is important to note that the community of pollinators, competitors, and antagonists differs between La Gamba and La Selva, so the adaptive landscape for both floral scent and coloration are likely to vary.

In conclusion, we have shown that in this highly specialized plant-pollinator mutualism, discrete variation in floral scent is associated with pollinator attraction and specificity, contributing to the maintenance of pre-pollination gene-flow barriers between sympatric lineages. We suspect that these microevolutionary processes are pervasive in *Gongora*, and other euglossine-pollinated plants, and that they have resulted in a macro-evolutionary pattern of adaptive radiations [71, 79, 127]. To understand the evolution and maintenance of these pollination ecotypes, future studies should focus on investigating the number and distribution of loci involved in floral scent differentiation. QTL analyses of floral traits involved in pollinator shifts from other systems have found that they involve a few large-effect mutations [9, 17, 75], which could be rapidly fixed by selection if they allow a lineage to occupy a new phenotypic optimum [10, 97]. Because we observe sympatric chemotypes with floral profiles of different pathways, we suspect that these genetic differences will mainly involve regulatory regions.

Chapter 2

The *Gongora* genome assembly provides new resources and insights to understand floral scent evolution

Abstract

Orchidaceae is one of the largest flowering plant families with many species exhibiting highly specialized reproductive and ecological adaptations. An estimated 10% of neotropical orchid species are pollinated by scent-collecting male euglossine bees; however, to date there are no published genomes of species with this pollination syndrome. Here we present the first draft genome of a neotropical epiphytic orchid from genus *Gongora*, a representative of the male euglossine bee pollinated subtribe Stanhopeinae. The 1.83 Gb *de novo* genome with a scaffold N50 of 1.7Mb was assembled using a combination of short- and long-read sequencing and chromosome capture (Hi-C) information. A total of 20,496 protein-coding genes were annotated and 83.36% of the genome was identified as repetitive content. We identified and manually annotated 21 terpene synthase (TPS) genes and performed a phylogenetic analysis with other published orchid TPS genes. Finally, we sequenced the genomes of 13 individual plants belonging to two closely related sympatric *Gongora* lineages and performed population genetic analyses to identify possible genomic islands of differentiation. The *Gongora* genome assembly will serve as the foundation for future research aimed at understanding the genetic basis of floral scent biosynthesis and diversification in orchids.

Introduction

With more than 25,000 species found in nearly all terrestrial habitats, Orchidaceae is one of the largest and most widespread families of angiosperms [41]. The highly specialized ecological and reproductive strategies of many orchid species may have contributed to the family's high speciation rates [30, 139] and successful adaptation to distinct environments [35, 150].

A striking example of reproductive adaptations are the more than 700 species of neotropical orchids that are pollinated by euglossine bees (Apidae; Euglossini) [39, 48, 105]. *Gongora* is one of the at least 22 orchid genera that exhibit specialized mutualistic associations with scent-collecting male euglossine bees (also referred to as orchid bees). Male bees pollinate *Gongora* plants while visiting inflorescences to collect volatile organic compounds (VOCs), which they store in hind-leg pockets for later use during courtship display [6, 44].

All *Gongora* species rely exclusively on male euglossine bees for sexual reproduction and lack any additional floral rewards like nectar or edible pollen. *Gongora* plants emit species-specific floral bouquets which typically consist of one to three main compounds with an additional one to ten minor compounds. The most common volatiles found in the flowers are monoterpenoids, sesquiterpenoids, and aromatic compounds [36, 133, 136, 137]. We have previously shown (see Chapter 1) that differences in the floral scent profile between sympatric species lead to the attraction of different sets of pollinators, mediating the extent of gene flow and maintaining reproductive isolation barriers.

Gongora contains between 60 and 70 recognized species; however, the taxonomic delimitation and systematics of the genus are notoriously difficult because multiple cryptic species, with little to no morphological variation, can coexist in a population and are only discernible by their floral scent profiles [38, 66, 133](Chapter 1). Based on extensive ecological research, several authors have hypothesized that pollinator-driven diversification has played a major role in the evolutionary history of *Gongora* [39, 60, 105, 136]; however, the molecular and genetic mechanisms underlying the origin and maintenance of reproductive barriers among *Gongora* lineages remains largely unexplored.

Generating high-quality genomic resources for *Gongora* is needed to elucidate the genetic basis of floral scent production and how divergent floral scent phenotypes evolve and lead to

the evolution of reproductive barriers. The sequencing and *de novo* assembly of a reference genome can help us to identify candidate genes involved in the biosynthesis and regulation of floral volatile compounds. To date, there are no reference genomes available for any male euglossine bee pollinated orchids, and the lack of a reference genome is a major obstacle towards studying the ecology and diversification of these neotropical orchids.

In this study, we report the genome of a chemotype A *Gongora*, the second largest genome of any orchid assembly reported so far (after *Cymbidium goeringii*). The assembly was constructed using a hybrid strategy combining Illumina HiSeq, PacBio Single Molecule Real-Time (SMRT) and Hi-C sequencing technologies. The estimated *Gongora* genome size is 2.228 Gb. We found that the genome possesses a large number of repeat sequences compared to other orchid genomes, but comparable to *C. goeringii* [27]. In addition, I conducted a high quality annotation of the terpene synthase (TPS) genes present in the genome of *Gongora*, which lays the foundation for further research on floral scent biosynthesis. I also analyzed 13 genomes from 2 different sympatric *Gongora* chemotypes (cryptic species that are only differentiated on the basis of floral scent) and performed comparative genomic analyses to identify regions of elevated differentiation.

Materials and Methods

Genome Assembly

Sample preparation and sequencing

All materials used for the genome assembly were obtained from a mature *Gongora* A-chemotype plant collected from the surroundings of the La Gamba Tropenstation in the province of Puntarenas, which is located in southwestern Costa Rica. The sample was imported to the USA under the CITES Certificate of Scientific Exchange permit 14US51372B/9 and is currently located in the Botanical Conservatory at the University of California, Davis. For genome sequencing, we collected fresh leaves and flash froze them in liquid nitrogen. Two tissue samples were sent to Dovetail Genomics (Santa Cruz, CA, USA) for the construction and sequencing of 2 Illumina libraries (Illumina HiSeq X, insert sizes 402 bp and 523 bp), 1 PacBio Sequel library (5 SMRT cells, 7,876 bp average read length), 1 Hi-C library (llumina HiSeq X) and 1 Chicago library (Illumina HiSeq X) (Table B.1). An additional sample was

used for DNA extraction with a DNEasy Plant Mini Kit from QIAGEN, followed by library construction and sequencing using Illumina HiSeq 2000 platform. In total, we generated 412 Gb of raw reads that were then filtered according to sequencing quality and adapter contamination.

Genome size estimation

To estimate genome size and heterozygosity, we analyzed the k-mer frequency distribution from the 402 bp insert size Illumina library with Jellyfish [88]. We also estimated genome size with flow cytometry. Briefly, a piece of fresh orchid leaf ($1.5cm^2$) was chopped with a fresh single edge razor blade in a cold Galbraith buffer along with a similar sized piece of fresh leaf from either tomato (*Solanum lycopersicum*, 1C = 1320.3) or pea (*Pisum sativum*, 1C = 4591.71). The released nuclei were then filtered and stained with a cold solution of 25 mg/mL prodidium iodide for 30 minutes in the dark. We quantified the relative fluorescence of 2C orchid and 2C standard nuclei using a Beckman Coulter Cytoflex flow cytometer. Ploidy level was determined by the relative position of the 2C orchid and 2C standard peaks and by the estimated genome size based on the ratio of the 2C peak positions of the sample and standard times the amount of DNA in the standard.

Genome assembly

Given the high levels of heterozygosity and repetitive content in the *Gongora* genome, we decided to use a hybrid strategy for the assembly. Long reads can improve the contiguity of an assembly; however, this technology is associated with high error rates [147]. We used FMLRC [131] with one of our Illumina libraries (SRCD2_S1_L00 Table B.1) to leverage the higher accuracy of the short reads to perform long-read error correction on the PacBio library. Before proceeding with the assembly, we used BLAST [8] to identify long reads not belonging to the nuclear genome by comparing against the *Oncidium* chloroplast (GQ324949.1) and mitochondrial (KJ501920.1) sequences downloaded from NCBI [132]. The resulting 85,553 reads were left out of the nuclear genome assembly and used separately to assemble the organelle genomes.

With the corrected and filtered PacBio reads as input, we used WTDBG2 [112] for the

de novo assembly. One of the short read libraries (SRCD2_S1_L00) was then mapped to the contigs using BWA [81]. Contigs with different levels of coverage were blasted against NCBI's nt library [132] to determine if they belong to exogenous DNA. Contigs with more than 70% of their length not covered by any Illumina reads matched bacterial DNA, so 70% was established as a cutoff point to remove exogenous contigs from the assembly.

SSPACE v3.0 [15] and the 523 bp insert size Illumina library were used for a preliminary scaffolding step. To improve accuracy of the assembly, pilon [130] and the 402 bp insert size library were then used to polish the assembly. This preliminary assembly was sent to Dovetail for further scaffolding with the Chicago and Hi-C libraries, which were used as input data for HiRise, a software pipeline designed for utilizing proximity ligation data to scaffold genome assemblies [102].

Using BLAST we identified scaffolds matching the *Oncidium* organelle genomes. These scaffolds, together with the previously identified chloroplast and mitochondrial PacBio reads, were used as input for Canu [77] to perform *de novo* assemblies. SSPACE v3.0 [15] and pilon [130] were used to improve the contiguity and accuracy of the assemblies. The final chloroplast genome was annotated using GeSeq [124] from Chlorobox and visualized with OGDRAW [54].

Transcriptome sequencing and assembly

Floral tissue from the labellum of 2 chemotype A and 2 chemotype M plants was collected, immediately frozen in liquid nitrogen, and stored at $-80°C$ until RNA extraction with TRIzol reagent and TurboDNAase DNA removal. The samples were then used in cDNA library construction and Illumina TruSeq sequencing. The cDNA libraries were constructed using the NEBNext Ultra RNA Library Prep Kit and sequenced on an Illumina TruSeq platform generating 100-bp paired-end reads. Before assembly, high-quality reads were obtained by removing adapter sequences and filtering out low-quality and putative rRNA reads. Using the *de novo* assembly of the *Gongora* genome, we performed a genome-guided *de novo* transcriptome assembly with hisat2 [74] and Trinity [58].

Repeat annotation

Tandem repeats and transposable elements were identified and annotated using the Extensive de-novo TE Annotator (EDTA) v1.9.8 pipeline [98], RepeatModeler v.2.0.1 [46] and RepeatMasker v.4.1.2 [25]. Briefly, the EDTA pipeline and RepeatModeler were used for both *ab initio* and homology-based identification of TEs and tandem repeats, producing a custom *Gongora* repeat library. We then used RepeatMasker with the custom library and with the monocots library from Repbase (20181026 release) [70] to combine the results and generate a complete repeat annotation.

Gene prediction

The MAKER v.3.01.04 pipeline was used to annotate the *Gongora* genome and to generate a consensus gene set based on a comprehensive strategy integrating homology-based and transcriptome-based predictions. We combined our custom repeat library with protein data from *O. sativa* and *A. officinalis* from EnsemblPlants [62], EST data from *P. aphrodite* from Orchidstra 2.0 [23] for homology-based prediction, and our *de novo* transcriptome assembly for chemotype A for *ab initio* gene prediction. These results were integrated into a final set of 20,496 protein-coding genes. BUSCO v3.0.2 [117] was used to evaluate the completeness and quality of the final set of gene models.

TPS genes

The highly conserved domains PF01397 (N terminal domain) and PF03936 (C terminal domain) were downloaded from the Pfam database [91] and were used, together with previously reported orchid TPS protein sequences [143], to generate hidden Markov model profiles and carry out HMM searches with augusuts v.3.3 [120] against the *Gongora* genome assembly. The results were manually examined and annotated, resulting in a final set of 19 *Gongora* TPS genes. These sequences were analyzed with BLASTP v.2.7.1 against MAKER's protein predictions to further improve the annotations.

The predicted *Gongora* TPS protein sequences, together with TPS sequences from other orchids (*Apostasia shenzhenica* [143], *Dendrobium catenatum* [64], *D. officinale* [143], *Pha-*

laenopsis aphrodite [64], *P. bellina* [64], *P. equestris* [143] and *Vanilla planifolia* [64]), *Arabidopsis thaliana* [143], *Oryza sativa* [143] and *Sorghum bicolor* [143] were aligned with MUSCLE [42] using default settings. Based on this alignment, we reconstructed an unrooted neighbor-joining phylogenetic tree using MEGA v.11.0.10 [123] with default parameters and 100 bootstrap replicates. Finally, all repetitive elements in the gene sequences and the 20kb up and downstream of the coding region were identified and annotated.

Whole genome re-sequencing

Sample preparation and sequencing

All plants used in the present study were collected from the surroundings of the La Gamba Tropenstation, in the Puntarenas province, Costa Rica. In total, we included 7 samples from Chemotype A and 6 from Chemotype M (Table B.10). Phenotypic data from these samples includes: floral scent analyses (Tables B.11 and B.12), pictures of the flowers (Figure B.5) and pollinator visitation field observations (Tables B.13 and B.14). DNA extractions were done with the Plant DNeasy Mini Kit (Qiagen, Germany) and then sent to Novogene for library construction and whole genome resequencing on NovaSeq 6000 sequencer to generate 150bp PE reads to a target coverage of 8x.

Population genetic analyses

After assessing the quality of the sequencing data with FASTQC v.0.11.7, the reads were mapped to the *Gongora* reference genome using BWA-MEM, as implemented in BWA v.0.7.17 [81]. Samtools v.1.8 [84] was then used to mark duplicates and index the bam files. We used bcftools v.1.6 [34] to call variants and filtered them with vcftools v.0.1.17 [33] to remove insertions and deletions, sites with minor allele frequency under 0.08, sites with over 5% missing data, sites with mapping quality below 30, and sites with a sequencing depth below 5x or over 20x.

To investigate the population structure, we used plink v1.90 [101] to perform a principal component analysis (PCA) based on 183,776 unlinked sites. To estimate divergence and nucleotide diversity, we used the popgenWindows.py script written by Simon Martin to

estimate Fst, dxy and pi over 20kb windows with a 10kb step size and with a minimum of 2kb sites covered. All results were plotted in R.

Results

Genome assembly

In this study, we report the first draft of the *Gongora* genome assembly. To overcome the high repeat content and heterozygosity, our assembly strategy consisted of a combination of short- and long-read sequencing together with chromosome conformation capture (Hi-C) technologies. Based on a k-mer analysis, the final genome size was estimated to be 2.228 Gb (2.6Gb with flow cytometry) with a heterozygosity of 5.9%. A total of 71 Gb of SMRT sequences were corrected with small reads and used for the initial contig assembly. SSPACE v3.0 [15] was used for scaffolding and, after polishing with pilon v.1.23 [101], the total length of the assembly was 1.831 Gb, with a corresponding contig N50 value of 0.382 Mb (Table B.3). To further improve the assembly, 35.1 Gb of Chicago and 32 Gb of Hi-C library reads were used to anchor, order and orient the contigs. The final assembly contains 9,024 scaffolds, with a total length of 1.832 Gb and an N50 of 1.756 Mb (Table 2.1). About 50% of the total assembled genome is contained in the 262 longest scaffolds.

Completeness of the genome assembly was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) v.3.0.1 [117] with default parameters and the embryophyta dataset. Of the 1614 conserved core embryophyta genes used to assess genome completeness, 1422 (88%) of core genes are represented in our genome assembly (Table 2.1).

The mitochondrial genome was assembled into 13 scaffolds with a total length of 462,164 bp. The chloroplast genome is contained within a single scaffold of 187,299 bp in length and contains a pair of inverted repeats named IRa and IRb of 26.6 kbp that divide the chloroplast genome into a large single copy (LSC) (84.8 kbp) and a small single copy (SSC) region (49.1 kbp). We identified and annotated a total of 156 genes, including 19 genes duplicated in the IR region, 30 distinct tRNAs and 4 distinct rRNA genes.

Gene prediction

Using both *de novo* and library-based repetitive sequence annotation, a total of 1.53 Gb repetitive elements occupying more than 83% of the *Gongora* genome were annotated. The

Table 2.1. Comparison between genome assemblies

	P. equestris	D. catenatum	P. aphrodite	C. sinense[a]	C. goeringii[a]	*Gongora*
Year of publication	2015	2016	2018	2021	2021	**2022**
Estimated genome size (Gb)	1.6	1.11	1.2	3.52	4.0	**2.228**
Assembled genome size (Gb)	1.086	1.01	1.025	3.45	3.99	**1.832**
Scaffold N50 (Mbp)	0.359	0.391	0.946	NA	NA	**1.756**
Contig N50 (kbp)	20.55	33.09	18.81	1110	377.6	**382.58**
Longest scaffold (Mbp)	81.76	2.59	10.39	NA	NA	**17.67**
Repeat content (%)	62	78.1	60.3	77.78	88.87	**83.36**
BUSCO assessment (%)	91	92.46	95	91	87.8	**88**
Gene number	29431	28910	28902	29638	29556	**20496**
Total scaffolds	236185	72901	13732	20	20	**9024**
Reference	[20]	[146]	[22]	[142]	[27]	

[a]Chromosome-level assembly.

repetitive content of *Gongora* is higher than most of the other sequenced orchids except for *C. goeringii* (88.87%). Retrotransposable elements, known to be the dominant form of repeats in angiosperm genomes, constitute a large part of the genome and include the most abundant subtypes, such as LTR/Copia (3.04%), LTR/Gypsy (2.82%), LINE/L1 (0.83%) and LINE/RTE-BovB (1.1%), among others (Table B.7). Of the repetitive elements, 31.11% could not be classified into any known families, which is consistent with previous reports from other orchid genomes suggesting that there may be new repetitive or transposable elements unique to the family Orchidaceae [140, 142, 146].

Protein-coding gene models were constructed using a pipeline combining *de novo* prediction and homology-based prediction methods. In total, 20,496 putative genes were identified in the *Gongora*. Using BUSCO v3.0 to assess the completeness of the genic regions with the lilliopsida database, we found that 42.3% of the 3,236 plant single-copy orthologues were present in the annotation.

TPS genes

A small TPS gene family size was observed in *Gongora* relative to other orchids (Table B.8). To resolve the phylogenetic relationship of the *Gongora* TPS genes and those of other orchids, we constructed a phylogenetic tree based on their amino acid sequences and included TPS gene sequences derived from *Arabidopsis thaliana*, *Oryza sativa* and *Sorghum bicolor*. The 19 putative TPS genes in *Gongora* were ascribed to four previously recognized TPS subfamilies in angiosperms: TPSa, TPSb, TPSc, and TPSe/f (Figure 2.1).

A detailed analysis of the repetitive elements in coding DNA sequences (CDSs) of the TPS genes showed that the repetitive content of the introns is similar to that of the 10kb and 10-20 kb regions around the CDSs (Figure B.4). This result is consistent with the particularly high average intron length found in other orchid genomes compared to other angiosperms [20, 144–146, 149]. It has been suggested that the presence of regulatory elements in the introns could play a role in alternative spacing events, gene regulation and functional diversification in Orchidaceae [149], but more research is needed to properly characterize the expression patterns and regulation of TPS genes in *Gongora*.

Whole genome re-sequencing

An average of 143 million paired-end reads were obtained per individual, with an average overall alignment rate of 99.4% to the *Gongora* reference genome. The average depth of coverage was relatively homogeneous at 9.2X and did not appear to be affected by chemotype. A total of 34,509,160 variable sites were found between all samples.

Genetic clustering, as indicated by the first two principal components of our PCA analysis, was consistent with previous research on the La Gamba population (Chapter 1). The first principal component explaining over 10% of the genetic variance clearly separates chemotype A and chemotype M samples (Figure B.6).

To identify any outlier regions potentially associated with reproductive isolation between chemotypes A and M, we estimated genome-wide patterns of divergence and nucleotide diversity across 20 kb windows spanning the whole genome (Figure 2.2). Overall genetic differentiation between the two species was low (Fst= 0.019), the mean genome-wide dxy was 0.026, and nucleotide diversity was higher in chemotype M (pi= 0.021) compared to
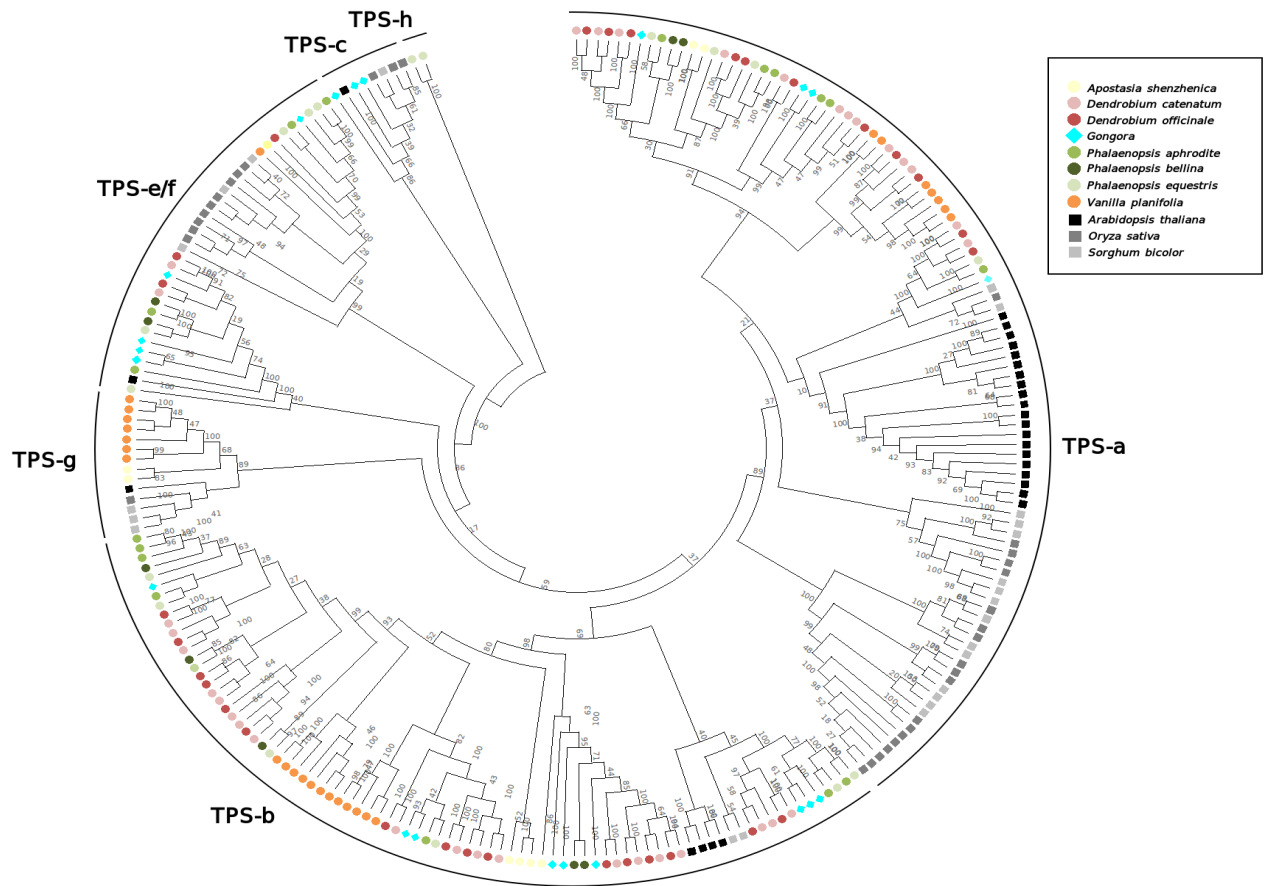
Figure 2.1: The evolutionary history was inferred using the Maximum Parsimony method. The most parsimonious tree with length = 28269 is shown. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm. The tree is drawn to scale, with branch lengths calculated using the average pathway method and are in the units of the number of changes over the whole sequence. This analysis involved 248 TPS amino acid sequences. There were a total of 1864 positions in the final dataset. Evolutionary analyses were conducted in MEGA11.
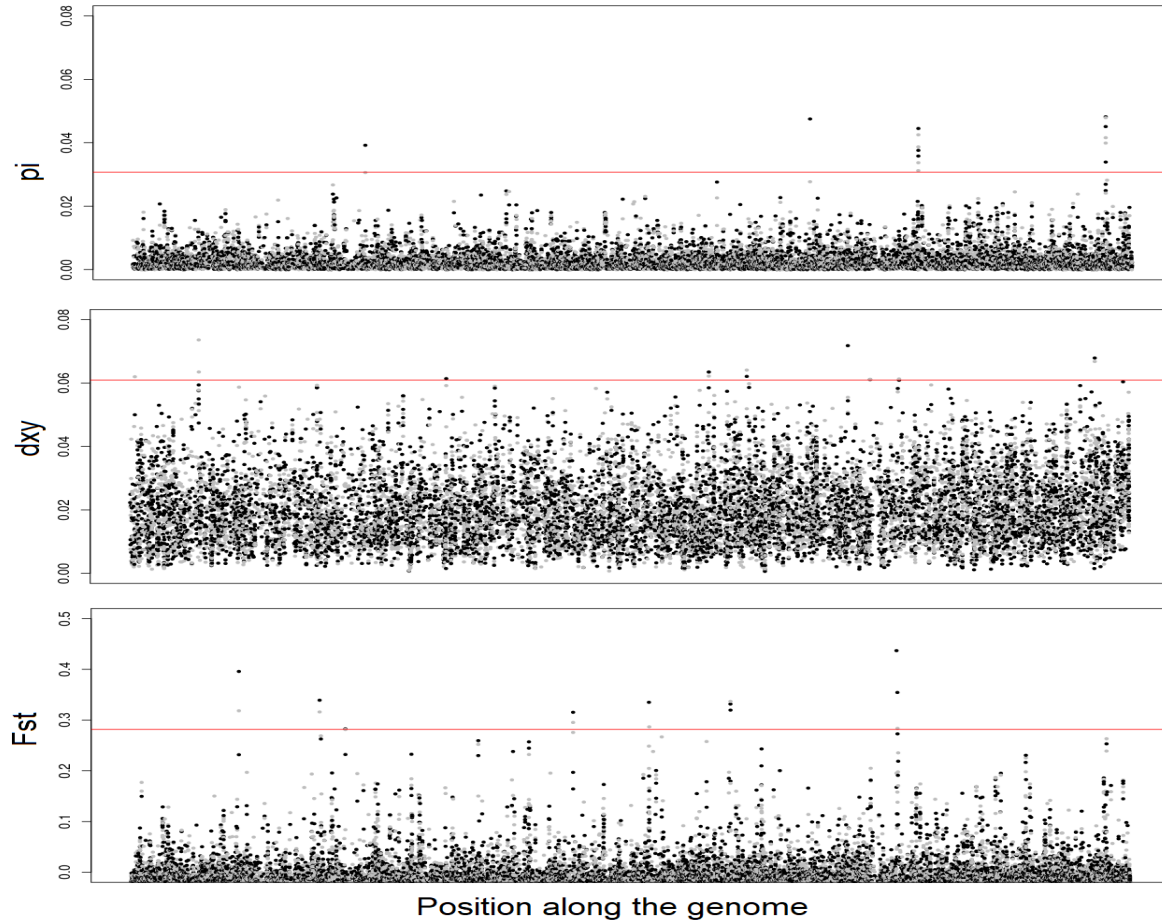
Figure 2.2: Information about the figure.

chemotype A (pi = 0.019). Genomic regions with intermediate levels of genetic differentiation (Fst > 0.1) were scattered across the genome, did not show significantly elevated dxy and their nucleotide diversity did not differ from that of the genetic background.

Discussion

Orchid species belonging to the neotropical subtribes Stanhopeinae and Catasetinae exhibit highly specialized pollination associations with fragrance-collecting male euglossine bees. Because in this system floral scent regulates pollinator attraction and specificity, differences in the floral scent profile of closely related lineages can mediate reproductive isolation [39, 136, 137] (Chapter 1). The study of speciation in these orchids thus involves understanding the molecular basis of floral scent emission as well as the evolutionary forces promoting its differentiation. So far, research in this area has been limited by a lack of omics data.

In this study, we used PacBio, Illumina paired-end, and Hi-C sequencing technologies to construct a genome assembly of *Gongora*, an orchid from the subtribe Stanhopeinae, with an assembled genome size of 1.83 Gb. With this new genome assembly it will be possible to further investigate the genetic architecture of floral scent and its role in reproductive isolation.

Most volatile compounds emitted by *Gongora* flowers belong to the terpene compound class [61]. The biosynthesis of terpenes is mainly mediated by the terpene synthase (TPS) gene family which is present in all land plants. TPS enzymes catalyze complex stereospecific cycloisomerization reactions converting a few central precursor molecules (prenyl diphosphates) into a diverse array of volatile compounds. TPSs evolve rapidly through gene duplication and sequence divergence resulting in an astounding diversity of often lineage-specific terpene compounds [24]. The total number of TPS genes in a genome differs between species and in orchids they have been found to range from 14 in *Apostasia shenzhenica* [143] to 48 in *Dendrobium chrysotoxum* [149]. TPS genes are classified into seven subfamilies: TPS-a, TPS-b, TPS-c, TPS-d, TPS-e/f, TPS-g, and TPS-h [24]. In the *Gongora* genome we identified and annotated 21 different TPS genes (Figure 2.1) with 5 belonging to the TPS-e/f subfamily and 9 to the TPS-b subfamily. Genes from these two subfamilies have been previously shown to be involved in monoterpene biosynthesis in the floral tissue of *Phalaenopsis bellina* [63].

Species-specific diversification of TPS functions is one possible mechanism mediating the differentiation in floral scent profiles among *Gongora* populations. Alteration of TPS product specificity through point mutations has been demonstrated in other systems. For example, point mutations in the limonene synthase of mint (*Mentha spicata*) resulted in the novel production of pinene and linalool [118]. Additionally, novel transcriptome diversity could be generated through transposable elements (TEs) by providing novel promoters, splice sites, or polyadenylation signals [28]. In *Gongora* and other orchid genomes, the TE content in introns has contributed to their increased average length compared to other angiosperm genomes [20, 143]. Here we found that the TE content in the TPS introns of *Gongora* was comparable to that of the surrounding genomic regions (Figure B.4). Further research is needed to functionally characterize *Gongora*'s TPS genes and to investigate their expression

patterns between different chemotypes.

The differential expression and regulation of biosynthetic pathways underlying the production of volatile organic compounds is expected to play a major role in scent differentiation. *Gongora* chemotypes A and M from La Gamba are closely related to each other, occur sympatrically and have overlapping flowering phenologies. The volatile compounds emitted by these two lineages are the product of different biosynthetic pathways: chemotype A produces mainly aromatic compounds and chemotype M emits monoterpenoids (Tables B.11, B.12). Through pollinator network reconstruction we have previously shown that each chemotype attracts a unique set of pollinator species, but reproductive isolation is not complete (Chapter 1). Despite the occurrence of gene flow, no plants with intermediate floral phenotypes have been observed so far. Therefore, we hypothesized that, through natural selection, the genomic regions associated with differential floral scent emission would withstand the homogenizing process of gene flow and remain highly differentiated.

To further understand the mechanisms involved in floral scent evolution, we generated whole-genome sequences from 13 individual plants (6 chemotype M and 7 chemotype A). Analysis of 38,694 unlinked single-nucleotide polymorphisms (SNPs) confirmed our previous finding that the population is structured by floral scent profile (Figre B.6). Genome-wide differentiation was low (Fst = 0.019) with three scaffolds containing regions with Fst values falling within the top 1% of the empirical Fst distribution. These regions contained no protein-coding genes, they did not exhibit elevated divergence (dxy) compared to the rest of the genome (genome-wide dxy = 0.02), and their levels of nucleotide diversity (pi) were not different from the genomic background, suggesting that they did not arise as a result of divergent selection advancing reproductive isolation [31].

Genomic landscapes of differentiation can be influenced by different processes such as genetic drift, sorting of ancient divergent haplotypes, recent selective sweeps in regions with low recombination, or background selection [31, 32, 65, 94, 107]. Moreover, if the genetic architecture of floral scent is complex and involves many genes of small effect, this could prove difficult to detect in genome scans [111].

Looking into the differential expression of genes in the labellum of these orchids might shed light into the regulatory networks involved in differential floral scent biosynthesis. An-

other possible way forward is to sample additional *Gongora* chemotypes. In our previous research we have identified another pair of chemotypes with similar floral scent composition to each other, but only differing in the presence/absence of a few additional modifier compounds. Contrasting different orchid lineages can be a powerful tool to unravel the selection pressures that have contributed to the floral scent differentiation in the genus [87, 129].

In conclusion, we generated a high-quality reference genome for *Gongora* which will serve as a crucial resource for understanding the evolution and maintenance of reproductive barriers in euglossine-bee pollinated orchids. Future studies may focus on elucidating the molecular mechanisms that control pollinator specialization in this group by improving our annotation and studying the expression and regulation of biosynthetic pathways involved in floral scent production.

Appendix A

Appendix Chapter 01

A.1   Supplementary Methods

Study Sites and Orchid Populations

The orchid genus *Gongora* is broadly distributed across tropical America in humid lowland forests where some species can reach relatively high abundances [60, 66]. We conducted our study in the surrounding forests of the La Gamba (8°42' 2.26" N -83°12' 3.68" W) and the La Selva (10°25' 19" N -84°00' 54" W) Biological Stations in Costa Rica. *Gongora* orchids are epiphytic and often grow in mature and secondary forests. Plants are easily identified in the field by the presence of angularly ribbed, ovoid pseudobulbs with two plicate leaves per pseudobulb bearing 5 longitudinal, prominent nerves, and lateral overhanging inflorescences arising from the base of the pseudobulbs [66]. We collected adult *Gongora* plants between 2013 and 2019 for La Gamba and between 2016 and 2019 for La Selva along the established trail systems adjacent to each station. Plants were potted and maintained at the stations until blooming. We photographed flowers of each individual plant to record morphological and color variation. Pollinaria were removed from each inflorescence and were preserved in vials filled with silica gel until subsequent DNA extraction.

Chemical Analysis of Floral Scents

We analyzed the chemical composition of *Gongora* floral scents using Gas Chromatography-Mass Spectrometry (GC-MS). Our approach aimed to elucidate the diversity of volatile compounds emitted by flowers in order to identify potential cryptic species. To extract

floral scents, we implemented a static headspace method in both the greenhouse and the field. Briefly, a single inflorescence was enclosed inside a nylon oven bag (Reynolds Kitchens, Richmond, VA, USA) closed at the top with metal wire for 30 minutes, for the scent to accumulate in the bag. Subsequently, we connected scent traps to a diaphragm electric vacuum pump (Parker, Cleveland, OH, USA) via Tygon tubing (ID 3.3 mm) and continuously extracted air from the bag through a small slit. We fabricated single-use scent traps with clear glass tubing (2.4mm ID, 3.5cm length) plugged at both ends with glass wool and filled with 20 mg of bulk carbide (charcoal) and 20mg of Tenax GC (Supelco, Bellefonte, PA, USA; mesh size 60/80). We conditioned scent traps by passing 5mL of hexane, which we dried by placing on a hot plate (at 50-60°C) for 30 minutes. We extracted scent volatiles by passing air from the headspace through the scent trap at a rate of 2.5 L per minute for 2 h using the vacuum pump. We eluted scent compounds by injecting 200 µL of clean hexane into the scent trap, which we stored in glass inserts within 2 mL GC vials. Samples were kept at -20° C until GC-MS analysis.

We analyzed scent extracts using an Agilent 7890B GC fitted with a 30 m × 0.25 mm × 0.25 mm HP-5 Ultra inert column, coupled to an Agilent 5977A mass spectrometer (Agilent Technologies). All scents were analyzed in the same instrument, which is housed at the University of California Davis. Whenever possible, we obtained multiple sample replicates from different inflorescences produced by the same plant. We obtained negative controls by simultaneously sampling empty bags filled with ambient air. We injected a 1 µL aliquot using an auto-sampler set to a 3:1 split ratio. The split ratio was adjusted for some samples to increase signal detection. Oven temperature was programmed at 60°C for 3 min and then increased by 10°C per minute until it reached 300°C; then the oven temperature was kept at 315°C for 1 min. The injector and transfer line temperatures were kept constant at 250°C. We used helium as the carrier gas with a constant flow rate set to 1.2 mL per minute. Electron Impact (EI) mass spectra were obtained by scanning between 30 and 550 m/z. GC-MS data were processed using MassHunter GC/MS Acquisition software vB.07.00 (Agilent) and MSD ChemStation Enhanced Data Analysis Software vF.01.00 (Agilent). We tentatively identified scent compounds by searching against the NIST05 mass spectral database using the NIST MS Search software v2.0. We confirmed compound identities by comparing against

authentic chemical standards run under identical conditions [2]. We calculated the total ion abundance of each peak using the MSD ChemStation software using the RTE integrator. Only those peaks with an area greater than 3% relative to the largest peak area were included in downstream analyses.

We used multivariate statistical methods to investigate the variation of chemical profiles between individual plants and chemotypes. We normalized raw peak areas by estimating their relative area (the area of each peak divided by the total chromatogram peak area). We calculated pairwise distance among individual orchid samples using the Bray–Curtis dissimilarity metric in the package ecodist v1.2.2 [50]. The Bray–Curtis dissimilarity metric is not affected by 'double zeros' and only considers compounds that are jointly shared between samples [13, 80, 151]. We used the dissimilarity matrices to perform a non-metric multi-dimensional scaling (nMDS) analysis. This method visually represents similarity between individuals in pre-specified reduced space dimensions, using a non-eigenvector method that is flexible with respect to the choice of distance metrics (e.g. Bray–Curtis). Individual points (samples) that cluster together share a more similar scent chemistry than those located further apart. We created two- and three-dimensional plots using the 'nmds' algorithm in ecodist v1.2.2.

Pollinator Observation and Visitation Rates

We reconstructed the pollination network of each *Gongora* chemotype present in our study site. Between 2013 and 2019, we documented visitation rates and the diversity of euglossine bee species visiting *Gongora* inflorescences at La Gamba and then the same was done at La Selva between 2016 and 2019. Immediately after extracting the headspace (as described above) we relocated each plant to a nearby forest site to observe pollinator behavior during the morning hours (from 8:00 AM until noon). A *Gongora* inflorescence consists of a pendant spike that contains 4-30 flowers, and all the flowers from the same spike begin anthesis simultaneously at dawn and remain open for three days before wilting. Thus, we conducted both headspace sampling and pollinator observation on the first and second day of anthesis. Whenever possible, multiple observations were conducted on the same individual plant if multiple inflorescences were produced. Male bees were allowed to land at least once on the

inflorescence to perform scent-collecting behavior before being captured. All bee visitors were captured for proper species identification. We deposited bee voucher specimens in our entomological reference collection at the University of California Davis. We visualized and analyzed bee-orchid networks using the R package bipartite v2.05 [37].

Population Genetic Analyses

We conducted a population genetic analysis of *Gongora* from the general area surrounding the La Gamba and La Selva Biological Stations. We obtained DNA sequence data from both field-collected adult plants and pollinaria samples recovered directly from male bees. Our analysis aimed to (i) identify co-occurring cryptic populations, (ii) reconstruct high-resolution bee-orchid pollination networks, and (iii) infer the chemotypes of pollinaria by matching pollinaria genotypes to those of field collected plants with known chemotypes. Pollinaria samples recovered from male bees were obtained by luring male bees with six broad-spectrum chemical baits (1,8-cineole, methyl salicylate, eugenol, terpinen-4-ol, vanillin, 1,4-dimethoxybenzene). We presented chemical baits on 10×10 cm filter paper squares that were attached to trees along an established trails near the stations. Baiting was conducted between 8:00 AM and noon, which corresponds to the peak time for activity of male euglossine bees. Upon capture, we removed the attached pollinaria and immediately transferred to 2mL vials filled with silica gel to ensure DNA preservation. Voucher bees were pinned and deposited in the entomological collection housed at the University of California Davis. Some bees carried multiple pollinaria, and in such cases we conducted DNA extractions separately for each pollinarium. We extracted DNA using DNeasy Plant Mini kits (Qiagen).

We implemented Genotype by Sequencing (GBS) following previously established protocols [43, 86]. We pooled 95 samples at a time and genotyped them using a single Illumina Hi Seq lane. Libraries were cleaned and sequenced at the QB3 Vincent J. Coates Genomics Sequencing Laboratory (UC Berkeley). The resulting sequencing reads were screened for Single Nucleotide Polymorphism (SNPs) using the UNEAK pipeline implemented in TASSEL [16] with error tolerance rates set to 0.03. This conservative approach reduces the number of SNP calls due to potential sequencing errors and requires that each site is identified by at least 5 sequence reads, that occur in more than 50% of individuals and are inferred strictly

as diploid individuals (SNPs with higher ploidy levels are considered artifacts).

For our analysis, we required SNPs to be called in at least 25% of individuals. Additionally, for an individual to be included in any downstream analysis it must have genotype information at 25% of SNP sites. We also excluded any sites that had data in one of our control lanes for each GBS plate. We used ADMIXTURE [5] to estimate ancestry in the combined set of individuals with chemotype information and pollinaria samples. To estimate between-chemotype visitation rates, we used a subset of bee pollinators that were caught while visiting orchid flowers and carried *Gongora* pollinaria. By comparing the genotype of these pollinaria against the genotypes of the plants being visited by the bee, we directly estimated cross-visitation rates.

A.2 Supplementary Tables

Table A.1. La Gamba chemotype profiles

Chemotype A	Chemotype M	Chemotype S	Chemotype X
trans-methyl-methoxy cinnamate 35.8%	terpinen-4-ol 70.5%	Unknown 51 48.1%	cineole 65.9%
estragole 30.9%	beta pinene 9.8%	alpha farnesene 26.6%	terpinen-4-ol 13.7%
cis-methyl-methoxy cinnamate 15.8%	alpha pinene 4.3%	linalool 9.23%	alpha pinene 8.2%
chavicol 8.04%	sabinene 3.7%	cineole 6.3%	beta-pinene 4.9%
beta elemene 3.5%	limonene 3.8%	beta elemene 2.3%	veratrol 3.6%
caryophyllene 2.1%	alpha thujene 2.8%	beta ocimene 1.9%	linalool 3.5%
anethole 1.9%	cineole 1.9%	beta pinene 1.1%	
cinnamic acid 1.7%	terpinolene 1.7%	Unknown 36 0.6%	

Note. — Average floral scent composition per chemotype from La Gamba based on the following number of floral scent extractions: 26 from chemotype A, 23 from chemotype M, 11 from chemotype S, 5 from chemotype X.

Table A.2. La Selva chemotype profiles

Chemotype P	Chemotype R	Chemotype T
Unknown 12 65.7%	estragole 54.8%	estragole 44.3%
Unknown 14 34.3%	eugenol 38.1%	trans-methyl-methoxy cinnamate 29.4%
	methyleugenol 3.6%	cis-methyl-methoxy cinnamate 16.3%
	estragole trans 2.2%	eugenol 9.3%
	p-methyl-anisole 1.22%	

Note. — Average floral scent composition per chemotype from La Selva based on the following number of floral scent extractions: 9 from chemotype T, 16 from chemotype R, 8 from chemotype

Table A.3. La Gamba Fst

Pop1	Pop2	Fst	Lower bound CI limit	Upper bound CI limit	p-value
A	M	0.0534	0.0514	0.0554	0
A	S	0.2599	0.2555	0.2624	0
A	X	0.2590	0.2570	0.2622	0
M	S	0.1929	0.1936	0.2001	0
M	X	0.1966	0.1897	0.1957	0
S	X	0.0743	0.0714	0.0775	0

Note. — Pairwise Fst values were calculated in R using the function gl.fst.pop from package dartR.

Table A.4. La Gamba Heterozygosity

Pop	Ho	He	Fis	Number of samples
A	0.0836	0.2078	0.4354	29
M	0.0946	0.2362	0.4307	28
S	0.0771	0.2351	0.5008	28
X	0.0601	0.2358	0.5782	4

Note. — Values were calculated using the function basic.stats from R package hierfstat.

Table A.5. La Gamba F3 Statistics

Pop1	Pop2	Pop3	f3	stderr	Zscore	nsnps
A	S	M	-0.0064	0.0007	-8.63	61564
A	X	M	0.0004	0.0006	0.662	60384
M	X	S	0.0638	0.0023	27.32	59769
A	S	X	0.0671	0.0018	37.17	59692
A	X	S	0.0712	0.0025	27.63	59692
M	S	X	0.0750	0.0017	43.52	59769
M	X	A	0.0930	0.0021	43.56	60384
M	S	A	0.1013	0.0019	52.09	61564
S	X	M	0.2184	0.0060	36.28	59769
A	M	S	0.3005	0.0100	29.99	61564
A	M	X	0.3237	0.0080	40.42	60384
S	X	A	0.3459	0.0050	69.13	59692

Note. — F3 values were calculated using the f3 function in the R package admixr.

Table A.6. La Selva Fst

Pop1	Pop2	Fst	Lower bound CI limit	Upper bound CI limit	p-value
P	T	0.0566	0.0536	0.0587	0
P	R	0.0748	0.0718	0.0779	0
T	R	0.0075	0.0051	0.0102	0

Note. — Pairwise Fst values were calculated in R using the function gl.fst.pop from package dartR.

Table A.7.   La Selva Heterozygosity

Pop	Ho	He	Fis	Number of samples
P	0.0854	0.1908	0.3582	20
T	0.1284	0.2251	0.2180	14
R	0.1007	0.2259	0.3167	11

Note. — Values were calculated using the function basic.stats from R package hierfstat.

Table A.8.   La Selva F3 Statistics

Pop1	Pop2	Pop3	f3	stderr	Zscore	nsnps
R	P	T	0.0172	0.0012	13.50	40461
T	P	R	0.0485	0.0010	44.74	40461
T	R	P	0.0897	0.0018	47.98	40461

Note. — F3 values were calculated using the f3 function in the R package admixr.

A.3   Supplementary Figures
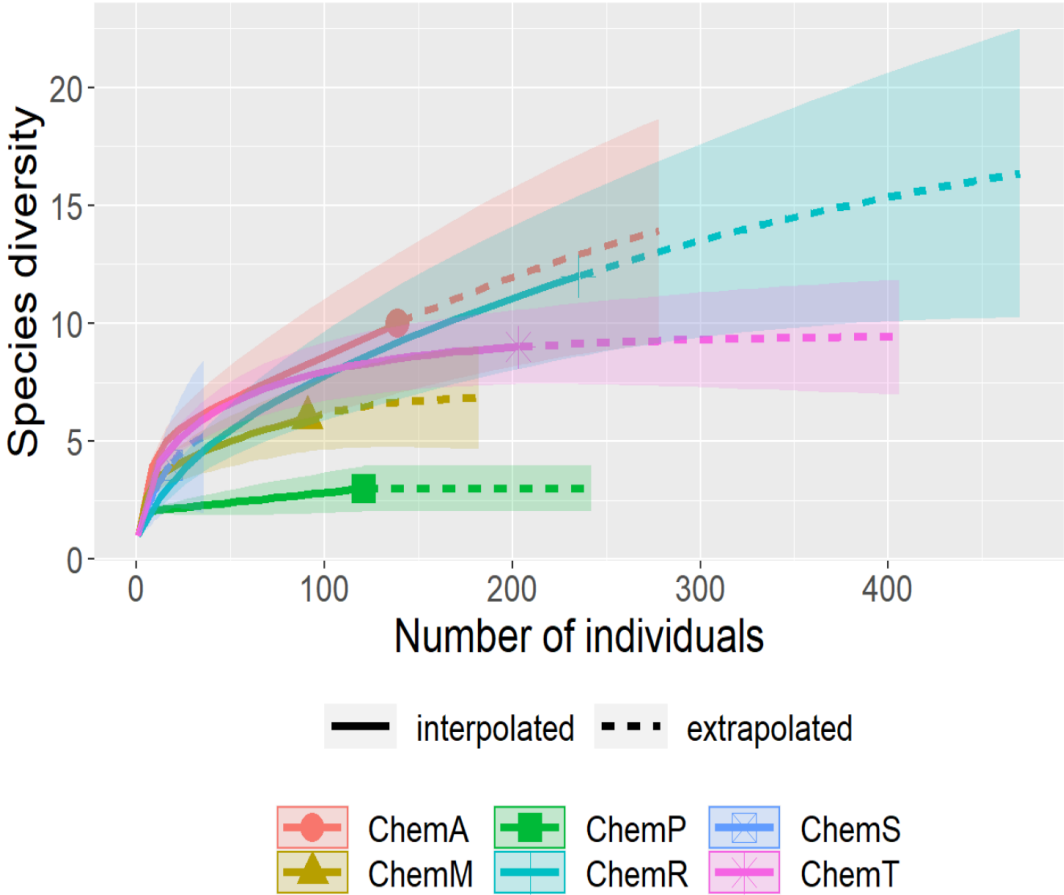
Figure A.1: Rarefaction curves showing the diversity of bee species observed visiting all seven chemotypes.
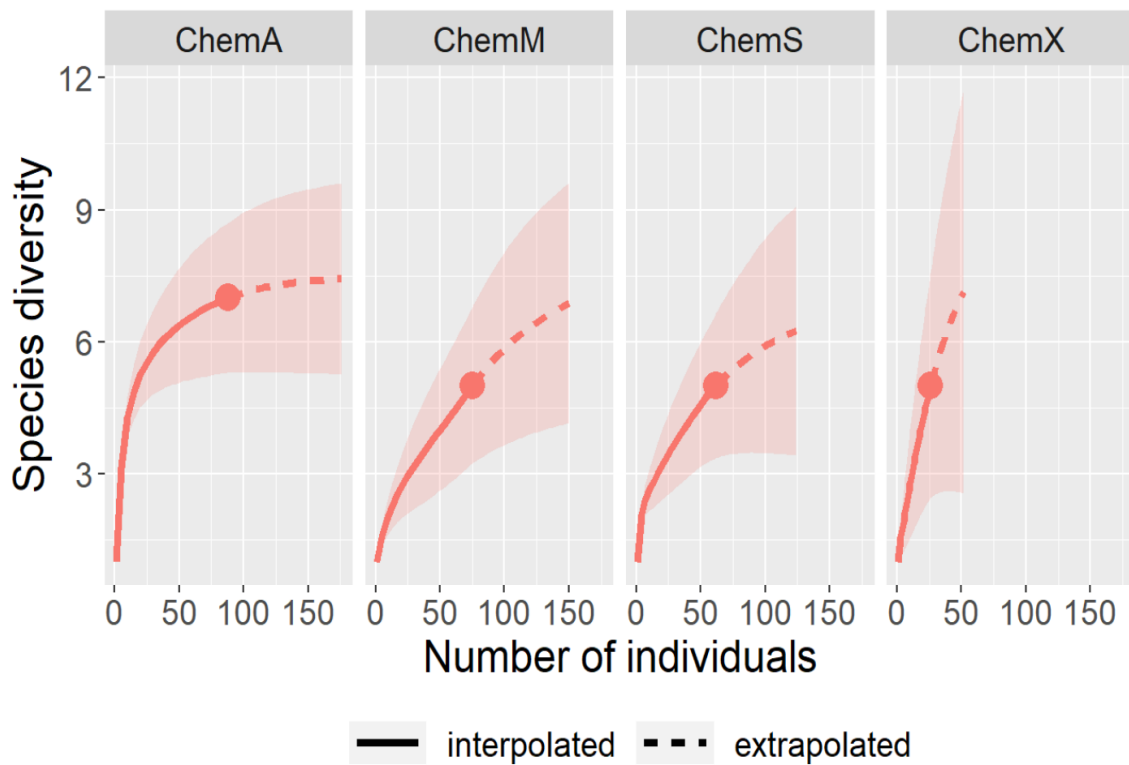
Figure A.2: Rarefaction curves showing the diversity of pollinator species known to carry pollinaria from the four chemotypes at La Gamba.
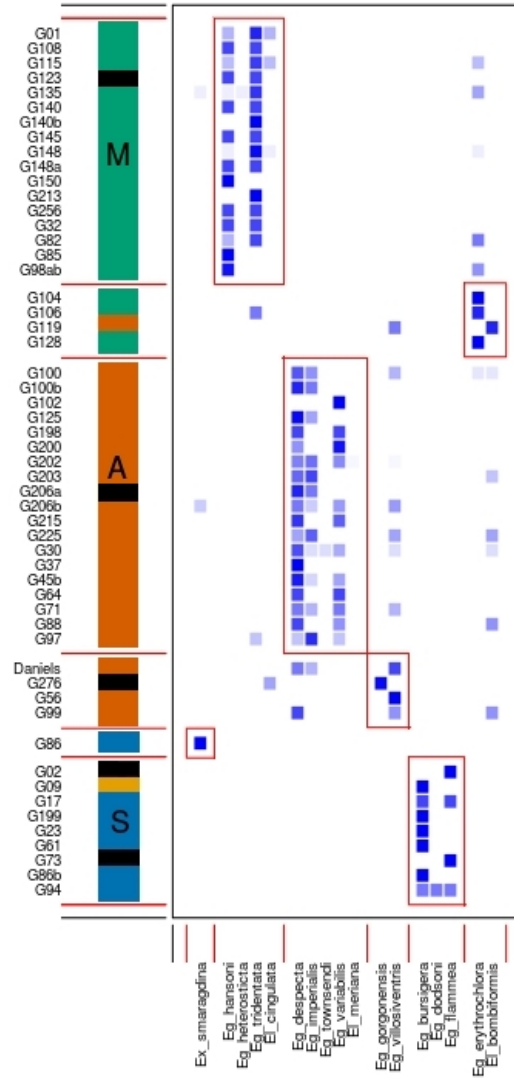
Figure A.3: Modularity analysis of bee visitors from La Gamba was performed using the function computeModules from package bipartite in R. The y-axis contains individual plant data colored by chemotype (black bars are from plants not phenotyped). The x-axis contains bee species and the blue squares are colored according to the relative amount of bee visitors from a given species each plant obtained, the darker the color the more visitors.
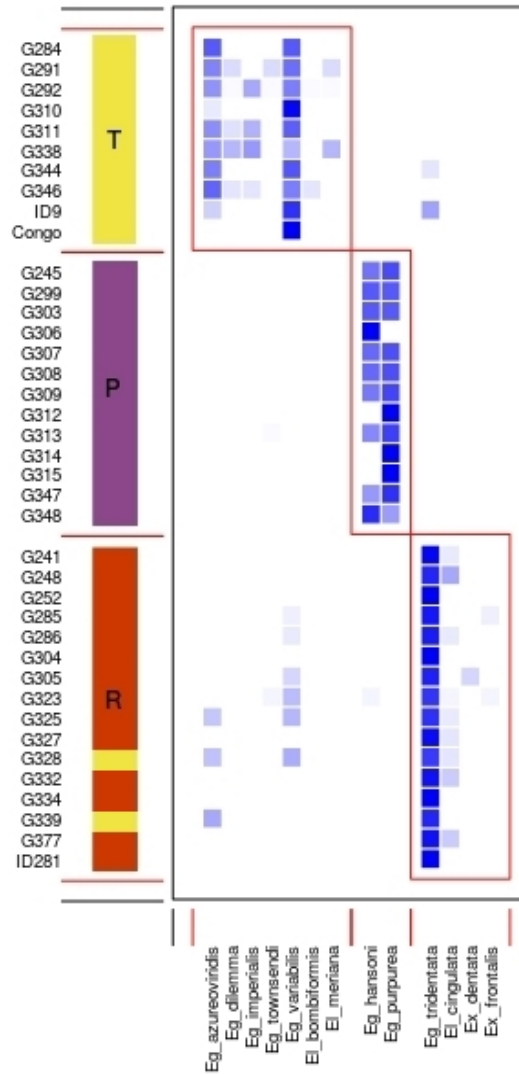
Figure A.4: Modularity analysis of bee visitors from La Selva was performed using the function computeModules from package bipartite in R. The y-axis contains individual plant data colored by chemotype. The x-axis contains bee species and the blue squares are colored according to the relative amount of bee visitors from a given species each plant obtained, the darker the color the more visitors.
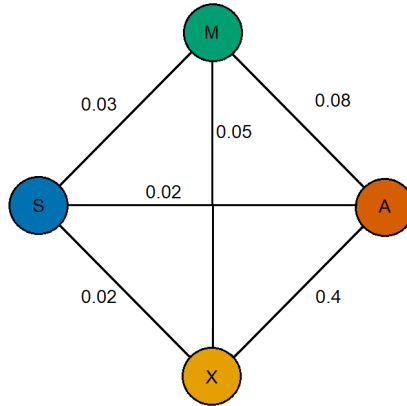
Figure A.5: Pollinator overlap (PS) estimates between chemotypes from La Gamba.
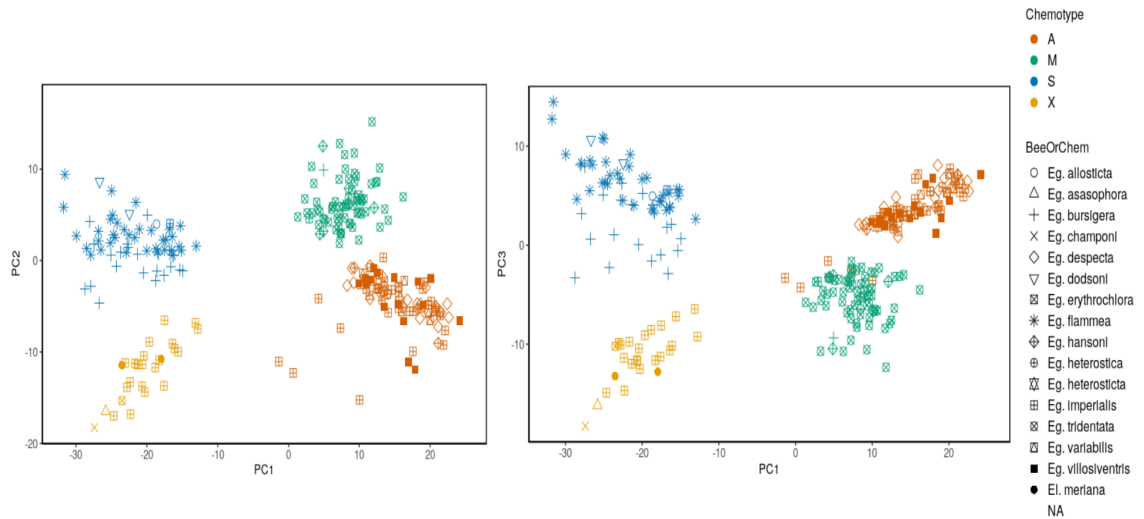
Figure A.6: Principal Component Analysis of 130 La Selva genotypes based on 79,675 SNPs with 45.5% missing data was performed with function glPca from package adegenet in R. The first principal component (PC1) explains 7.58% of variance, PC2 explains 1.18% and PC3 1.09%.
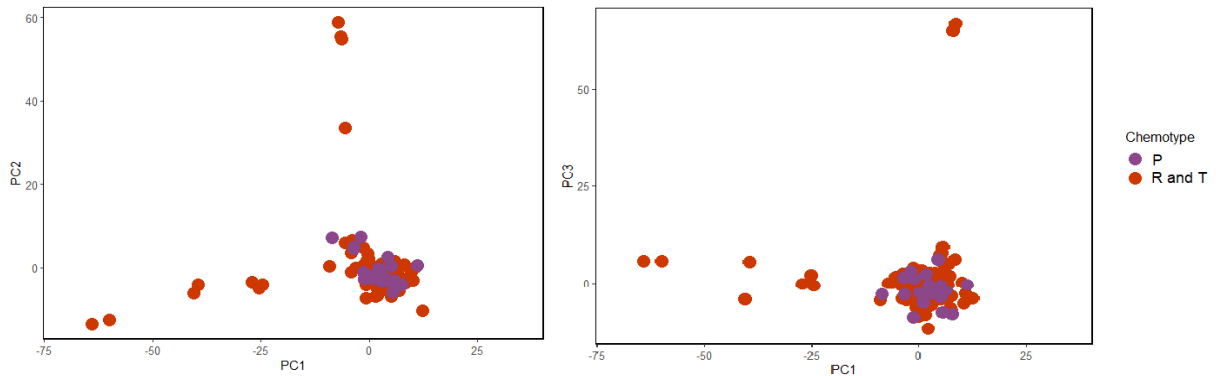
Figure A.7: Principal Component Analysis of 324 La Selva genotypes based on 71,948 SNPs with 53.31% missing data was performed with function glPca from package adegenet in R. The first rincipal component (PC1) explains 2.58% of variance, PC2 explains 2.066% and PC3 1.763%.
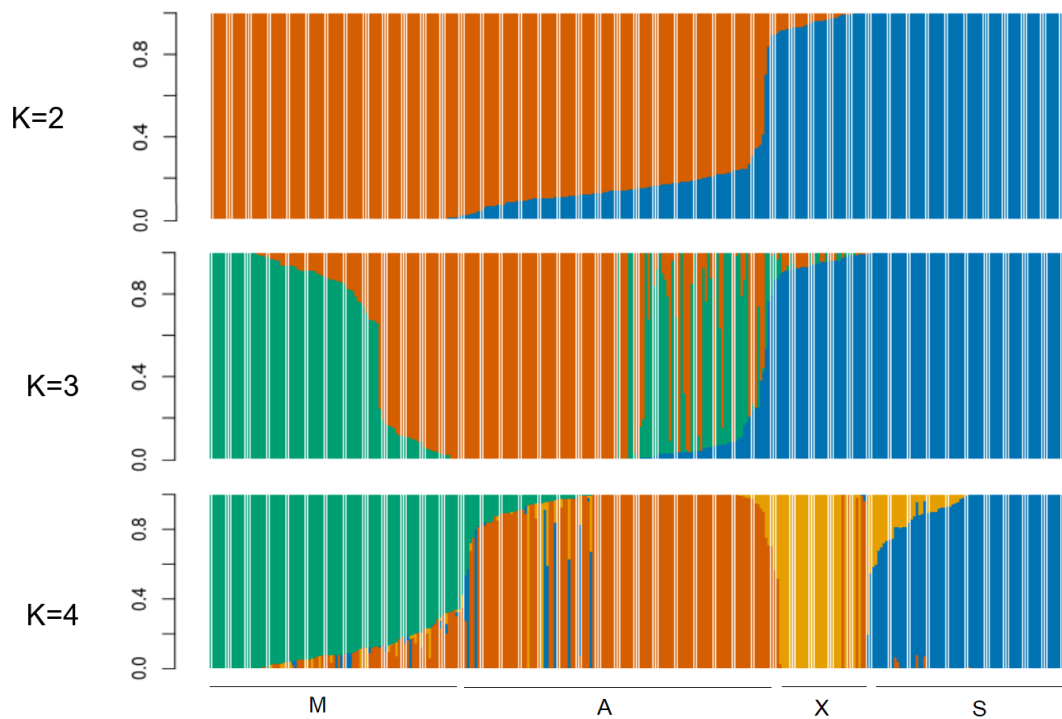
Figure A.8: ADMIXTURE analysis (K=2 to 4) for ancestry estimation of the La Gamba chemotypes: A, M, S and X. Each individual is represented by a thin vertical line, which is partitioned into K colored segments that represent the individual's estimated membership fractions in K clusters.
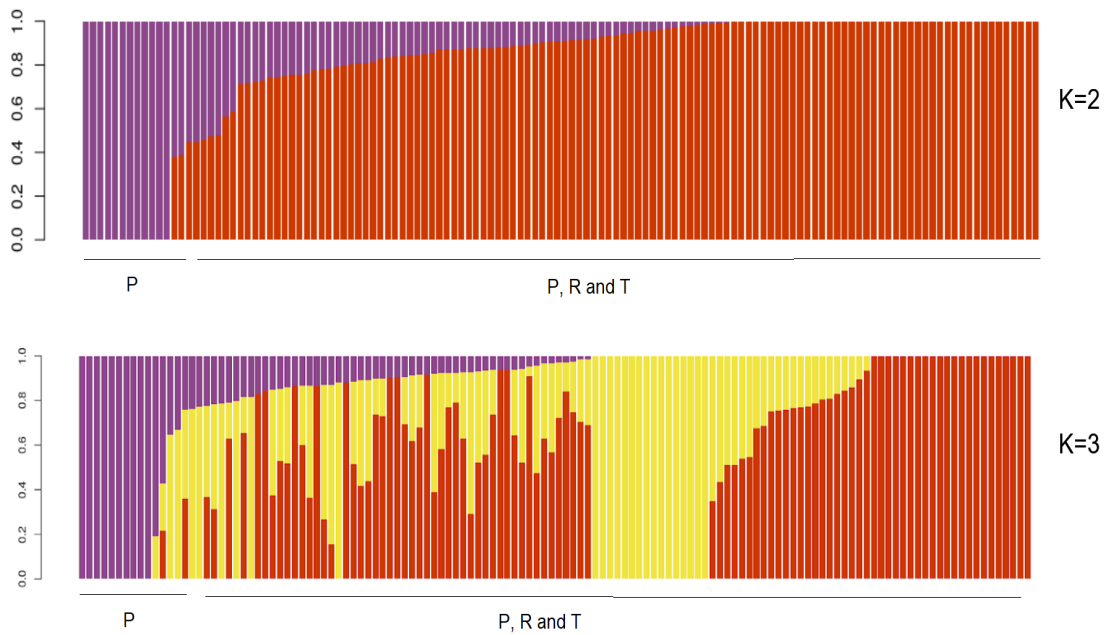
Figure A.9: ADMIXTURE analysis (K=2 and K=3) for ancestry estimation of the La Gamba chemotypes: A, M, S and X. Each individual is represented by a thin vertical line, which is partitioned into K colored segments that represent the individual's estimated membership fractions in K clusters.
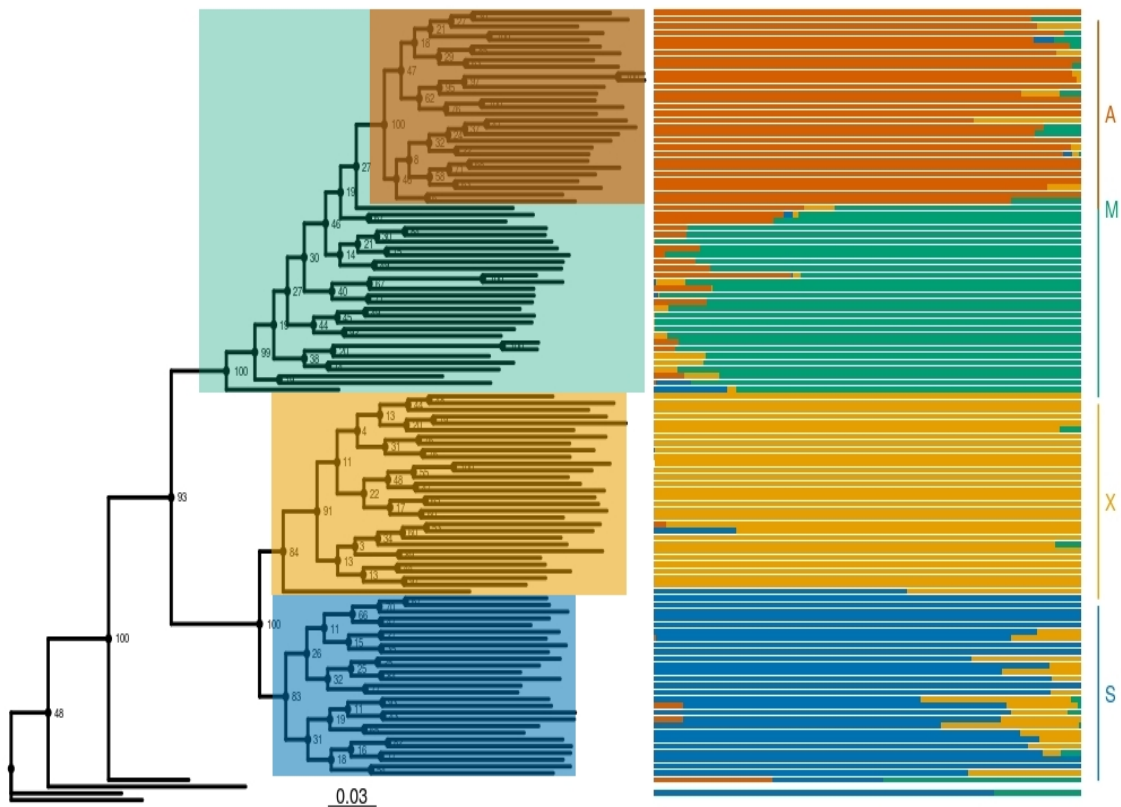
Figure A.10: Phylogenetic tree inferred from maximum likelihood analysis using RaxML. Node values represent bootstrap support and the scale bar represents substitutions per site.
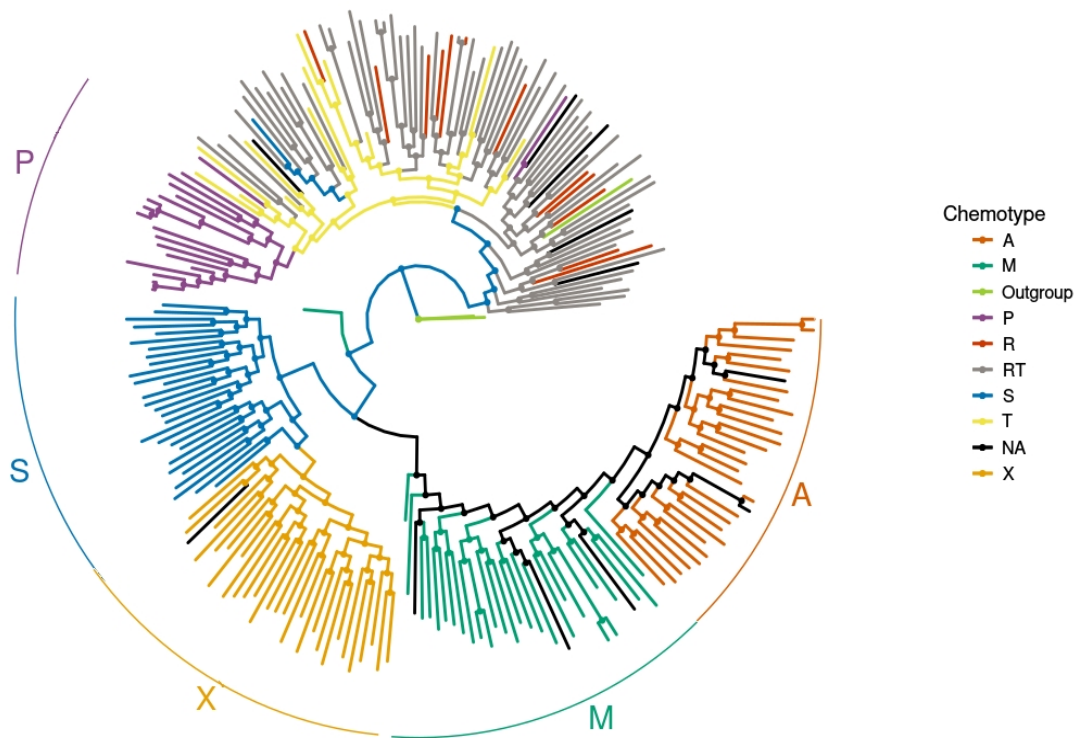
Figure A.11: Phylogenetic tree inferred from maximum likelihood analysis using RaxML. Samples shown in black are samples from which we do not have information about their chemotype, samples shown in grey are samples from La Selva from which we are not sure about their chemotype.
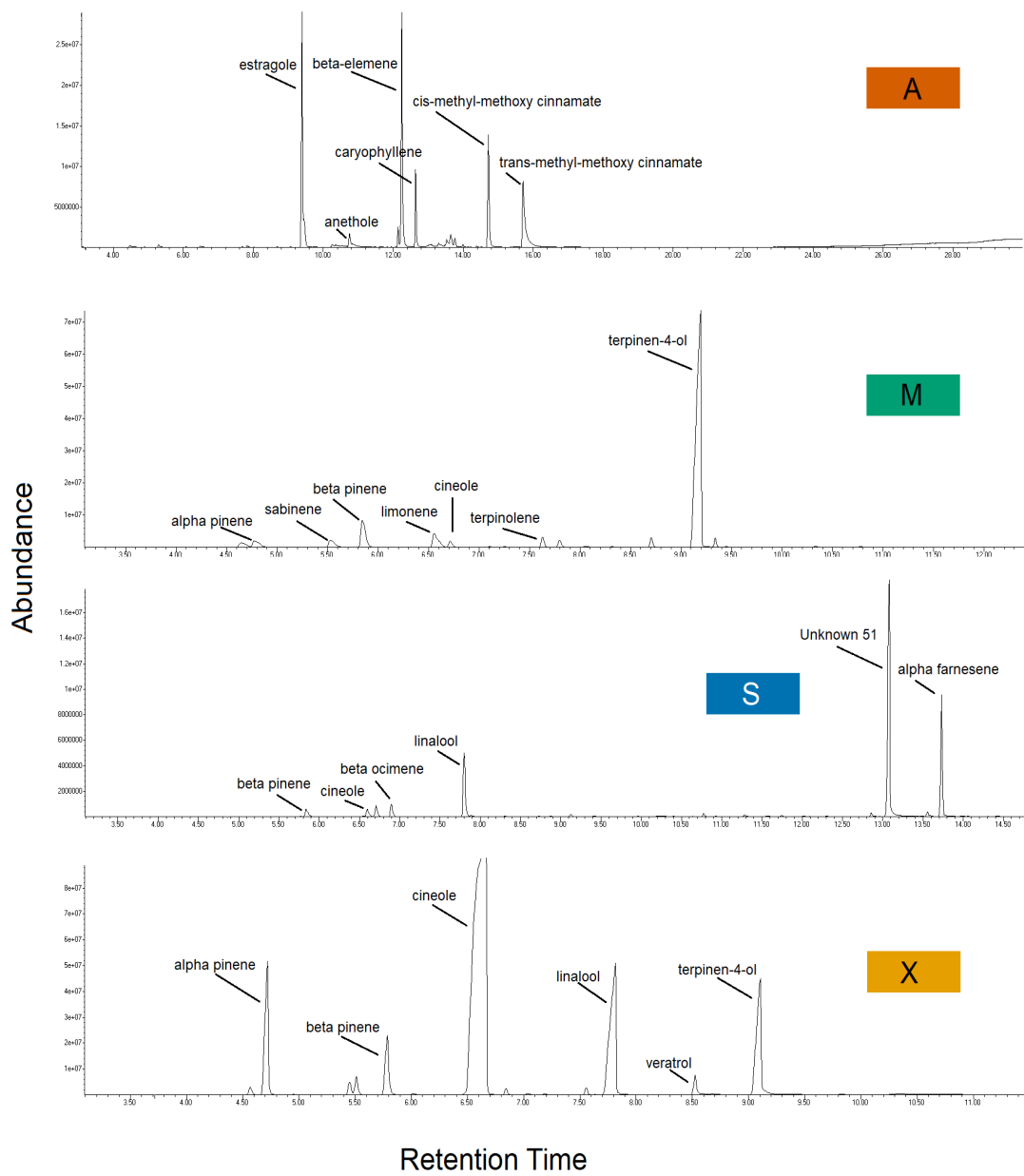
Figure A.12: Representative Gas Chromatogram (GC) plots of the four *Gongora* chemotypes from La Gamba.
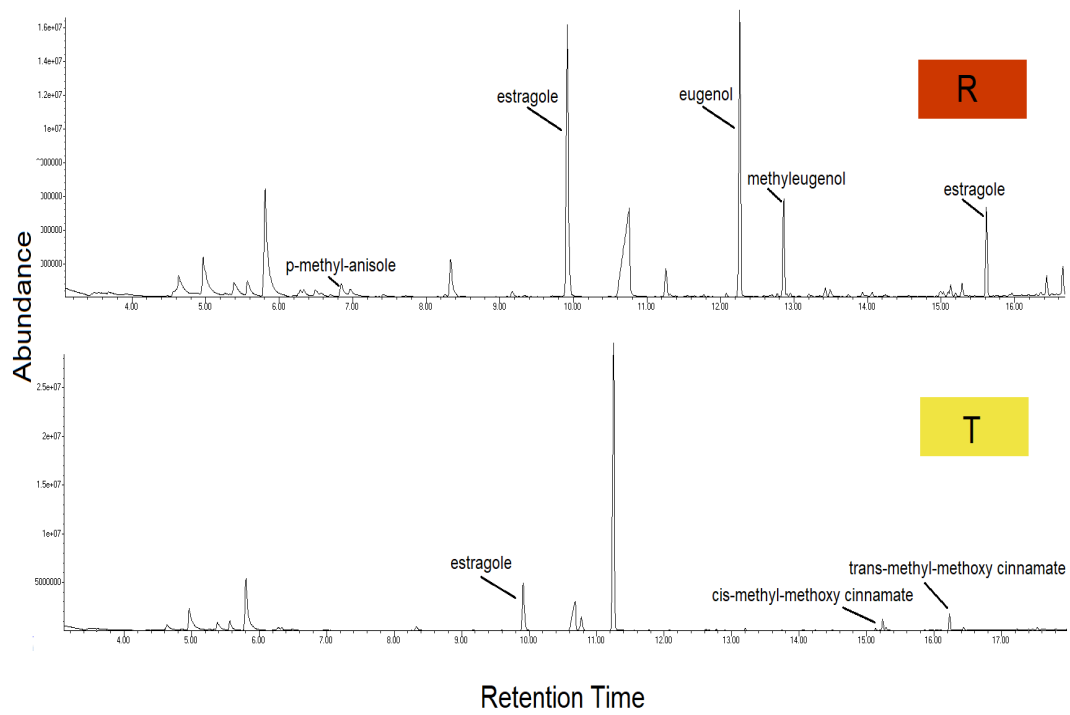
Figure A.13: Representative Gas Chromatogram (GC) plots of chemotypes R and T from La Selva.
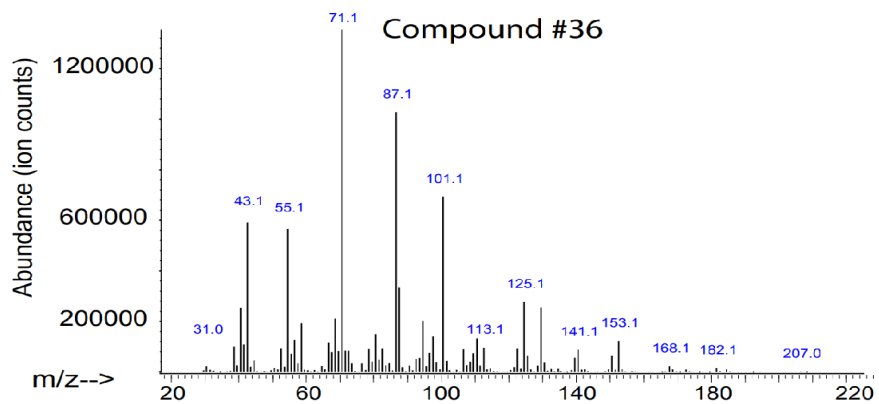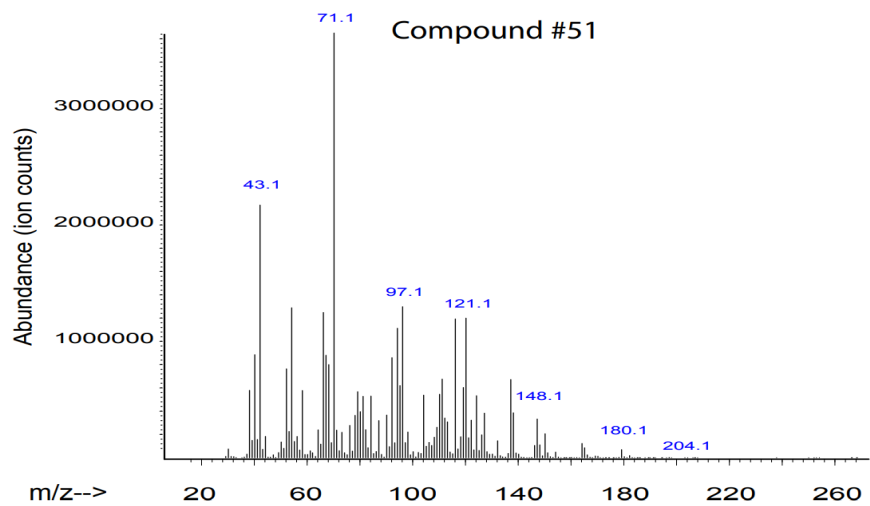
Figure A.14: Mass spectra from Chemotype S unknown compounds 51 and 36.
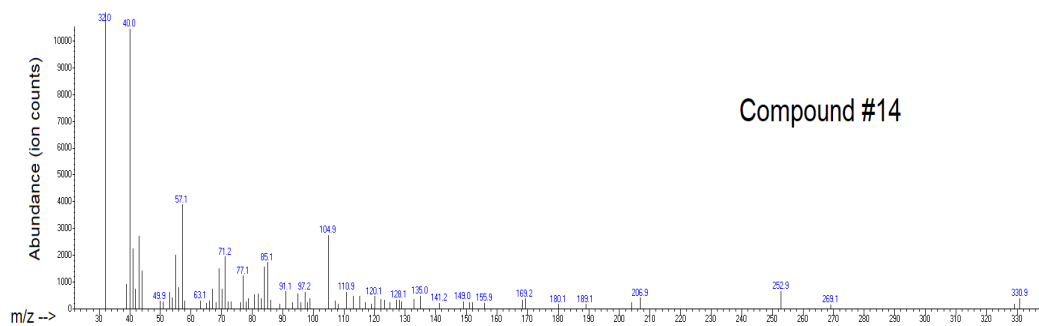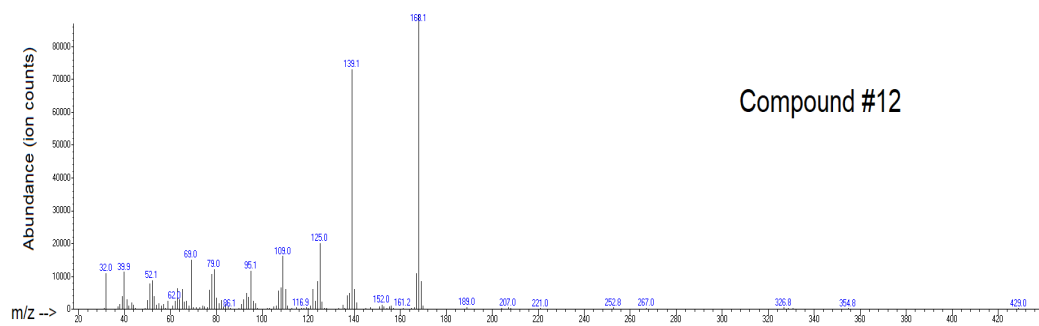
Figure A.15: Mass spectra from Chemotype P unknown compounds 12 and 14.

Appendix B

Appendix Chapter 02

B.1   Supplementary Methods

Genome assembly

FMLRC read correction

Long reads can improve the contiguity of an assembly; however, this technology is associated with high error rates [147]. We used FMLRC [131] with one of our Illumina libraries to leverage the higher accuracy of the short reads to perform long-read error correction on the PacBio library. Prior to running FMLRC we constructed a BWT of the short-read sequencing data using RopeBWT2 [82]:

```
gunzip -c SRCD2_S1_L001_R?_001.fastq.gz | awk "NR % 4 == 2" | sort
 sort -T ./temp | tr NT TN | ./ropebwt2/ropebwt2 -LR | tr NT TN |
 fmlrc-convert gongora_comp_msbwt.npy
```

The output file was: **16G gongora_comp_msbwt.npy**

Then we ran FMLRC in the XSEDE cluster in one node, with 56 tasks per node and 2000 GB memory:

```
./fmlrc -k 21 -K 59 -p 41 gongora_comp_msbwt.npy                  \
     Pacbio_reads.fasta Pacbio_corrected.fasta
```

The output file was: **69G Pacbio_corrected.fasta**

Chloroplast and mitochondria read filtering

To filter out reads corresponding to the chloroplast and the mitochondria we looked for the closest orchid relative with mitochondria and chloroplast sequences available in NCBI [132] at the moment:

- GQ324949.1 *Oncidium* chloroplast genome

- KJ501920.1 *Oncidium* mitochondrial genome

We used BLAST [8] to find corrected PacBio reads that could belong to the mitochondria or the chloroplast.

```
makeblastdb -in GQ324949.1_Oncidium_chloroplast_genome.fasta      \
            -dbtype nucl -parse_seqids -out Oncidium_chloroplast
makeblastdb -in KJ501920.1_Oncidium_mitochondrial_genome.fasta    \
            -dbtype nucl -parse_seqids -out Oncidium_mitochondria


blastn -db Oncidium_chloroplast -query Pacbio_corrected.fasta     \
       -out Pacbio_chloroplast_hits -sum_stats TRUE               \
       -outfmt "6 qseqid sseqid pident length qgi mismatch        \
       gapopen qstart qend sstart send evalue bitscore"           \
       -num_threads 11
blastn -db Oncidium_mitochondria -query Pacbio_corrected.fasta    \
       -out Pacbio_mitochondria_hits -sum_stats TRUE              \
       -outfmt "6 qseqid sseqid pident length qgi mismatch        \
       gapopen qstart qend sstart send evalue bitscore"           \
       -num_threads 11
```

Then we filtered out reads that had a hit with the mitochondria or the chloroplast and that showed:

- pident (percentage of identical matches) above 90%

- length (alignment length) above 50% of the read's total length

- evalue under 0.0001

With this filtering strategy, we identified a total of 85,553 reads with high probability of belonging to the chloroplast or the mitochondria. We removed these reads from the fasta file that was used in the assembly.

The output file was: **69G Pacbio_corrected_noplastids.fasta**

Genome assembly with WTDGB2

For the initial assembly we used wtdbg2 v.2.5 [112]. We calculated the consensus sequences and then polished this consensus using the Illumina library SRCD2_S1_L001 with minimap2 [83], samtools v1.8 [84] and bwa v0.7.17 [81]:

```
./wtdbg2 -i Pacbio_corrected_noplastids.fasta -o Gongorav1 -f     \
        -t 16 -g 2.228g -X 20.0 -L 3000 -p 21 -S 4 -s 0.05
./wtpoa-cns -t 16 -i Gongorav1.ctg.lay.gz -fo Gongorav1.raw.fa
./minimap2 -t16 -ax map-pb -r2k Gongorav1.raw.fa                  \
        Pacbio_corrected_noplastids.fasta | samtools sort        \
        -@4 > Gongorav1.bam
samtools view -F0x900 Gongorav1.bam | ./wtpoa-cns -t 16          \
        -d Gongorav1.raw.fa -i - -fo Gongorav1.cns.fa
bwa index Gongorav1.cns.fa
bwa mem -t 16 Gongorav1.cns.fa SRCD2_S1_L001_R1_001.fastq.gz     \
        SRCD2_S1_L001_R2_001.fastq.gz | samtools sort -O SAM | .
wtpoa-cns -t 16 -x sam-sr -d Gongorav1.cns.fa -i - -fo Gongorav1.srp.fa
```

The resulting files were:

- 1.9G Gongorav1.raw.fa contianing 24,194 contigs

- 32G Gongorav1.bam

- 1.9G Gongorav1.cns.fa containing 24,194 contigs

- 1.9G Gongorav1.srp.fa containing 24,194 contigs

We mapped the reads from the Illumina library SRCD2_S1_L001 to the assembly Gongorav1.srp.fa using bwa v0.7.17, samtools v.1.8 and bedtools v.2.28.0 [103]:

```
bwa index Gongorav1.srp.fa
bwa mem -t 16 Gongorav1.cns.fa SRCD2_S1_L001_R1_001.fastq.gz     \
    SRCD2_S1_L001_R2_001.fastq.gz | samtools sort               \
    -o gongorav1_SRCD2.bam -
samtools index gongorav1_SRCD2.bam
samtools stats gongorav1_SRCD2.bam
bedtools genomecov -ibam gongorav1_SRCD2.bam                     \
    -bga > gongorav1_SRCD2.coverage
```

The resulting files were:

- 69G gongorav1_SRCD2.bam the sorted and indexed bam file.

- 14G gongorav1_SRCD2.coverage contains the genome coverage data.

We use the coverage file to generate some summary statistics and assess the state of the genome assembly (Table B.2). The average coverage is around 94.57 but we observe several base pairs that had zero coverage (Figure B.1). Half of the regions with no coverage were shorter than 8,514bp.

We extracted all contigs with at least one base pair with 0-coverage and calculated the percentage of their total length that was not covered by any reads. We observe that the mean percentage of a contig's length with no-coverage seems to be correlated with contig length. In other words, contigs with low coverage tend to be shorter (Figure B.2).

A BLAST search of the longest contigs with low coverage against NCBI's nt database confirmed them to be bacterial contamination. We proceeded to sample contigs with different percentages of non-covered length and compared them with BLAST against the nt database. Contigs with more than 70% of their length not covered by any Illumina reads matched bacterial DNA; therefore, we decided to eliminate contigs with more than 70% of their length not covered by any Illumina reads. Filtered out contigs can be found in file: **53K contamination_contigs.csv**.

We filtered out a total of 6,107 contigs. The filtered assembly is **1.8G Gongorav1.ctg.fa** and contains 18,087 contigs.

Scaffolding

SSPACE v3.0 [15] was used for scaffolding with the mate-pair libraries DTG-SG-126-NS_S1 and bwa v.0.7.17:

```
SSPACE_Standard_v3.0.pl -l Libraries.txt -s Gongorav1.ctg.fa     \
      -b Gongorav1_Scaffolding -T 12 -p 1
```

The output file was **1.8G Gongorav1.scf.fa** containing 17,920 scaffolds. Then we used the paired-end library DTG-SG-122 for polishing with pilon v.1.23 [130]. First we mapped the Illumina library to our fasta file:

```
bwa index Gongorav1.scf.fa
bwa mem -t 8 Gongorav1.scf.fa DTG-SG-122_R1_001.fastq.gz          \
   DTG-SG-122_R2_001.fastq.gz | samtools sort -o gongorav1_DTG122.bam -
samtools index gongorav1_DTG122.bam
```

The output file was **185G gongorav1_DTG122.bam**.

Because pilon v.1.23 has high memory requirements and the bam file is large, we divided the assembly and the bam file into 35 smaller files to allow for at least 1Gb of memory for every 1 Mbp, and we merged all files at the end.

```
java -jar -Xms124g -Xmx124g  pilon-1.23.jar  --genome Group0.fa  \
    --frags Group0.bam  --output Gongorav1_Gr0 --outdir Group0  \
    --diploid --threads 8
```

The polished output file was **1.8G Gongorav1.pil.fa** containing 17,920 scaffolds. This version of the assembly was submitted to Dovetail for further scaffolding resulting in the final assembly version: **Gongorav1.chi.fa**.

Chloroplast and mitochondria assembly

From the final assembly we ran BLAST v.2.7.1 to compare every scaffold to the *Oncidium* chloroplast and mitochondrial genomes.

```
makeblastdb -in Gongorav1.chi.fa -parse_seqids -out Gongorav1.chi\
      -dbtype nucl
blastn -db Gongorav1.chi -sum_stats TRUE -outfmt  6                  \
      -query GQ324949.1_Oncidium_chloroplast_genome.fasta        \
      -out CHI_chloroplast_hits  -num_threads 8
blastn -db Gongorav1.chi -sum_stats TRUE -out fmt 6                  \
      -query KJ501920.1_Oncidium_mitochondrial_genome.fasta     \
      -out CHI_mitochondria_hits  -num_threads 8
```

We identified 4 scaffolds that matched to the *Oncidium* chloroplast genome in at leas 50% of their lenght and 1 scaffold that matched the mitochondria. Using these scaffolds together with the previously 85,553 filtered PacBio reads matching the mitochondria and chloroplast, we assembled both genomes with Canu v.1.6 [77]:

```
canu -assemble -p cu.1 -d Plast.cu genomeSize=250K               \
    -pacbio-corrected Plastids_all.fa useGrid=false
```

The assembly resulted in 76 contigs totalling 2,183,254 bp (including 6 repeats of total length 240,243 bp). The NG50 was 187,376. Every contig was blasted against the NCBI nt database to determine if it belongs to the chloroplast or the mitochondria. In total, 63 contigs matched chloroplast sequences and 13 contigs matched mitochondrial sequences.

For scaffolding of the chloroplast we used SSPACE v3.0 using the mate-pair library DTG-SG-126-NS_S1.

```
SSPACE_Standard_v3.0.pl -l Libraries.txt  -s cpDNA_ctg.fasta      \
      -b cpDNA_scf -T 8 -p 1
```

We polished the chloroplast and mitochondrial assemblies with pilon v.1.23 and the paired-end library SRCD2_S1_L002 :

```
bwa index cpDNA_scf.fasta

bwa mem -t 8 cpDNA_scf.fasta SRCD2_S1_L002_R1_001.fastq.gz        \
    SRCD2_S1_L002_R2_001.fastq.gz | samtools sort -o cpDNA_temp.bam -

samtools index cpDNA_temp.bam

samtools view -b -F 4 cpDNA_temp.bam -o cpDNA_SRCD2S1.bam

samtools index cpDNA_SRCD2S1.bam

pilon --genome cpDNA_scf.fasta --frags  cpDNA_SRCD2S1.bam         \
      --output Gongorav1_cpDNA --outdir Gongorav1_cpDNA           \
      --fix all --threads 8

bwa index mtDNA_ctg.fasta

bwa mem -t 8 mtDNA_ctg.fasta SRCD2_S1_L002_R1_001.fastq.gz        \
    SRCD2_S1_L002_R2_001.fastq.gz | samtools sort -o mtDNA_temp.bam -

samtools index mtDNA_temp.bam

samtools view -b -F 4 mtDNA_temp.bam -o mtDNA_SRCD2S1.bam

samtools index mtDNA_SRCD2S1.bam

pilon --genome mtDNA_ctg.fasta --frags  mtDNA_SRCD2S1.bam         \
      --output Gongorav1_mtDNA --outdir Gongorav1_mtDNA           \
      --fix all --threads 8
```

The output assembly for the mitochondria contains 13 contigs and 462,164 bp. The chloroplast assembly was contained in a single scaffold 187,299 bp in length.

Assembly quality control

The final assembly and all intermediate assemblies were evaluated using QUAST v.5.0.2 [57] and BUSCO v.3.0.2 :

```
python quast-5.0.2/quast.py Gongorav1.pil.fa Gongorav1.chi.fa    \
      -o Gongorav1 --threads 8 --labels Pil,Chi --eukaryote      \
      --large --min-contig 1000 --k-mer-stats                    \
      --est-ref-size 2228000000

python $BUSCO_DIR/scripts/run_BUSCO.py -i Gongorav1.pil.fa        \
```

```
        -c 8 -o Gongorav1.pil.fa -m geno -l liliopsida_odb10
python $BUSCO_DIR/scripts/run_BUSCO.py -i Gongorav1.chi.fa        \
        -c 8 -o Gongorav1.chi.fa -m geno -l liliopsida_odb10
```

Genome annotation

Custom repeat library

For the whole-genome *de novo* repeat annotation we used the Extensive de-novo TE Anno-
tator (EDTA) v.1.9.8 [98]. This pipeline results in a species-specific repeat library that can
be used to mask the genome previous to gene annotation.

The names of the scaffolds in the genome assembly were simplified to be compatible with
EDTA. The analysis consists of different steps and we ran every individually. The first step
identifies all long terminal repeat (LTR) transposable elements (TEs):

```
perl ../EDTA_raw.pl --genome Gongorav1.fa --type ltr           \
    --overwrite 1 --threads 8
```

The second and third steps identify tandem inverted repeats (TIR) and Helitrons, re-
spectively:

```
perl ../EDTA_raw.pl --genome Gongorav1.fa  --type tir          \
    --overwrite 0 --threads 8


perl ../EDTA_raw.pl --genome Gongorav1.fa  --type helitron      \
    --overwrite 0 --threads 8
```

The last step filters the results from all previous steps and outputs the custom repeat
library:

```
perl ../EDTA.pl --genome Gongorav1.fa --overwrite 0            \
    --step filter  --threads 8
```

To complement the results from the EDTA pipeline we ran RepeatModeler v.2.0.1 [46],
which is another *de novo* transposable element family identidication and modeling software.

```
BuildDatabase -name Gongorav1 Gongorav1.fa
RepeatModeler -database Gongorav1 -pa 12
```

To merge the results from both approaches we used USEARCH [7] to cluster the identified repeat sequences with at least 80% identity and to remove all but one sequence from each cluster.

```
cat Gongorav1.fa.mod.EDTA.TElib.fa                        \
    Gongorav1-families.fa > Gongorav1_lib1.fa
usearch -cluster_fast Gongorav1_lib1.fa -id 0.8           \
        -centroids Gongorav1_lib2.fa
mv Gongorav1_lib2.fa Gongorav1_lib1.fa
```

The output file **Gongorav1_lib1.fa** is the custom repeat library for *Gongora*. It is non-redundant and contains 6,423 elements. Using this library and RepeatMasker v.4.1.2-p1 [25] we masked 83.26% of the genome:

```
./RepeatMasker -pa 8 -e ncbi -lib Gongorav1_lib1.fa       \
    -dir Denovo_mask Gongorav1.fa -libdir Libraries        \
    -trf_prgm trf409.linux64
```

We ran RepeatMasker a second time using the monocots library from Repbase:

```
./RepeatMasker -pa 8 -e ncbi -species monocots            \
              -dir Gongora_mask Gongorav1.fa.masked       \
              -libdir Libraries -trf_prgm trf409.linux64
```

This step masked an additional 0.1% of the genome. We merged the results of both masking steps to produce a final repeat annotation for *Gongora*.

Annotation with MAKER

For the annotation of the *Gongora* genome assembly we used the MAKER pipeline [21] with the following evidence data sets:

- A *de novo* transcript assembly of RNA-seq data generated from floral tissue of a *Gongora* chemotype A sample (Trinity-GG.fasta)

- An EST sequence file from *Phalaenopsis aphrodite* downloaded from Orchidstra 2.0 [23] (PATC.fasta)

- A protein sequence file from *Oryza sativa* and *Arabidopsis officinalis* downloaded from EnsemblPlants [62] (protein.fa)

- A custom repeat library (Gongorav1.full_mask.complex.reformat.gff3)

We merged the output files from MAKER into GFF and FASTA outputs:

```
gff3_merge -s -d  Gongora_rnd1_master_datastore_index.log >        \
      Gongora_rnd1.all.maker.gff
fasta_merge -d Gongora_rnd1_master_datastore_index.log
```

This resulted in the annotation of 20,496 genes.

TPS genes

TPS gene annotation

From Pfam [91] we downloaded multiple sequence alignments for the conserved TPS domains: PF03936 (C terminal domain) and PF01397 (N-terminal domain). Both alignment files were converted into block profiles to be used with augustus:

```
msa2prfl.pl --prefix_from_seqnames --max_entropy=0.75            \
          --blockscorefile=PF03936_seed.blocks.txt            \
           PF03936_seed.txt > PF03936_seed.prfl
msa2prfl.pl --prefix_from_seqnames --max_entropy=0.75            \
          --blockscorefile=PF01397_seed.blocks.txt            \
           PF01397_seed.txt > PF01397_seed.prfl
```

Using the profiles for both domains we ran a preliminary fast block search to identify which scaffolds contain possible profile hits:

```
fastBlockSearch --cutoff=1.1 Gongorav1.fa PF03936_seed.prfl
fastBlockSearch --cutoff=1.1 Gongorav1.fa PF01397_seed.prfl
```

From the genome assembly we extracted all scaffolds that could contain a TPS gene and saved them into a separate FASTA file called ScYo1bC_torun.fasta. Then we ran augustus [120] with the maize training parameter set:

```
augustus --optCfgFile=ppx.cfg --proteinprofile=PF03936_seed.prfl \
        ScYo1bC_torun.fasta > Scaffolds.ppx.gff
augustus --optCfgFile=ppx.cfg --proteinprofile=PF01397_seed.prfl \
        ScYo1bC_torun.fasta > PF01397.ppx.gff
```

Together both profiles identified 2,266 candidate TPS genes, we expect most of these sequences to be false positives. To narrow down the results we downloaded orchid TPS sequences from [143] and used them together with the candidate TPS genes to perform multiple sequence alignments and phylogenetic analyses with MEGA [123]. Only candidate TPS sequences clustering together with the previously reported orchid TPS sequences were considered for further analysis. After manual inspection and annotation of these sequences, we identified 21 final candidate TPS genes.

Using the results from the genome annotation pipeline we further improved the gene annotations and identified the number and type of transposable elements in the introns and exons of each TPS genes, as well as the TEs present in the 0-10kb and 10kb-20kb regions around each gene.

Whole genome re-sequencing

We selected 7 and 6 individuals of chemotypes A and M, respectively, for genotyping using a shotgun sequencing approach with 150bp paired-end Illumina libraries and a target depth of coverage of 8X. Individual plants were selected based on available phenotypic data (B.10, B.11, B.12, B.13, B.14,B.5). Data quality was assessed with FastQC v.0.11.7 [45].

The libraries were mapped to the *Gongora* reference genome using bwa:

```
bwa index Gongorav1.fa
bwa mem -t 8 Gongorav1.fa                                         \
    G01/G01_CSFP210002974-1a_H57FNDSX2_L1_1.fq.gz                 \
    G01/G01_CSFP210002974-1a_H57FNDSX2_L1_2.fq.gz                 \
    | samtools \sort -o G01.bam -
samtools sort -n -o Sorted_G01.bam -O BAM G01.bam
samtools fixmate -m Sorted_G01.bam Fixmate_G01.bam
samtools sort -o Sorted_G01.bam Fixmate_G01.bam
samtools markdup -r -s Sorted_G01.bam G01.1.bam
samtools index G01.1.bam
samtools stats G01.1.bam > G01_bam.stats
```

We then estimated inbreeding coefficients with ANGSD [78] and ngsF from ngsTools [47]. We first calculated genotype likelihoods for the A and M chemotype samples separately:

```
angsd -P 8 -b ChemA.bamlist -ref Gongorav1.fa                    \
      -out Results/ChemA -uniqueOnly 1 -remove_bads 1            \
      -only_proper_pairs 1 -trim 0 -C 50 -baq 1 -minMapQ 23      \
      -minQ 20 -minInd 3 -setMinDepth 27 -setMaxDepth 148        \
      -doCounts 1 -GL 1 -doMajorMinor 1 -doMaf 1                 \
      -skipTriallelic 1 -doGlf 3 -SNP_pval 1e-3
angsd -P 8 -b ChemM.bamlist -ref Gongorav1.fa                    \
      -out Results/ChemM -uniqueOnly 1 -remove_bads 1            \
      -only_proper_pairs 1 -trim 0 -C 50 -baq 1 -minMapQ 23      \
      -minQ 20 -minInd 3 -setMinDepth 27  -setMaxDepth 148       \
      -doCounts 1 -GL 1 -doMajorMinor 1 -doMaf 1                 \
      -skipTriallelic 1 -doGlf 3 -SNP_pval 1e-3
```

Then we estimated starting values for the inbreeding coefficients and used these to perform a deep search with ngsF:

```
zcat ChemA.glf.gz | ngsF --n_ind 7 --n_sites 27347619 --glf -   \
```

```
    --out ChemA.approx_indF --approx_EM --init_values u        \
    --n_threads 8
zcat ChemA.glf.gz | ngsF --n_ind 7 --n_sites 27347619  --glf -   \
    --out ChemA.indF --init_values ChemA.approx_indF.pars        \
    --n_threads 8
zcat ChemM.glf.gz | ngsF --n_ind 6 --n_sites 27493444           \
    --glf - --out ChemM.approx_indF --approx_EM                  \
    --init_values u --n_threads 8
zcat ChemM.glf.gz | ngsF --n_ind 6 --n_sites 27493444           \
    --glf - --out ChemM.indF                                    \
    --init_values ChemM.approx_indF.pars --n_threads 8
```

Population genetic differentiation

We first calculated the site frequency spectrum (SFS) with ANGSD. The SFS records the proportions of sites at different allele frequencies. We calculated the sample allele frequency likelihoods at each site for each chemotype separately:

```
angsd -P 4 -b ChemA.bamlist -ref Gongorav1.fa  -anc Gongorav1.fa \
    -out Results/ChemA -uniqueOnly 1 -remove_bads 1              \
    -only_proper_pairs 1 -trim 0 -C 50 -baq 1 -minMapQ 23        \
    -minQ 20 -minInd 3 -setMinDepth 27 -setMaxDepth 148          \
    -doCounts 1 -GL 1 -doSaf 1
angsd -P 4 -b ChemM.bamlist -ref Gongorav1.fa -anc Gongorav1.fa  \
    -out Results/ChemM -uniqueOnly 1 -remove_bads 1              \
    -only_proper_pairs 1 -trim 0 -C 50 -baq 1 -minMapQ 23        \
    -minQ 20 -minInd 3 -setMinDepth 27 -setMaxDepth 148          \
    -doCounts 1 -GL 1 -doSaf 1
```

We then used these likelihoods to estimate the overall SFS using realSFS:

```
realSFS ChemA.saf.idx -P 4 2> /dev/null > ChemA.sfs
realSFS ChemM.saf.idx -P 4 2> /dev/null > ChemM.sfs
```

To calculate the joint SFS:

```
realSFS -P 4 ChemA.saf.idx ChemM.saf.idx 2> /dev/null > ChemA.ChemM.sfs
```

Using the joint SFS we estimate the allele frequency differentiation between A and M individuals by computing Fst values across the genome in sliding windows of 50kbp and a step size of 10kbp:

```
realSFS fst index ChemA.saf.idx ChemM.saf.idx                    \
        -sfs ChemA.ChemM.sfs -fstout ChemA.ChemM -whichFST 1
realSFS fst stats2 ChemA.ChemM.fst.idx -win 50000               \
        -step 10000 -whichFST 1 > ChemA.ChemM.fst.txt
```

We also used vcftools v.0.1.17 [33] to call variants and perform additional analyses. After aligning the libraries to the genome assembly, we called variants with the bcftools v.1.6 mpileup tool [34]:

```
bcftools mpileup -a AD,DP,SP -Ou -f Gongorav1.fa --max-depth 148 \
                 -b All.bamlist | bcftools call -f GQ,GP -mv      \
                 -Oz -o gongora.vcf.gz
bcftools index gongora.vcf.gz
```

We filtered the resulting VCF file with vcftools:

```
vcftools --gzvcf gongora.vcf.gz --remove-indels --maf 0.08        \
        --max-missing 0.95 --minQ 30 --min-meanDP 5               \
        --max-meanDP 20 --minDP 5 --maxDP 20 --recode             \
        --stdout | gzip -c > gongora_filtered.vcf.gz
```

After filtering, we kept 3,655,988 sites out of a possible 92,387,041 sites.

To estimate the mean genome-wide Fst, we used vcftools:

```
vcftools --gzvcf gongora_filtered.vcf.gz                          \
        --weir-fst-pop ChemA.bamlist                              \
        --weir-fst-pop ChemM.bamlist --out ChemA_ChemM
```

The mean genome-wide Fst was 0.01915. From these results we identified 3 scaffolds with regions showing high levels of genetic differentiation (Fst=1)(Table B.15).

To calculate sliding window estimates for Fst, dxy and pi across the genome, we repeated the genotype call with mpileup but we retained all monomorphic sites:

```
bcftools mpileup -a AD,DP,SP -Ou -f Gongorav1.fa --max-depth 148  \
        -b All.bamlist | bcftools call -f GQ,GP -m -Oz            \
        -o gongora_all.vcf.gz
bcftools index gongora_all.vcf.gz
vcftools --gzvcf gongora_all.vcf.gz --remove-indels               \
        --max-alleles 2 --max-missing 0.95 --minQ 30              \
        --min-meanDP 5 --max-meanDP 20 --minDP 5 --maxDP 20       \
        --recode --stdout | gzip -c > gongora_filtered2.vcf.gz
parseVCF.py -i gongora_filtered2.vcf.gz | bgzip > gongora_filtered.geno.gz
```

Last script from Simon Martin. Then we calculate the sliding window estimates:

```
popgenWindows.py -g gongora_filtered.geno.gz                     \
                -o gongora.Fst.Dxy.pi.ANGSD.csv.gz               \
                -f phased -w 50000 -m 5000 -s 10000 -p ChemA     \
                -p ChemM  --popsFile pop.info
```

Principal Component Analysis

To perform a principal component analysis (PCA) we used the filtered VCF gongora_filtered.vcf.gz and PLINK v1.90p [101]. We first pruned the dataset of variants that are in possible linkage:

```
plink --vcf gongora_filtered.vcf.gz --double-id --allow-extra-chr \
     --set-missing-var-ids @:# --indep-pairwise 50 10 0.1         \
     --out gongoras
```

After the linkage pruning we removed 3,472,212 variants. With the remaining variants we ran the PCA:

```
plink --vcf gongora_filtered.vcf.gz --double-id --allow-extra-chr \
    --set-missing-var-ids @:# --extract gongoras.prune.in        \
    --make-bed --pca --out gongoras
```

For this analysis, 183,776 variants were used with a total genotyping rate of 99.1%. The results were plotted in R using packages tidyverse [134] and ggplot2 [135].

B.2 Supplementary Figures
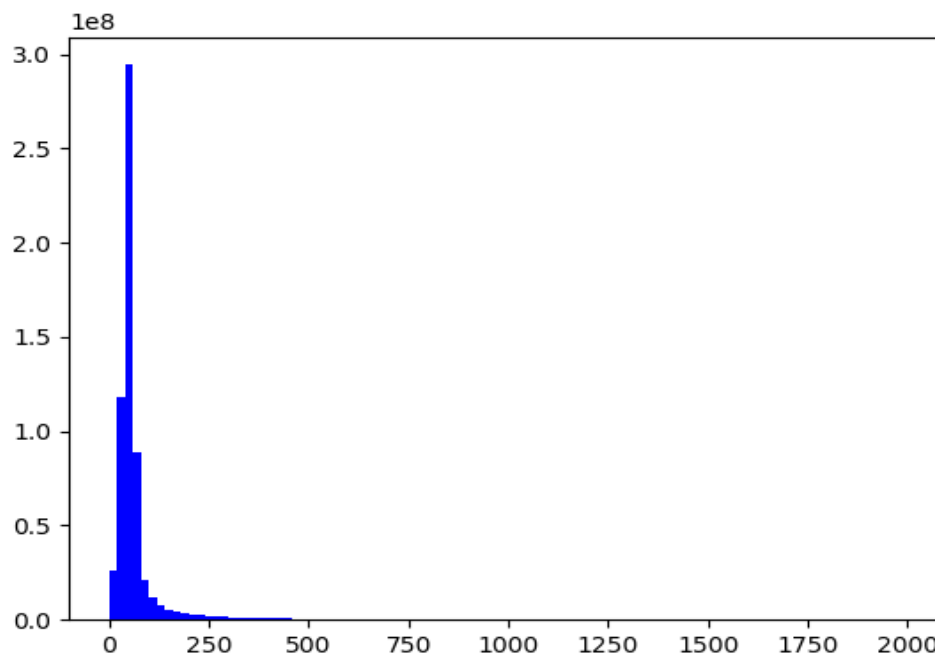
Figure B.1: Per base coverage histogram after mapping the Illumina SRCD2_S1_L001 library to the assembly Gongorav1.srp.fa.
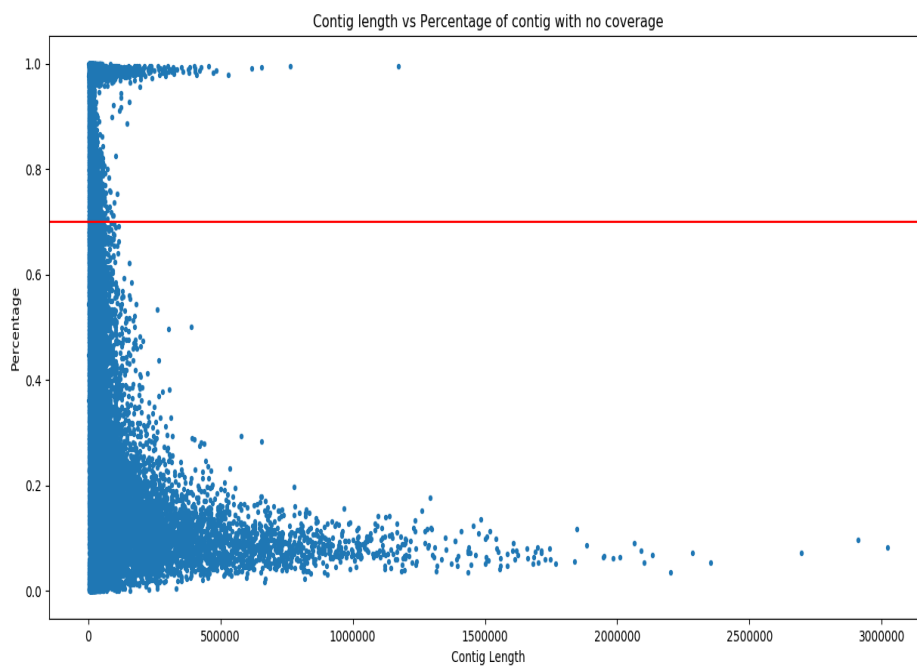
Figure B.2: Scatterplot showing the relationship between contig size (bp) and the percentage of it's length that was not covered by any reads from the Illumina SRCD2_S1_L001 library. Regions with no to low coverage could be indicative of misassemblies or exogenous DNA.
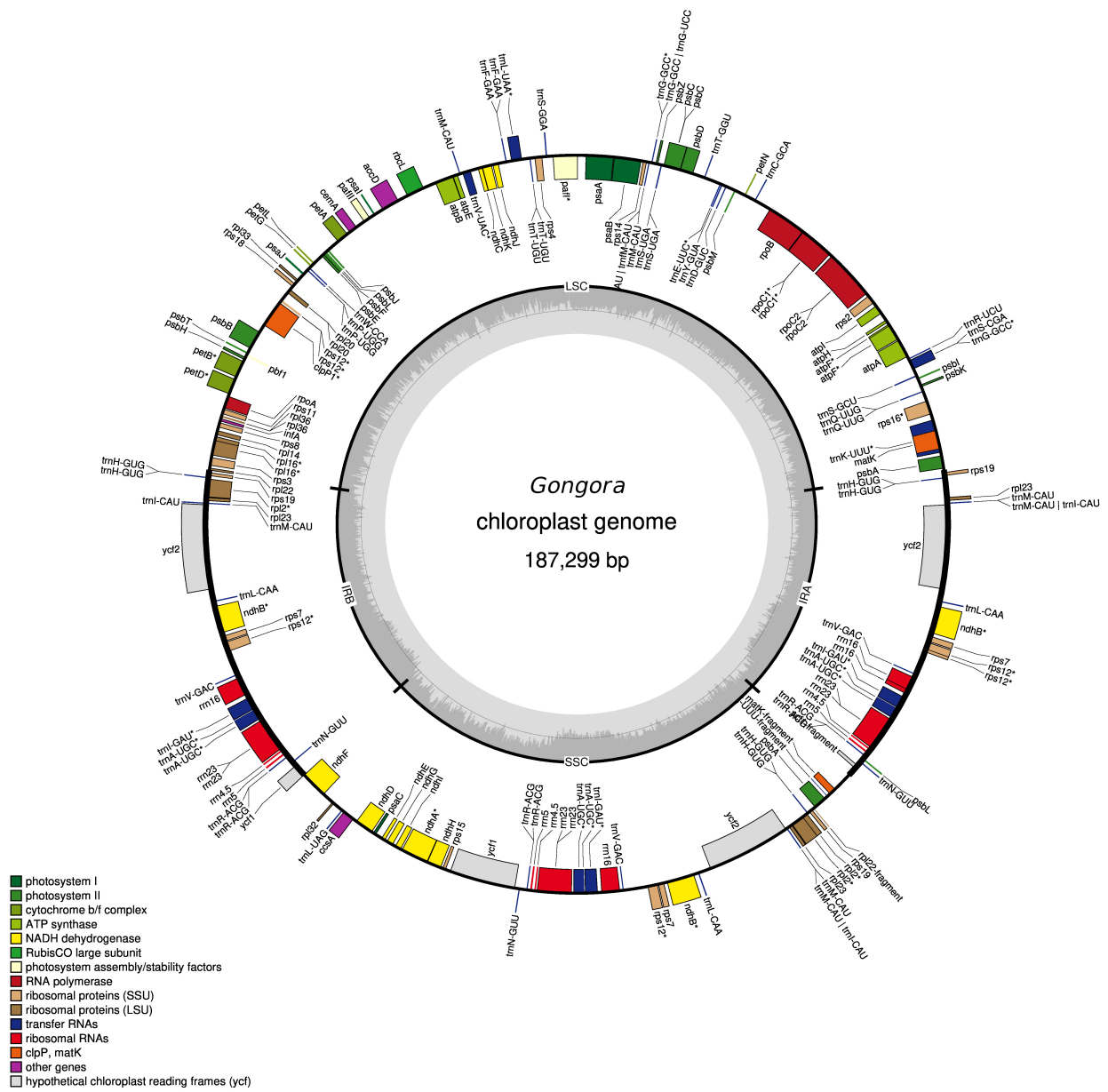
Figure B.3: Chloroplast genome assembly.
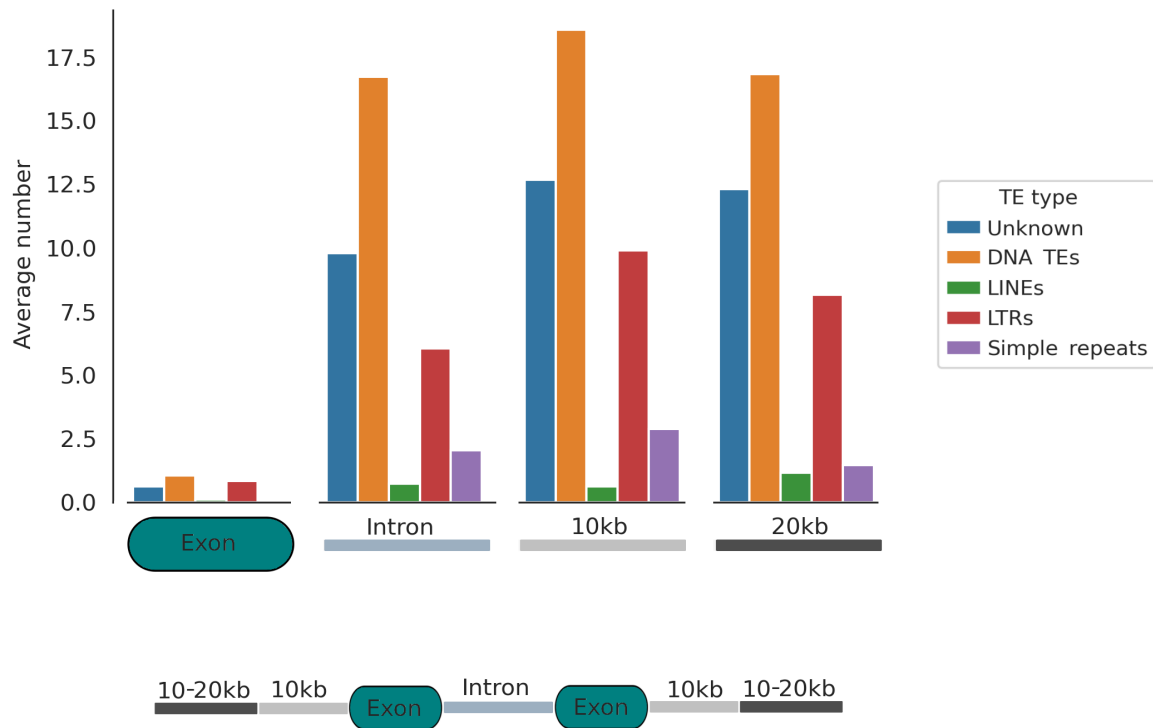
Figure B.4: Repetitive elements associated with terpene synthase (TPS) genes in *Gongora*. Repetitive elements were categorized as Unknown, DNA TEs, LINEs, LTRs or Simple repeats. All elements were identified in the exons and introns of TPS genes, as well as in the 0-10kb and 10kb-20kb regions surrounding the coding sequence.
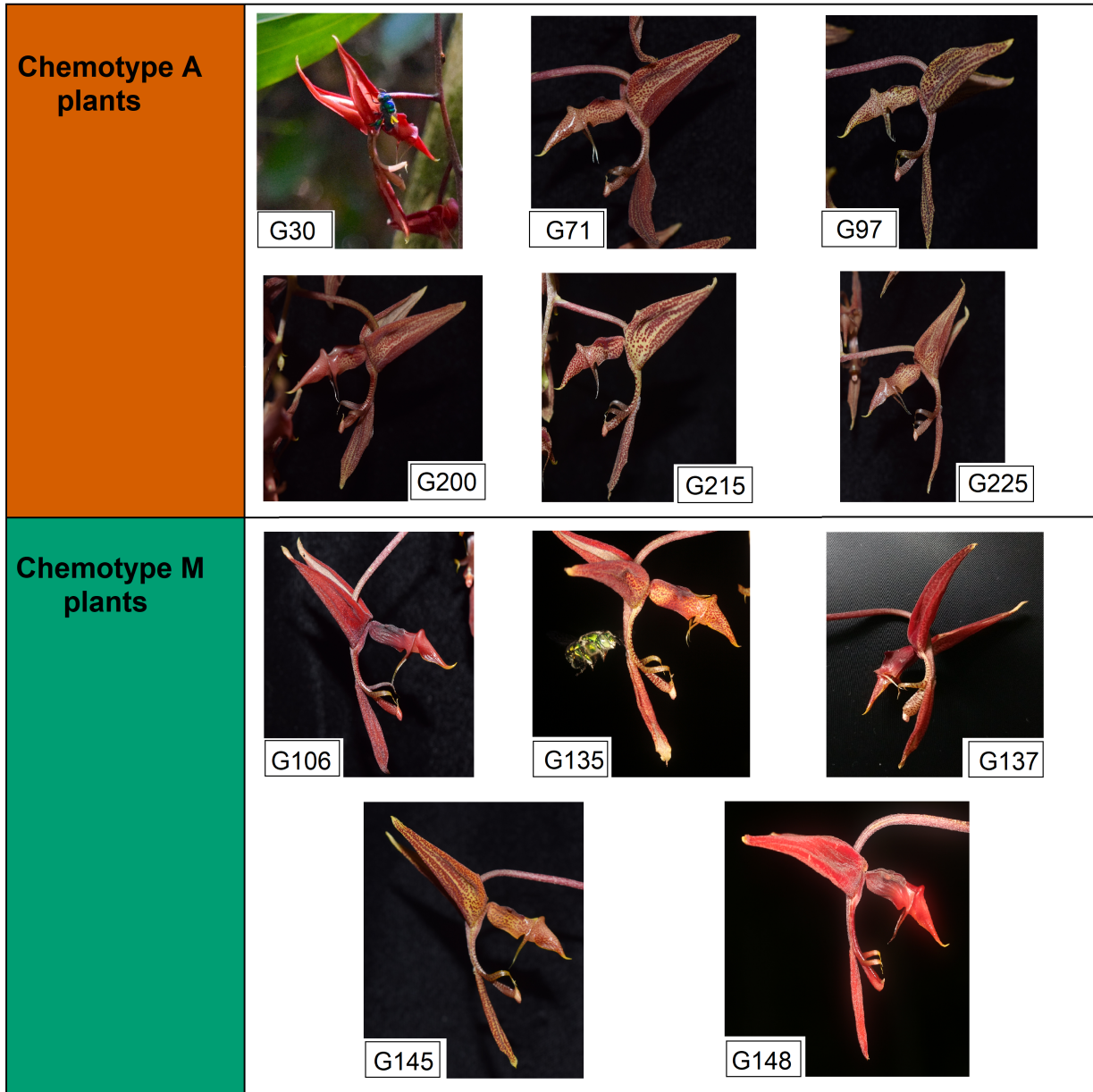
Figure B.5: Photographs of flowers from the A and M chemotype individuals used for whole genome sequencing.
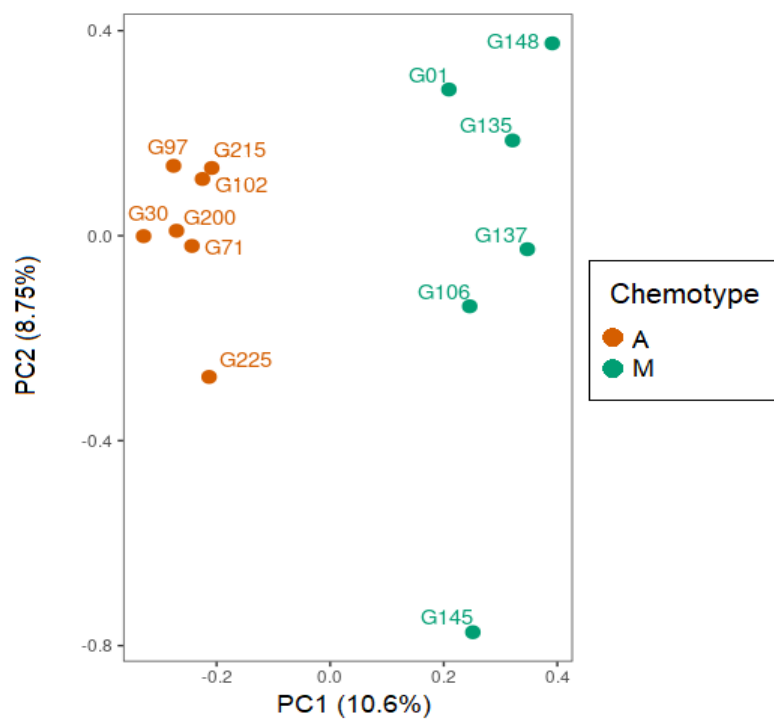
Figure B.6: Principal component analysis based on single nucleotide polymorphisms (SNPs) from chemotype A and M individuals collected from La Gamba.

B.3 Supplementary Tables

Table B.1.   Libraries used for genome assembly

Library name	Type	File size
Pacbio_reads.fasta	PacBio	71Gb
SRCD2_S1_L00_R1	Illumina paired-end	29Gb
SRCD2_S1_L00_R2	Illumina paired-end	34Gb
DTG-SG-122_R1	Illumina paired-end	94Gb
DTG-SG-122_R2	Illumina paired-end	102Gb
DTG-SG-126-NS_S1_R1	Illumina mate-pair	40Gb
DTG-SG-126-NS_S1_R2	Illumina mate-pair	42Gb

Table B.2.   Coverage statistics for Gongorav1.srp

	Gongorav1.srp Coverage
Mean coverage	94.57
Standard deviation	489.84
Minimum	0
25%	40
50%	50
75%	62
Maximum	417435

Note. — Summary statistics for the genomic coverage of Gongorav1.srp using the SRCD2_S1_L001 library.

Table B.3.   Genome assembly quality control

	Gongorav1.pil.fa	Gongorav1
N50	382,585	1,756,489
L50	1,308	262
Total length (Gbp)	1.831	1.833
GC%	32.6	32.6
Number of Ns per 100 kbp	0.09	73.04
Longest contig/scaffold (Mbp)	3.02	17.68
Number of contigs/scaffolds	17,920	9,028

Table B.4.   Orchid genome assemblies part 1

	P. equestris	D. officinale	D. catenatum	A. shenzhenica	G. elata	P. aphrodite
Year published	2015	2015	2016	2017	2018	2018
Estimated genome size (Gb)	1.6	1.27	1.11	0.471	1.18	1.2
Assembled genome size (Gb)	1.086	1.66	1.01	0.349	1.06	1.025
Scaffold N50 (Mbp)	0.359	0.076	0.391	3.029	4.9	0.946
Contig N50 (kbp)	20.55	25.12	33.09	80.06	68.9	18.81
Longest scaffold (Mbp)	81.76	1.15	2.59	12.42	24.7	10.39
Repeat content (%)	62	63.33	78.1	42.05	66.18	60.3
BUSCO/CEGMA assessment (%)	91	91.5	92.46	93.6	96.37	95
Gene number	29431	35567	28910	21841	18969	28902
Total scaffolds	236185	751466	72901	2986	3779	13732
Reference	[20]	[140]	[146]	[145]	[144]	[22]

Table B.5.   Orchid genome assemblies part 2

	A. ramifera	C. sinense[a]	C. goeringii[a]	D. chrysotoxum[a]	D. nobile[a]	Gongora
Year published	2021	2021	2021	2021	2022	**2022**
Estimated genome size (Gb)	0.332	3.52	4.0	1.38	1.16	**2.228**
Assembled genome size (Gb)	0.365	3.45	3.99	1.37	1.19	**1.832**
Scaffold N50 (Mbp)	0.287	NA	NA	NA	64.46	**1.756**
Contig N50 (kbp)	0.030	1110	377.6	1540.9	1618.31	**382.58**
Longest scaffold (Mbp)	1.39	NA	NA	NA	95.36	**17.67**
Repeat content (%)	44.9	77.78	88.87	62.81	61.07	**83.36**
BUSCO/CEGMA assessment (%)	90.9	91	87.8	90.3	96.22	**88**
Gene number	22841	29638	29556	30044	29476	**20496**
Total scaffolds	1494	20	20	19	57	**9024**
Reference	[148]	[142]	[27]	[149]	[138]	

[a]Chromosome-level assembly.

Table B.6.   BUSCO results

Database	eukaryota odb10	viridiplantae odb10	embryophyta odb10	liliopsida odb10
Complete	253 (92.1%)	378 (89%)	1422 (88%)	2673 (82.6%)
Complete and single-copy	228 (89.4%)	373 (87.8%)	1393 (86.3%)	2628 (81.2%)
Complete and duplicated	7 (2.7%)	5 (1.2%)	29 (1.8%)	45 (1.4%)
Fragmented	13 (5.1%)	22 (5.2%)	72 (4.5%)	261 (8.1%)
Missing	7 (2.8%)	25 (5.8%)	120 (7.4%)	302 (9.3%)
Total searched	255 (100%)	425 (100%)	1614 (100%)	3236 (100%)

Note. — Average floral scent composition per chemotype from La Gamba based on the following number of floral scent extractions: 26 from chemotype A, 23 from chemotype M, 11 from chemotype S, 5 from chemotype X.

Table B.7. Repeat elements from the custom repeat library

Element type	Number of elements
Retroelements	
LINEs:	21
CRE-AMbal	1
L1	11
RTE-BovB	9
LTR elements:	1546
Caulimovirus	3
Copia	884
Gypsy	436
Unknown	223
DNA transposons	
CMC-EnSpm	7
DTA	229
DTC	373
DTH	160
DTM	2033
DTT	52
Helitron	80
MULE-MuDR	10
PIF-Harbinger	14
hAT-Ac	7
hAT-Tag1	17
hAT-Tip100	4
MITEs:	802
DTA	222
DTC	61
DTH	50
DTM	459
DTT	10
Rolling circles:	2
Helitron	2
Simple	16

Note. — Transposable elements identified by
EDTA and RepeatModeler.

Table B.8.   TPS genes annotated in *Gongora*

Gene ID	Scaffold	Start	End	BLAST hit
GonTPS1	ScYo1bC_3528;HRSCAF=4727	1028513	1069864	myrcene synthase
GonTPS2	ScYo1bC_3528;HRSCAF=4727	2251803	2265234	ent-copalyl diphosphate synthase
GonTPS3	ScYo1bC_4211;HRSCAF=5619	7505544	7539237	(-)-alpha-terpineol synthase
GonTPS4	ScYo1bC_4211;HRSCAF=5619	7603290	7607317	(-)-alpha-terpineol synthase
GonTPS5	ScYo1bC_4211;HRSCAF=5619	7677564	7682469	(-)-alpha-terpineol synthase
GonTPS6	ScYo1bC_4646;HRSCAF=6182	1515176	1537592	(-)-alpha-terpineol synthase
GonTPS7	ScYo1bC_4646;HRSCAF=6182	1598454	1610497	(-)-alpha-terpineol synthase
GonTPS8	ScYo1bC_4646;HRSCAF=6182	2822316	2834776	ent-copalyl diphosphate synthase
GonTPS9	ScYo1bC_4646;HRSCAF=6182	2857636	2870743	ent-copalyl diphosphate synthase
GonTPS10	ScYo1bC_5319;HRSCAF=7094	645903	649679	terpene-synthase 10
GonTPS11A	ScYo1bC_7603;HRSCAF=10930	16396029	16406215	S-linalool synthase
GonTPS11B	ScYo1bC_7603;HRSCAF=10930	16377246	16385543	S-linalool synthase
GonTPS11C	ScYo1bC_7603;HRSCAF=10930	16335665	16338712	S-linalool synthase
GonTPS12	ScYo1bC_7658;HRSCAF=11206	2250660	2261905	alpha-humulene synthase
GonTPS13	ScYo1bC_7658;HRSCAF=11206	2460439	2474615	alpha-humulene synthase
GonTPS14	ScYo1bC_7842;HRSCAF=12061	55811	58171	S-linalool synthase
GonTPS15	ScYo1bC_7951;HRSCAF=12757	2132842	2143968	alpha-humulene synthase
GonTPS16	ScYo1bC_8127;HRSCAF=13344	583407	591480	ent-kaur-16-ene synthase
GonTPS17	ScYo1bC_8293;HRSCAF=13780	1184456	1199630	(-)-germacrene D synthase
GonTPS18	ScYo1bC_8627;HRSCAF=14165	3924	12557	terpene-synthase 10
GonTPS19	ScYo1bC_8627;HRSCAF=14165	213172	262565	terpene-synthase 10

Note. — Start and End columns show start and end coordinates of the gene annotation within the scaffold.

Table B.9.   TPS genes in orchids

Species	TPS-a	TPS-b	TPS-c	TPS-ef	TPS-g	TPS-h
Gongora	**4**	**9**	**3**	**5**	**0**	**0**
Apostasia shenzhenica	2	4	0	1	2	0
Dendrobium catenatum	13	19 (18[b])	0	2 (4[b])	0	0
Dendrobium officinale	11 (14[b])	16	0 (1[b])	3	0	0
Phalaenopsis aphrodite	6	7	0	4	0	0
Phalaenopsis bellina	2 (1[b])	5 (7[b])	0	2 (3[b])	0	0
Phalaenopsis equestris	4 (5[a])	7	0 (1[a])	5 (4[a])	0	2 (3[a])
Vanilla planifolia	7	12	0	1	7	0

[a]Yu, 2020 reported a different number.

[b]Huang, 2021 reported a different number.

Note. — Number of TPS genes per orchid species.

Table B.10.   Sequenced orchid samples

PlantID	Chemotype	Inbreeding coefficient
G30	A	0.0012
G71	A	0.0010
G97	A	0.0014
G102	A	0.0019
G200	A	0.0014
G215	A	0.0019
G225	A	0.0020
G01	M	0.0018
G106	M	0.0012
G135	M	0.0010
G137	M	0.0006
G145	M	0.0023
G148	M	0.0023

Table B.11.   Scent profiles for chemotype A plants

R.T.	Compound	G30	G71a	G71b	G97	G102	G200	G215	G225
6.5	limonene	5.7	0	0	0	0	0	0	0
9.4	estragole	11.2	27.3	14.4	18.1	18.8	44.1	30.1	23.2
10.2	chavicol	5.7	15.9	10.1	4.4	7.3	2.8	2.3	0
10.7	anethole	1.1	2.9	0	0	1.9	0	0	0
12.1	cinnamic acid	0	3.6	1.9	5.7	1.6	0	0	0
12.2	beta-elemene	29.6	4.7	0	0	7.1	9.6	9.6	0
12.7	caryophyllene	21.4	1.3	0	0	21.4	0	0	0
14.7	cis methyl-methoxycinnamate	5.4	12.9	13.6	15.8	4.4	6.3	6.3	7.7
15.7	trans methyl-methoxycinnamate	19.8	31.4	60.1	60.1	37.5	51.2	51.7	69

Note. — R.T. refers to the retention time during the gas chromatography analysis. Amounts represent relative abundances (%) of each compound.

Table B.12.   Scent profiles for chemotype M plants

R.T.	Compound	G01	G106	G135	G145	G148a	G148b
4.6	alpha-thujene	0	3.4	3.4	3.2	3.3	4.9
4.7	alpha-pinene	2.8	4.4	3.9	4.9	4.3	5.8
5.5	sabinene	2.7	4.2	4.2	4.1	4.2	5.9
5.8	beta pinene	8.7	9.4	8.8	11.5	9.6	11.8
6.5	limonene	3.9	3.1	2.8	2.8	2.9	4.4
6.6	eucalyptol	1.2	2.3	2	2.1	2.2	2.8
6.7	beta ocimene	0	0	0	2	0	2.2
7.5	terpinolene	0	0	0	0	2.5	3.5
9.2	L-4-terpineol	80.5	73.3	74.8	69.4	70.9	58.7

Note. — R.T. refers to the retention time during the gas chromatography analysis. Amounts represent relative abundances (%) of each compound.

Table B.13.　Bee visitor observations for chemotype A plants

PlantID	Date	Day of blooming	Eg. despecta	Eg. imperialis	Eg. townsendi	Eg. tridentata	Eg. variabilis	Eg. villosiventris	El. bombiformis
G30	03/19/2016	1	3	1	0	0	1	1	1
G30	03/20/2016	2	2	0	1	0	1	0	0
G71	04/15/2015	NA	0	1	0	0	0	1	0
G71	04/10/2016	1	2	0	0	0	1	0	0
G71	04/11/2016	2	0	0	0	0	1	0	0
G97	04/19/2015	1	0	2	0	0	0	0	0
G97	04/20/2015	2	1	3	0	1	1	0	0
G102	03/08/2016	2	0	0	0	0	4	0	0
G200	05/18/2016	2	1	0	0	0	3	0	0
G215	05/18/2016	2	3	0	0	0	2	0	0
G225	05/18/2016	2	1	2	0	0	0	1	1

Note. — Date corresponds to the date when the observations were made and day of blooming refers to number of days after anthesis. Values are given in absolute numbers of bees observed visiting a plant.

Table B.14. Bee visitor observations for chemotype M plants

PlantID	Date	Day of blooming	Eg. erythrochlora	Eg. hansoni	Eg. heterosticta	Eg. tridentata	El. cingulata	Ex. smaragdina
G01	03/16/2013	NA	0	1	0	3	1	0
G01	03/21/2014	NA	0	0	0	1	0	0
G106	04/12/2016	1	2	0	0	1	0	0
G135	03/25/2016	1	2	0	0	4	0	0
G135	03/26/2016	2	1	0	1	5	0	1
G135	05/04/2016	2	0	1	0	0	0	0
G145	06/09/2016	1	0	1	0	1	0	0
G148	03/24/2016	2	0	1	0	7	1	0
G148	03/28/2016	1	1	0	0	5	0	0
G148	05/04/2016	2	0	1	0	1	0	0

Note. — Date corresponds to the date when the observations were made and day of blooming refers to number of days after anthesis. Values are given in absolute numbers of bees observed visiting a plant.

85

Table B.15.   Scaffolds with high Fst values

Scaffold ID	Start	End	Region size (bp)
ScYo1bC_2192;HRSCAF=2933	3725184	3831842	106,658
ScYo1bC_6298;HRSCAF=8377	322338	774872	452,534
ScYo1bC_6298;HRSCAF=8377	1233961	1441693	207,732
ScYo1bC_7675;HRSCAF=11265	46956	778313	731,357
ScYo1bC_7675;HRSCAF=11265	2403033	2645881	242,848
ScYo1bC_7675;HRSCAF=11265	3409904	3471819	61,915

References

[1] Ackerman, J. D. (1989). Geographic and seasonal variation in fragrance choices and preferences of male euglossine bees. *Biotropica*, pages 340–347.

[2] Adams, R. P. et al. (2007). *Identification of essential oil components by gas chromatography/mass spectrometry*, volume 456. Allured publishing corporation Carol Stream.

[3] Ågren, J., Hellström, F., Toräng, P., and Ehrlén, J. (2013). Mutualists and antagonists drive among-population variation in selection and evolution of floral display in a perennial herb. *Proceedings of the National Academy of Sciences*, 110(45):18202–18207.

[4] Alexander, D. H. and Lange, K. (2011). Enhancements to the admixture algorithm for individual ancestry estimation. *BMC bioinformatics*, 12(1):246.

[5] Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9):1655–1664.

[6] Allen, P. H. (1954). Pollination in gongora maculata. *Ceiba*, 4(2):121–125.

[7] Alloui, T., Boussebough, I., Chaoui, A., Nouar, A. Z., and Chettah, M. C. (2015). Usearch: A meta search engine based on a new result merging strategy. In *2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)*, volume 1, pages 531–536. IEEE.

[8] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410.

[9] Amrad, A., Moser, M., Mandel, T., de Vries, M., Schuurink, R. C., Freitas, L., and Kuhlemeier, C. (2016). Gain and loss of floral scent production through changes in structural genes during pollinator-mediated speciation. *Current Biology*, 26(24):3303–3312.

[10] Armbruster, W. S. (2014). Floral specialization and angiosperm diversity: phenotypic divergence, fitness trade-offs and realized pollination accuracy. *AoB Plants*, 6.

[11] Atwood, J. T. (1987). The vascular flora of la selva biological station, costa rica orchidaceae. *Selbyana*, 10(1):76–145.

[12] Baack, E., Melo, M. C., Rieseberg, L. H., and Ortiz-Barrientos, D. (2015). The origins of reproductive isolation in plants. *New Phytologist*, 207(4):968–984.

[13] Beals, E. W. (1984). Bray-curtis ordination: an effective strategy for analysis of multivariate ecological data. In *Advances in ecological research*, volume 14, pages 1–55. Elsevier.

[14] Benitez-Vieyra, S., Pérez-Alquicira, J., Sazatornil, F. D., Domínguez, C. A., Boege, K., Pérez-Ishiwara, R., and Fornoni, J. (2019). Evolutionary transition between bee pollination and hummingbird pollination in salvia: Comparing means, variances and covariances of corolla traits. *Journal of evolutionary biology*, 32(8):783–793.

[15] Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using sspace. *Bioinformatics*, 27(4):578–579.

[16] Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). Tassel: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19):2633–2635.

[17] Bradshaw, H. and Schemske, D. W. (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, 426(6963):176–178.

[18] Brandt, K., Dötterl, S., Ramírez, S. R., Etl, F., Machado, I. C., Navarro, D. M. d. A. F., Dobler, D., Reiser, O., Ayasse, M., and Milet-Pinheiro, P. (2021). Unraveling the olfactory biases of male euglossine bees: species-specific antennal responses and their evolutionary significance for perfume flowers. *Frontiers in Ecology and Evolution*, page 690.

[19] Broyles, S. B., Vail, C., and Sherman-Broyles, S. L. (1996). Pollination genetics of hybridization in sympatric populations of asclepias exaltata and a. syriaca (asclepiadaceae). *American Journal of Botany*, 83(12):1580–1584.

[20] Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W.-C., Liu, K.-W., Chen, L.-J., He, Y., Xu, Q., Bian, C., et al. (2015). The genome sequence of the orchid phalaenopsis equestris. *Nature genetics*, 47(1):65–72.

[21] Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S., and Yandell, M. (2008). Maker: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome research*, 18(1):188–196.

[22] Chao, Y.-T., Chen, W.-C., Chen, C.-Y., Ho, H.-Y., Yeh, C.-H., Kuo, Y.-T., Su, C.-L., Yen, S.-H., Hsueh, H.-Y., Yeh, J.-H., et al. (2018). Chromosome-level assembly, genetic and physical mapping of phalaenopsis aphrodite genome provides new insights into species adaptation and resources for orchid breeding. *Plant biotechnology journal*, 16(12):2027–2041.

[23] Chao, Y.-T., Yen, S.-H., Yeh, J.-H., Chen, W.-C., and Shih, M.-C. (2017). Orchidstra 2.0—a transcriptomics resource for the orchid family. *Plant and Cell Physiology*, 58(1):e9–e9.

[24] Chen, F., Tholl, D., Bohlmann, J., and Pichersky, E. (2011). The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant Journal*, 66(1):212–229.

[25] Chen, N. (2004). Using repeat masker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics*, 5(1):4–10.

[26] Christie, K. and Strauss, S. Y. (2019). Reproductive isolation and the maintenance of species boundaries in two serpentine endemic jewelflowers. *Evolution*, 73(7):1375–1391.

[27] Chung, O., Kim, J., Bolser, D., Kim, H.-M., Jun, J. H., Choi, J.-P., Jang, H.-D., Cho, Y. S., Bhak, J., and Kwak, M. (2022). A chromosome-scale genome assembly and annotation of the spring orchid (cymbidium goeringii). *Molecular Ecology Resources*, 22(3):1168–1177.

[28] Cowley, M. and Oakey, R. J. (2013). Transposable elements re-wire and fine-tune the transcriptome. *PLoS genetics*, 9(1):e1003234.

[29] Coyne, J. A., Orr, H. A., et al. (2004). *Speciation*, volume 37. Sinauer Associates Sunderland, MA.

[30] Cozzolino, S. and Widmer, A. (2005). Orchid diversity: an evolutionary consequence of deception? *Trends in Ecology & Evolution*, 20(9):487–494.

[31] Cruickshank, T. E. and Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular ecology*, 23(13):3133–3157.

[32] Cutter, A. D. and Payseur, B. A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics*, 14(4):262–274.

[33] Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., et al. (2011). The variant call format and vcftools. *Bioinformatics*, 27(15):2156–2158.

[34] Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., et al. (2021). Twelve years of samtools and bcftools. *Gigascience*, 10(2):giab008.

[35] Dodson, C. H. (2003). Why are there so many orchid species? *Lankesteriana*.

[36] Dodson, C. H., Dressler, R. L., Hills, H. G., Adams, R. M., and Williams, N. H. (1969). Biologically active compounds in orchid fragrances. *Science*, 164(3885):1243–1249.

[37] Dormann, C. F., Gruber, B., and Fründ, J. (2008). Introducing the bipartite package: analysing ecological networks. *interaction*, 1(0.2413793).

[38] Dressler, R. L. (1966). Some observations on gongora. *Orchid Digest*, 30:220–223.

[39] Dressler, R. L. (1968). Observations on orchids and euglossine bees in panama and costa rica. *Revista de Biologia Tropical*, 15(1):143–183.

[40] Dressler, R. L. (1990). *The orchids: natural history and classification*. Harvard University Press.

[41] Dressler, R. L. (2005). How many orchid species? *Selbyana*, pages 155–158.

[42] Edgar, R. C. (2004). Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5):1792–1797.

[43] Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species. *PloS one*, 6(5):e19379.

[44] Eltz, T., Sager, A., and Lunau, K. (2005). Juggling with volatiles: exposure of perfumes by displaying male orchid bees. *Journal of Comparative Physiology A*, 191(7):575–581.

[45] FastQC (2016). Fastqc: a quality control tool for high throughput sequence data. *Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.*

[46] Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., and Smit, A. F. (2020). Repeatmodeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, 117(17):9451–9457.

[47] Fumagalli, M., Vieira, F. G., Linderoth, T., and Nielsen, R. (2014). ngstools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*, 30(10):1486–1487.

[48] Gerlach, G. and Schill, R. (1991). Composition of orchid scents attracting euglossine bees. *Botanica Acta*, 104(5):379–384.

[49] Gervasi, D. D. and Schiestl, F. P. (2017). Real-time divergent evolution in plants driven by pollinators. *Nature Communications*, 8(1):1–8.

[50] Goslee, S. C. and Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, 22(1):1–19.

[51] Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical f-statistics. *Molecular Ecology Notes*, 5(1):184–186.

[52] Grant, V. (1949). Pollination systems as isolating mechanisms in angiosperms. *Evolution*, pages 82–97.

[53] Grant, V. and Grant, K. A. (1965). *Flower pollination in the Phlox family.* Columbia University Press.

[54] Greiner, S., Lehwark, P., and Bock, R. (2019). Organellargenomedraw (ogdraw) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic acids research*, 47(W1):W59–W64.

[55] Gross, K., Sun, M., and Schiestl, F. P. (2016). Why do floral perfumes become different? region-specific selection on floral scent in a terrestrial orchid. *PloS one*, 11(2):e0147975.

[56] Gruber, B., Unmack, P. J., Berry, O. F., and Georges, A. (2018). dartr: An r package to facilitate analysis of snp data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, 18(3):691–699.

[57] Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). Quast: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8):1072–1075.

[58] Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from rna-seq using the trinity platform for reference generation and analysis. *Nature protocols*, 8(8):1494–1512.

[59] Harder, L. D. and Johnson, S. D. (2009). Darwin's beautiful contrivances: evolutionary and functional evidence for floral adaptation. *New Phytologist*, 183(3):530–545.

[60] Hetherington-Rauth, M. C. and Ramírez, S. R. (2015). Evolutionary trends and specialization in the euglossine bee-pollinated orchid genus gongora. *Annals of the Missouri Botanical Garden*, 100(4):271–299.

[61] Hetherington-Rauth, M. C. and Ramírez, S. R. (2016). Evolution and diversity of floral scent chemistry in the euglossine bee-pollinated orchid genus gongora. *Annals of Botany*, 118(1):135–148.

[62] Howe, K. L., Contreras-Moreira, B., De Silva, N., Maslen, G., Akanni, W., Allen, J., Alvarez-Jarreta, J., Barba, M., Bolser, D. M., Cambell, L., et al. (2020). Ensembl genomes 2020—enabling non-vertebrate genomic research. *Nucleic acids research*, 48(D1):D689–D695.

[63] Huang, H., Kuo, Y.-W., Chuang, Y.-C., Yang, Y.-P., Huang, L.-M., Jeng, M.-F., Chen, W.-H., and Chen, H.-H. (2021a). Terpene synthase-b and terpene synthase-e/f genes produce monoterpenes for phalaenopsis bellina floral scent. *Frontiers in Plant Science*, page 1422.

[64] Huang, L.-M., Huang, H., Chuang, Y.-C., Chen, W.-H., Wang, C.-N., and Chen, H.-H. (2021b). Evolution of terpene synthases in orchidaceae. *International journal of molecular sciences*, 22(13):6947.

[65] Irwin, D. E., Alcaide, M., Delmore, K. E., Irwin, J. H., and Owens, G. L. (2016). Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Molecular Ecology*, 25(18):4488–4507.

[66] Jenny, R. (1993). *Monograph of the genus Gongora Ruiz & Pavon*. Koeltz Scientific Books.

[67] Johnson, S. D. (2010). The pollination niche and its role in the diversification and maintenance of the southern african flora. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1539):499–516.

[68] Johnson, S. D., Harder, L., and Barrett, S. (2006). *Pollinator-driven speciation in plants*. Ecology and evolution of flowers. Oxford: Oxford University Press.

[69] Jombart, T. (2008). adegenet: a r package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11):1403–1405.

[70] Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research*, 110(1-4):462–467.

[71] Kay, K. M. and Sargent, R. D. (2009). The role of animal pollination in plant speciation: integrating ecology, geography, and genetics. *Annual Review of Ecology, Evolution, and Systematics*, 40:637–656.

[72] Kay, K. M., Woolhouse, S., Smith, B. A., Pope, N. S., and Rajakaruna, N. (2018). Sympatric serpentine endemic monardella (lamiaceae) species maintain habitat differences despite hybridization. *Molecular ecology*, 27(9):2302–2316.

[73] Kephart, S. and Theiss, K. (2004). Pollinator-mediated isolation in sympatric milkweeds (asclepias): do floral morphology and insect behavior influence species boundaries? *New Phytologist*, 161(1):265–277.

[74] Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with hisat2 and hisat-genotype. *Nature biotechnology*, 37(8):907–915.

[75] Klahre, U., Gurba, A., Hermann, K., Saxenhofer, M., Bossolini, E., Guerin, P., and Kuhlemeier, C. (2011). Pollinator choice in petunia depends on two major genetic loci for floral scent production. *Current Biology*, 21(9):730–739.

[76] Knudsen, J. T., Eriksson, R., Gershenzon, J., and Ståhl, B. (2006). Diversity and distribution of floral scent. *The botanical review*, 72(1):1.

[77] Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research*, 27(5):722–736.

[78] Korneliussen, T. S., Albrechtsen, A., and Nielsen, R. (2014). Angsd: analysis of next generation sequencing data. *BMC bioinformatics*, 15(1):1–13.

[79] Lagomarsino, L. P., Forrestel, E. J., Muchhala, N., and Davis, C. C. (2017). Repeated evolution of vertebrate pollination syndromes in a recently diverged andean plant clade. *Evolution*, 71(8):1970–1985.

[80] Laliberté, E. and Legendre, P. (2010). A distance-based framework for measuring functional diversity from multiple traits. *Ecology*, 91(1):299–305.

[81] Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with bwa-mem. *arXiv preprint arXiv:1303.3997*.

[82] Li, H. (2014). Fast construction of fm-index for long sequence reads. *Bioinformatics*, 30(22):3274–3275.

[83] Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100.

[84] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079.

[85] Lowry, D. B., Modliszewski, J. L., Wright, K. M., Wu, C. A., and Willis, J. H. (2008). The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1506):3009–3021.

[86] Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., Buckler, E. S., and Costich, D. E. (2013). Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based snp discovery protocol. *PLoS genetics*, 9(1):e1003215.

[87] Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., Feng, J., Jiang, D., Ahani, H., Abbott, R. J., et al. (2018). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proceedings of the National Academy of Sciences*, 115(2):E236–E243.

[88] Marcais, G. and Kingsford, C. (2012). Jellyfish: A fast k-mer counter. *Tutorialis e Manuais*, 1:1–8.

[89] Michener, C. D. (2000). *The bees of the world*, volume 1. JHU press.

[90] Milet-Pinheiro, P., Silva, J. B. F., Navarro, D. M., Machado, I. C., and Gerlach, G. (2018). Notes on pollination ecology and floral scent chemistry of the rare neotropical orchid catasetum galeritum r chb. f. *Plant Species Biology*, 33(2):158–163.

[91] Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., Tosatto, S. C., Paladin, L., Raj, S., Richardson, L. J., et al. (2021). Pfam: The protein families database in 2021. *Nucleic acids research*, 49(D1):D412–D419.

[92] Moeller, D. A. (2004). Facilitative interactions among plants via shared pollinators. *Ecology*, 85(12):3289–3301.

[93] Morjan, C. L. and Rieseberg, L. H. (2004). How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular ecology*, 13(6):1341–1356.

[94] Noor, M. A. and Bennett, S. M. (2009). Islands of speciation or mirages in the desert? examining the role of restricted recombination in maintaining species. *Heredity*, 103(6):439–444.

[95] Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P., O'hara, R., Simpson, G., Solymos, P., Stevens, M., Wagner, H., et al. (2013). Community ecology package. *R package version*, pages 2–0.

[96] Ollerton, J., Winfree, R., and Tarrant, S. (2011). How many flowering plants are pollinated by animals? *Oikos*, 120(3):321–326.

[97] Orr, H. A. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, 52(4):935–949.

[98] Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R., Hellinga, A. J., Lugo, C. S. B., Elliott, T. A., Ware, D., Peterson, T., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome biology*, 20(1):1–18.

[99] Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. *Genetics*, 192(3):1065–1093.

[100] Petr, M., Vernot, B., and Kelso, J. (2019). admixr—r package for reproducible analyses using admixtools. *Bioinformatics*, 35(17):3194–3195.

[101] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J., et al. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 81(3):559–575.

[102] Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., Troll, C. J., Fields, A., Hartley, P. D., Sugnet, C. W., et al. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome research*, 26(3):342–350.

[103] Quinlan, A. R. and Hall, I. M. (2010). Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.

[104] Raguso, R. A. (2008). Wake up and smell the roses: the ecology and evolution of floral scent. *Annual Review of Ecology, Evolution, and Systematics*, 39:549–569.

[105] Ramirez, S. R., Eltz, T., Fujiwara, M. K., Gerlach, G., Goldman-Huertas, B., Tsutsui, N. D., and Pierce, N. E. (2011). Asynchronous diversification in a specialized plant-pollinator mutualism. *Science*, 333(6050):1742–1746.

[106] Ramírez, S. R., Gravendeel, B., Singer, R. B., Marshall, C. R., and Pierce, N. E. (2007). Dating the origin of the orchidaceae from a fossil orchid with its pollinator. *Nature*, 448(7157):1042–1045.

[107] Ravinet, M., Faria, R., Butlin, R., Galindo, J., Bierne, N., Rafajlović, M., Noor, M., Mehlig, B., and Westram, A. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of evolutionary biology*, 30(8):1450–1477.

[108] Rieseberg, L. H., Wendel, J. F., et al. (1993). Introgression and its consequences in plants. *Hybrid zones and the evolutionary process*, 70:109.

[109] Rieseberg, L. H. and Willis, J. H. (2007). Plant speciation. *science*, 317(5840):910–914.

[110] Riffell, J. A., Lei, H., Abrell, L., and Hildebrand, J. G. (2013). Neural basis of a pollinator's buffet: olfactory specialization and learning in manduca sexta. *Science*, 339(6116):200–204.

[111] Rockman, M. V. (2012). The qtn program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution: International Journal of Organic Evolution*, 66(1):1–17.

[112] Ruan, J. and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nature methods*, 17(2):155–158.

[113] Sandring, S., RiihimäKi, M.-A., Savolainen, O., and Ågren, J. (2007). Selection on flowering time and floral display in an alpine and a lowland population of arabidopsis lyrata. *Journal of evolutionary biology*, 20(2):558–567.

[114] Sargent, R. D. and Ackerly, D. D. (2008). Plant–pollinator interactions and the assembly of plant communities. *Trends in Ecology & Evolution*, 23(3):123–130.

[115] Schiestl, F. P. and Schlüter, P. M. (2009). Floral isolation, specialized pollination, and pollinator behavior in orchids. *Annual review of entomology*, 54:425–446.

[116] Schoener, T. W. (1970). Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology*, 51(3):408–418.

[117] Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). Busco: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19):3210–3212.

[118] Srividya, N., Davis, E. M., Croteau, R. B., and Lange, B. M. (2015). Functional analysis of (4s)-limonene synthase mutants reveals determinants of catalytic outcome in a model monoterpene synthase. *Proceedings of the National Academy of Sciences*, 112(11):3332–3337.

[119] Stamatakis, A. (2014). Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9):1312–1313.

[120] Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). Augustus: ab initio prediction of alternative transcripts. *Nucleic acids research*, 34(suppl_2):W435–W439.

[121] Stebbins, G. L. (1970). Adaptive radiation of reproductive characteristics in angiosperms, i: pollination mechanisms. *Annual Review of Ecology and Systematics*, 1(1):307–326.

[122] Strauss, S. Y. and Whittall, J. B. (2006). Non-pollinator agents of selection on floral traits. *Ecology and evolution of flowers*, pages 120–138.

[123] Tamura, K., Stecher, G., and Kumar, S. (2021). Mega11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*, 38(7):3022–3027.

[124] Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., and Greiner, S. (2017). Geseq–versatile and accurate annotation of organelle genomes. *Nucleic acids research*, 45(W1):W6–W11.

[125] Tremblay, R. L. and Ackerman, J. D. (2001). Gene flow and effective population size in lepanthes (orchidaceae): a case for genetic drift. *Biological Journal of the Linnean Society*, 72(1):47–62.

[126] Van Der Cingel, N. A. (2001). *An atlas of orchid pollination: European orchids*. CRC Press.

[127] van der Niet, T. and Johnson, S. D. (2012). Phylogenetic evidence for pollinator-driven diversification of angiosperms. *Trends in ecology & evolution*, 27(6):353–361.

[128] Van der Niet, T., Peakall, R., and Johnson, S. D. (2014). Pollinator-driven ecological speciation in plants: new evidence and future perspectives. *Annals of Botany*, 113(2):199–212.

[129] Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., and Wolf, J. B. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature communications*, 7(1):1–10.

[130] Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one*, 9(11):e112963.

[131] Wang, J. R., Holt, J., McMillan, L., and Jones, C. D. (2018). Fmlrc: Hybrid long read error correction using an fm-index. *BMC bioinformatics*, 19(1):1–11.

[132] Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., et al. (2007). Database resources of the national center for biotechnology information. *Nucleic acids research*, 36(suppl_1):D13–D21.

[133] Whitten, W. M. (1985). *Variation in floral fragrances and pollinators in the Gongora quinguenervis complex (Orchidaceae) in Central Panama*. PhD thesis, University of Florida.

[134] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. *Journal of open source software*, 4(43):1686.

[135] Wickham, H., Chang, W., and Wickham, M. H. (2016). Package 'ggplot2'. *Create elegant data visualisations using the grammar of graphics. Version*, 2(1):1–189.

[136] Williams, N. H. and Dodson, C. H. (1972). Selective attraction of male euglossine bees to orchid floral fragrances and its importance in long distance pollen flow. *Evolution*, pages 84–95.

[137] Williams, N. H. and Whitten, W. M. (1983). Orchid floral fragrances and male euglossine bees: methods and advances in the last sesquidecade. *The Biological Bulletin*, 164(3):355–395.

[138] Xu, Q., Niu, S., Li, K., Zheng, P., Zhang, X., Jia, Y., Liu, Y., Niu, Y., Yu, L., Chen, D., et al. (2022). Chromosome-scale assembly of the dendrobium nobile genome provides insights into the molecular mechanism of the biosynthesis of the medicinal active ingredient of dendrobium. *Frontiers in genetics*, 13.

[139] Xu, S., Schlüter, P. M., and Schiestl, F. P. (2012). Pollinator-driven speciation in sexually deceptive orchids. *International Journal of Ecology*, 2012.

[140] Yan, L., Wang, X., Liu, H., Tian, Y., Lian, J., Yang, R., Hao, S., Wang, X., Yang, S., Li, Q., et al. (2015). The genome of dendrobium officinale illuminates the biology of the important traditional chinese orchid herb. *Molecular plant*, 8(6):922–934.

[141] Yang, C.-F., Gituru, R. W., and Guo, Y.-H. (2007). Reproductive isolation of two sympatric louseworts, pedicularis rhinanthoides and pedicularis longiflora (orobanchaceae): how does the same pollinator type avoid interspecific pollen transfer? *Biological Journal of the Linnean Society*, 90(1):37–48.

[142] Yang, F.-X., Gao, J., Wei, Y.-L., Ren, R., Zhang, G.-Q., Lu, C.-Q., Jin, J.-P., Ai, Y., Wang, Y.-Q., Chen, L.-J., et al. (2021). The genome of cymbidium sinense revealed the evolution of orchid traits. *Plant biotechnology journal*, 19(12):2501.

[143] Yu, Z., Zhao, C., Zhang, G., Teixeira da Silva, J. A., and Duan, J. (2020). Genome-wide identification and expression profile of tps gene family in dendrobium officinale and the role of dotps10 in linalool biosynthesis. *International journal of molecular sciences*, 21(15):5419.

[144] Yuan, Y., Jin, X., Liu, J., Zhao, X., Zhou, J., Wang, X., Wang, D., Lai, C., Xu, W., Huang, J., et al. (2018). The gastrodia elata genome provides insights into plant adaptation to heterotrophy. *Nature Communications*, 9(1):1–11.

[145] Zhang, G.-Q., Liu, K.-W., Li, Z., Lohaus, R., Hsiao, Y.-Y., Niu, S.-C., Wang, J.-Y., Lin, Y.-C., Xu, Q., Chen, L.-J., et al. (2017). The apostasia genome and the evolution of orchids. *Nature*, 549(7672):379–383.

[146] Zhang, G.-Q., Xu, Q., Bian, C., Tsai, W.-C., Yeh, C.-M., Liu, K.-W., Yoshida, K., Zhang, L.-S., Chang, S.-B., Chen, F., et al. (2016). The dendrobium catenatum lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Scientific reports*, 6(1):1–10.

[147] Zhang, H., Jain, C., and Aluru, S. (2020). A comprehensive evaluation of long read error correction methods. *BMC genomics*, 21(6):1–15.

[148] Zhang, W., Zhang, G., Zeng, P., Zhang, Y., Hu, H., Liu, Z., and Cai, J. (2021a). Genome sequence of apostasia ramifera provides insights into the adaptive evolution in orchids. *BMC genomics*, 22(1):1–12.

[149] Zhang, Y., Zhang, G.-Q., Zhang, D., Liu, X.-D., Xu, X.-Y., Sun, W.-H., Yu, X., Zhu, X., Wang, Z.-W., Zhao, X., et al. (2021b). Chromosome-scale assembly of the dendrobium chrysotoxum genome enhances the understanding of orchid evolution. *Horticulture Research*, 8.

[150] Zhou, X., Liu, Q., Han, J. Y., and Gao, J. (2016). Different pollinator assemblages ensure reproductive success of cleisostoma linearilobatum (orchidaceae) in fragmented holy hill forest and traditional tea garden. *Scientific Reports*, 6(1):1–7.

[151] Zimmermann, Y., Ramírez, S. R., and Eltz, T. (2009). Chemical niche differentiation among sympatric species of orchid bees. *Ecology*, 90(11):2994–3008.