The convergent evolution of hummingbird pollination results in repeated floral scent loss through gene downregulation

Kathy Darragh[1,2,+], Kathleen M. Kay[3], Santiago R. Ramírez[1]

[1]Department of Evolution and Ecology, University of California, Davis, CA, USA

[2]Department of Biology, Indiana University, Bloomington, IN, USA

[3]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA, USA

[+]Corresponding author (Email: kdarrag@iu.edu)

ORCID Kathy Darragh 0000-0002-9215-014X

ORCID Kathleen M. Kay 0000-0001-8858-110X

ORCID Santiago R. Ramírez 0000-0003-1306-1315

Keywords

Convergent evolution, floral scent, gene family evolution, hummingbird, orchid bee, pollination, spiral gingers, transcriptomics, trait evolution

Abstract

The repeated evolution of the same trait in different lineages provides powerful natural experiments to study the phenotypic and genotypic predictability of how traits are gained and lost. A fascinating example of this is the repeated evolution of hummingbird pollination in plant lineages in the Americas, a widespread and often unidirectional phenomenon. The spiral gingers in the genus *Costus* are ancestrally bee-pollinated, and hummingbird pollination has evolved multiple times independently in the tropical Americas. These pollinator transitions are accompanied by predictable morphological and color changes, but the changes in floral scent have not been described. In this study, we describe the floral scent composition of 30 species of *Costus* sampled across the phylogeny to understand how floral scent has evolved across the genus with respect to pollinator transitions. We then combine transcriptomics and genomics to identify genetic expression differences and gene family evolution associated with pollinator transitions. We show that hummingbird-pollinated species have mostly lost their floral scent, whereas bee-pollinated species exhibit either floral scent maintenance or in some cases, gains of more diverse scent profiles. We find the floral scent loss appears to be due to gene downregulation rather than pseudogenization. The remarkable consistency of scent loss in hummingbird-pollinated species highlights the shared strong selection pressures experienced by these lineages. Even species with more recent transitions from bee to hummingbird pollination exhibit scent loss, highlighting the rapid breakdown of scent production following pollinator transitions. This research highlights the capacity for rapid changes when selection pressures are strong through downregulation of floral scent genes.

Resumen

La evolución independiente de un rasgo en diferentes linajes ofrece la oportunidad de estudiar la predictibilidad de fenotipos y sus bases genéticas. Un ejemplo fascinante es la evolución repetida de polinización por colibríes en varios linajes de plantas en las Américas, un fenómeno generalizado y típicamente unidireccional. Las especies de caña agria del género *Costus* son polinizadas ancestralmente por abejas y la polinización por colibríes ha evolucionado varias veces de forma independiente en América tropical. Cada transición de polinizador ha sido acompañada de varios cambios morfológicos y colores predecibles, pero los cambios en el aroma floral no han sido descritos. En este estudio describimos la composición del aroma floral de 30 especies de *Costus* para comprender cómo el aroma floral ha evolucionado con respecto a las transiciones de polinizadores. Combinamos transcriptómica y genómica en un clado de *Costus* para identificar diferencias en la expresión genética y la evolución de la familia de genes asociadas con las transiciones de polinizadores. Nuestros resultados muestran que las especies polinizadas por colibríes han perdido en su mayoría su aroma floral, mientras que las especies polinizadas por abejas ya sea mantienen el aroma floral y, en algunos casos, muestran ganancias de perfiles de aroma más diversos. La consistencia en la pérdida de olor en especies polinizadas por colibríes resalta las fuertes presiones de selección natural que experimentan estos linajes. Incluso las especies con transiciones más recientes de polinización de abejas a colibríes exhiben pérdida de olor, lo que destaca la rápida interrupción de la producción de olores después de las transiciones de los polinizadores. También encontramos que la pérdida de aroma floral parece estar controlada por regulación génica negativa y no por la pérdida de genes de producción de aromas (*e.g.* pseudogenización). Esta investigación destaca la rapidez de cambios a través de la

regulación negativa de genes que controlan aromas florales cuando las presiones de selección son fuertes.

Significance Statement

The independent evolution of the same trait in different lineages provides a natural experiment to investigate the predictability of evolution. A fascinating example is the evolution of hummingbird pollination, which has occurred many times across flowering plants, and often involves convergence in multiple traits. In this study, we combine comparative chemical ecology with genomics and transcriptomics to study floral scent evolution in the genus *Costus*. We show that species exhibit loss of floral scent following transitions to hummingbird pollination suggesting a shared strong selection pressure favoring scent loss. We demonstrate the floral scent loss appears to be due to gene downregulation. This research highlights the capacity for rapid changes when selection pressures are strong through downregulation of scent genes.

Introduction

The repeated evolution of the same traits in different lineages is thought to reflect shared selection pressures. These cases of repeated evolution apply not only to the gain of traits, such as the evolution of antifreeze proteins in fish (Chen et al. 1997) or carnivory in plants (Albert et al. 1992), but also the loss of traits, as seen in the eye regression of cavefish (Sifuentes-Romero et al. 2023), or reduction of attractive traits such as flower size or floral scent in self-fertilizing plants (Tsuchimatsu and Fujii 2022). Repeated evolution provides us with systems where we can study the process of adaptation and ask questions about the predictability of evolution, and the extent of constraint, or flexibility, of the underlying molecular mechanisms (Orr 2005; Losos 2011; Sackton and Clark 2019).

One such example of repeated evolution is pollinator transitions in flowering plants. For example, the evolution of hummingbird pollination has occurred many times across flowering plants, and often involves convergence in a suite of traits. Flowers pollinated by hummingbirds are characterized by traits thought to attract and fit hummingbirds (such as a narrow shape and large amounts of nectar) and in addition, traits thought to deter bees (red coloration, lack of nectar guides, lack of scent) (Schemske and Bradshaw 1999; Castellanos et al. 2004; Bergamo et al. 2016). Interestingly, the evolution of hummingbird pollination has primarily been documented in one direction, from ancestral bee pollination to derived hummingbird pollination (Rosas-Guerrero et al. 2014; Abrahamczyk and Renner 2015; Kay and Grossenbacher 2022). However, there is clade-specific variation in this directionality, with the reasons for pollinator transitions remaining unclear (Tripp and Manos 2008; Rosas-Guerrero et al. 2014; Abrahamczyk and Renner 2015; Kay and Grossenbacher 2022; Barreto et al. 2024).

Hummingbird-pollinated flowers often lack floral scent and therefore provide an excellent system to study repeated trait loss (Knudsen et al. 2004; Wessinger 2024). Genomic studies have shown that, even for birds, hummingbirds have a low number of olfactory receptors (Driver and Balakrishnan 2021; Driver 2022), suggesting at least a reduced investment in this sensory modality. Behavioral experiments have also suggested that hummingbirds rely more on visual than olfactory cues when foraging (Goldsmith and Goldsmith 1982). In fact, floral scent analyses of hummingbird pollinated flowers have shown either complete lack of floral scent or low emission rates of floral scent compounds. (Knudsen et al. 2004; Byers et al. 2014; Amrad et

al. 2016). The lack of floral scent in hummingbird-pollinated species suggested that there is a cost to producing scent for these species. This could be due to energetic costs of compound production, but this is predicted to be minimal (Raguso 2016; Pichersky and Raguso 2018). An alternative is that there is an ecological cost for hummingbird-pollinated species attracting bees, also called the bee avoidance hypothesis (Castellanos et al. 2004; Bergamo et al. 2016).

A fascinating group in which to study the evolution of hummingbird pollination and potential trait loss is the genus *Costus* (Costaceae). The genus contains over 80 species of perennial understory herbaceous monocots, the majority of which have diverged rapidly within the last 3 million years in Central and South America (Maas 1972; Maas 1977; Vargas et al. 2020). Since establishing in the American tropics, hummingbird pollination has evolved repeatedly and independently in *Costus* from an ancestral state of bee pollination (Vargas et al. 2020). Specifically, bee-pollinated species are pollinated primarily by orchid bees (Apidae: Euglossini), with females visiting flowers for nectar and pollen, and males visiting for nectar (Kay and Schemske 2003). Hummingbird-pollinated species are primarily pollinated by non-territorial hermit hummingbirds (Kay and Schemske 2003). While different species were initially classified as hummingbird or bee pollinated (Maas 1972; Maas 1977), many species do not fit the classic paradigm of long tubular red flowers associated with hummingbird pollination (Grant and Grant 1968). For example, many bee-pollinated flowers have some red coloration (Kay and Schemske 2003), and hummingbird-pollinated flowers are short (Yost and Kay 2009). Despite this variation, a single phenotypic optimum is shared across most of the independent transitions to hummingbird pollination in the clade (Kay and Grossenbacher 2022). Traits including small flower size, the absence of nectar guides, and brightly colored floral bracts are all predictive of hummingbird pollination (Kay and Grossenbacher 2022). Furthermore, *Costus* species show extensive sympatry between different pollination syndromes (Kay and Schemske 2003; Vargas et al. 2020; Kay and Grossenbacher 2022) and even the specialist herbivores are shared across *Costus* species (García-Robledo et al. 2013), allowing us to isolate the effects of pollination syndrome from geographic location or herbivory. The combination of recent genomic resources, a well-resolved phylogeny, documented pollinators, and the possibility of collecting samples both in the field and in greenhouses, makes *Costus* a powerful system to study the evolution of convergence on hummingbird pollination (Vargas et al. 2020; Kay and Grossenbacher 2022; Harenčár et al. 2023).

While the morphological and color changes involved in the evolution of hummingbird pollination have been well-studied in *Costus*, the expected scent changes have not been characterized. In this study, we describe floral scent in 30 species of *Costus* sampled across the phylogeny and including eight independent origins of hummingbird pollination. We test for differences in floral scent between bee-pollinated and hummingbird-pollinated species and investigate how floral scent has evolved across the genus. We then focus on one clade of *Costus* with variation in floral scent production and generate floral transcriptomic data. We use this data to identify gene expression differences associated with floral scent production to better understand the genetic changes underlying floral scent evolution in this group. Finally, we annotated a family of genes important for the production of terpenoid floral scent compounds (the terpene synthase genes (TPS)) in three previously available *Costus* genomes, and a high-quality genome assembly for *Costus allenii* generated in this study, encompassing two independent transitions to hummingbird pollination. High-quality annotations allow us to analyze the evidence of gene loss or pseudogenization following transition to hummingbird pollination in this gene family that plays an important role in floral scent evolution. Our study, therefore, combines multiple approaches to characterize repeated phenotypic evolution and the molecular mechanisms of these changes with shifts to hummingbird pollination in *Costus*.

Results

<u>Bee-pollinated species produce more compounds and more diverse floral scents</u>

We collected 100 samples from 30 *Costus* species including species previously described as both bee- and hummingbird-pollinated (Fig 1A, Table S1). Using multivariate techniques, we visualized the floral scent profiles to identify the broad patterns in our dataset. Before filtering our data, we found that samples from hummingbird-pollinated species did not differ from ambient control samples (Fig 1B). In contrast, samples from bee-pollinated species spread across a wider section of chemospace, with many samples visually separated from both ambient and hummingbird-pollinated samples (Fig. 1B). We found that both pollination group, and species (nested within pollination group) explained variation in floral scent (PERMANOVA, pollination group: $F_{2,167}=8.0$, $p<0.001$; species: $F_{28,167}=2.5$, $p<0.001$). Interestingly, bee-pollinated species were significantly different from both hummingbird-pollinated species and ambient control

samples (Pairwise PERMANOVA, p=0.003, whereas hummingbird-pollinated species did not differ from ambient control samples (Pairwise PERMANOVA, p=0.07) (Figure S1).

After filtering likely contaminants from the ambient air, we obtained a final dataset with 60 floral scent compounds (Table S2). The majority of these compounds were terpenoid (87%, 52/60), with a few aromatics and compounds of unknown class. The dataset included both monoterpenes and sesquiterpenes. No one compound class was restricted to a specific pollinator group. After removing the ambient compounds, we found that one bee-pollinated sample and eighteen hummingbird-pollinated samples had no detectable floral scent, confirming the similarity of the hummingbird-pollinated samples to the ambient air samples. The unscented hummingbird-pollinated samples were from six species. For two species, *C. montanus* and *C. stenophyllus*, all individuals produced no detectable floral scent (4/4 individuals). Furthermore, *most C. chartaceus* (4/5), *C. lasius* (4/7), and *C. wilsonii* (5/8), and a single *C. productus* individual (1/5) lacked floral scent.

We found that bee-pollinated flowers produced more floral scent compounds, averaging 9 (s.d. 10) compounds in comparison to only 2 (s.d. 4) for hummingbird-pollinated flowers (Figure 1C). Bee-pollinated flowers contained more compounds than hummingbird-pollinated flowers (Fig 2A). We confirmed that this was the case for both plants raised in a greenhouse and those sampled in the field (Fig. S2). Similar to the number of compounds, we used a measure of biochemical and structural diversity of the compounds present in a sample to calculate the overall chemical diversity. The Functional Hill Diversity index considers not only the total number of compounds but also their biosynthetic complexity. We found that the bee-pollinated floral scents had a higher chemical diversity (Fig 2B). Bee-pollinated flowers also produced a higher abundance of floral scent, as measured by total ion abundance in each sample, a proxy for compound amount (Fig. 2C).

To see how individual compounds varied between bee- and hummingbird-pollinated species we calculated the percentage of individuals in each pollination group that produced the most common floral scent compounds in the dataset. We found that these compounds are present in a higher proportion of bee-pollinated flowers than hummingbird-pollinated flowers (Fig. 2D). In particular, (*E*)-beta-ocimene and beta-caryophyllene show the largest percentage differences

193 between the two pollination groups (Fig. 2D) and are the only two compounds to exhibit a
194 significant difference in frequency between pollinator groups (Table S2).

195 To facilitate visualization of our final filtered dataset using NMDS we removed the
196 nineteen samples with no detectable floral scent. When this filtered dataset was plotted, we
197 observed more overlap between the hummingbird- and bee-pollinated samples, but the two sets
198 of samples significantly different (Figure S1).

199

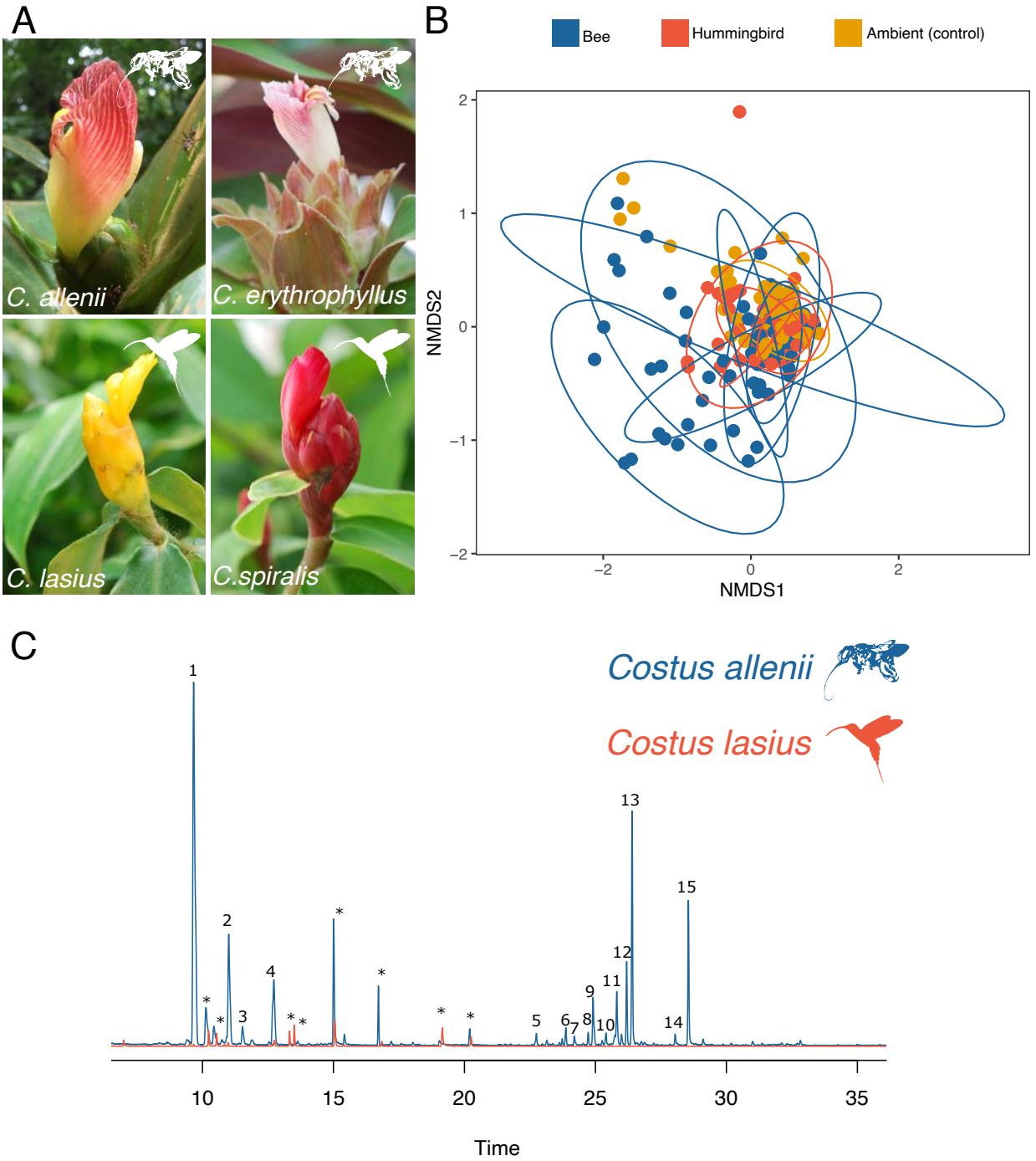


Figure 1. A) Representative images of two pairs of closely related bee- and hummingbird-pollinated species: *Costus allenii* and *C. lasius* and *C. erythrophyllus* and *C. spiralis*. B) NMDS (nonmetric multidimensional scaling) plot illustrating in the variation in the unfiltered dataset (Stress=0.12). Points are labelled as bee-pollinated, hummingbird-pollinated, or ambient samples. Ambient samples are control samples of the air taken in the

207 **same area as the focal plant individual. C) Total ion chromatogram of floral scent from**
208 ***Costus allenii* and *C. lasius*. We chose the individuals with the second-highest amount of**
209 **scent from each species to avoid plotting an outlier for either group. The *C. lasius* sample**
210 **was shifted on the x-axis by 1.5 minutes to match the *C. allenii* sample due to changes in the**
211 **GC/MS column over time. Abundance is scaled to the highest peak present in *C. allenii*. 1,**
212 **alpha-pinene; 2, beta-pinene; 3, beta-myrcene; 4, cineole; 5, alpha-copaene; 6, (*E*)-beta-**
213 **caryophyllene; 7, gamma-elemene; 8, alpha-humulene; 9, (*Z*)-beta-caryophyllene; 10,**
214 **germacrene D; 11, alpha-elemene; 12, gamma-cadinene; 13, delta-cadinene; 14,**
215 **germacrene-d-4-ol; 15, unidentified sesquiterpene (RI=1616). Contaminants are identified**
216 **with an asterisk (*).**

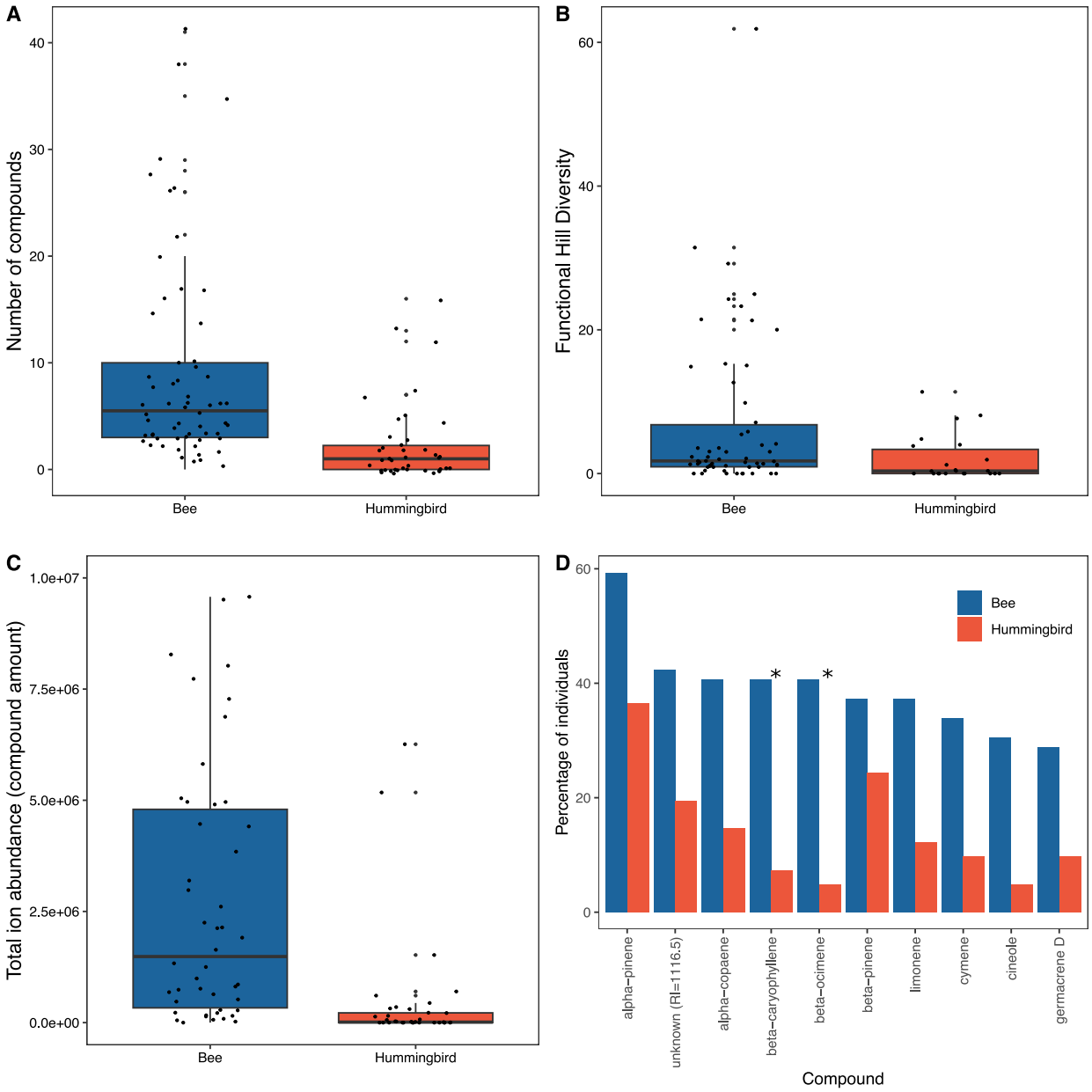


Figure 2. Differences in bee-pollinated and hummingbird-pollinated *Costus* species in terms of A) Number of scent compounds produced (ANOVA, pollination group: $F_{1,70}$=60.4, p<0.0001, species: $F_{28,70}$=9.4, p<0.0001), B) The diversity of scent compounds produced (ANOVA, pollination group: $F_{1,51}$=5.3, p=0.03, species: $F_{27,51}$=1.4, p=0.1), C) Total compound amount produced (measured by proxy of total ion abundance) (ANOVA, pollination group: $F_{1,70}$=109, p=<0.0001, species: $F_{28,70}$=2.8, p<0.001) and D) the percentage of individuals that produced the ten most common floral scent compounds in the dataset. Significant differences in frequency between bee- and hummingbird-pollinated flowers as

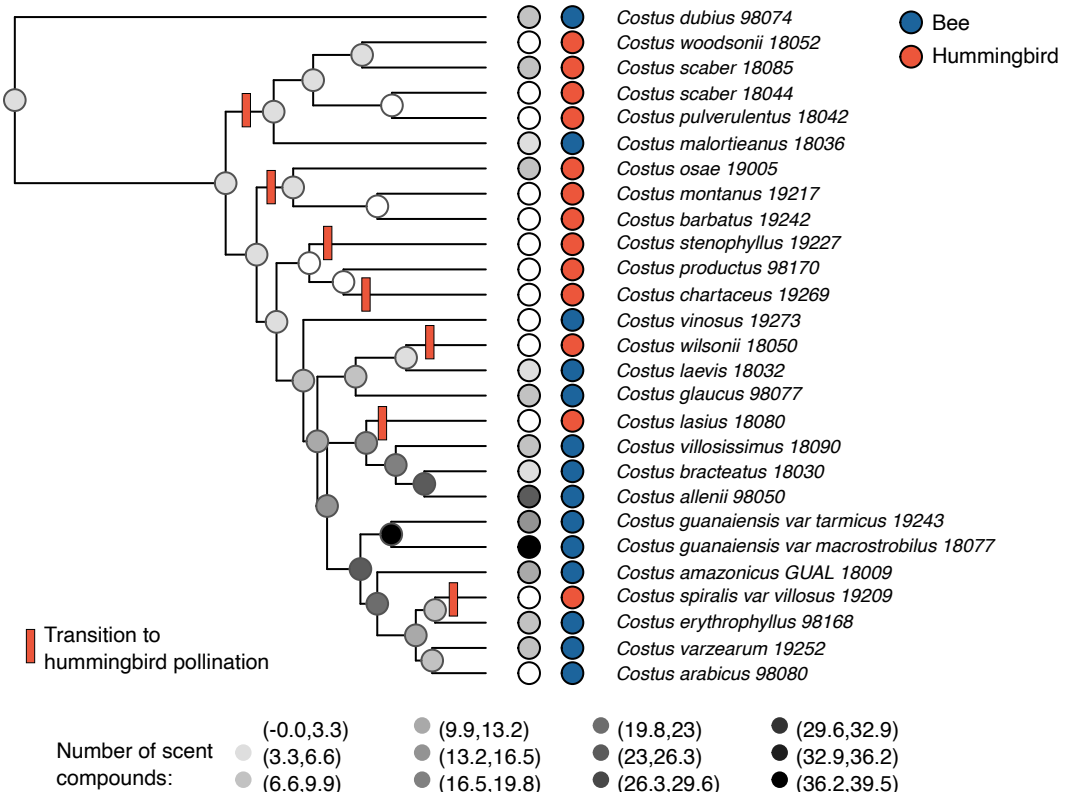
226 **determined by Chi-squared testing are highlighted with an asterisk. Fifteen individuals**
227 **with the highest compound production were removed from C for visualization.**

228 Phylogenetic analyses show that hummingbird-pollinated species have lost floral scent

229 To understand the evolution of floral scent in *Costus*, we carried out phylogenetic
230 analyses using a previously published phylogeny (Vargas et al. 2020). In our dataset, we found a
231 strong correlation between pollination group and the phylogenetic covariance matrix using two-
232 block partial least squares analysis (rPLS = 0.56; P=0.01). This reduces the statistical power of
233 testing group differences using phylogenetic ANOVA due to the high aggregation of pollinator
234 groups on the phylogeny (Adams and Collyer 2018). We tested for evidence of phylogenetic
235 signal in compound number, compound diversity, compound amount (total area of floral scent
236 peaks), (*E*)-beta-ocimene production, and (*E*)-beta-caryophyllene production, detecting weak and
237 non-significant phylogenetic signal overall (λ= 4.14795e-05, p=1). We also tested for evidence
238 of phylogenetic signal in the multivariate floral scent dataset, detecting non-significant almost
239 zero phylogenetic signal (Kmult=0.43, p=0.68). We therefore decided against using phylogenetic
240 ANOVA with simulations based on Brownian motion and instead used a method of randomizing
241 residuals in a permutation procedure (Adams and Collyer 2018). We found that compound
242 number, compound diversity, compound amount, and (*E*)-beta-ocimene production were all
243 significantly different between bee-pollinated and hummingbird-pollinated species (compound
244 number, $F_{1,26}$=7.2, p=0.008; compound diversity, $F_{1,26}$=5.8, p=0.01; compound amount,
245 $F_{1,26}$=627.5, p=0.008; (*E*)-beta-ocimene production, $F_{1,26}$=4.7, p=0.03). (*E*)-beta-caryophyllene
246 production did not differ between pollinator groups ($F_{1,26}$=1.0, p=0.4). Our comparative analyses
247 complement our previous analyses, again showing that bee-pollinated species produce higher
248 amounts of more diverse and complex floral scents and have higher (*E*)-beta-ocimene
249 production.

250 To further investigate how floral scent has evolved across the *Costus* phylogeny, we
251 carried out maximum likelihood ancestral state reconstruction of mean compound production.
252 We overlaid the eight independent origins of hummingbird pollination represented by our dataset
253 as determined in the full phylogenetic analysis of Vargas (2020) (Vargas et al. 2020). The
254 ancestral state of all *Costus* species found in tropical America is to produce more compounds
255 than average hummingbird-pollinated species (4.5 vs 2). In this reconstruction, the ancestral

256 *Costus* population produced an average floral scent – more compounds than hummingbird-
257 pollinated species but less than the average bee-pollinated species (Figure 3). Many
258 hummingbird-pollinated species have since completely lost floral scent, as shown by the white
259 circles (Figure 3). We see evidence for scent reduction or loss in all eight origins of
260 hummingbird pollination in our dataset. In contrast, some bee-pollinated species, such as *Costus*
261 *allenii*, have evolved much stronger floral scents than the ancestral state.



262

263
264
265 **Figure 3. Ancestral state reconstruction of floral scent in the genus *Costus*. Pollinator for**
266 **each species is illustrated by the right-hand dots. Mean compound number produced by**
267 **each species is shown by the left-hand dots with each shade of gray representing a range of**
268 **means. Darker dots indicate more compounds are produced. Predicted compound**
269 **production at each node is shown following the same color scale. Tip labels following**
270 **(Vargas et al. 2020). Transitions to hummingbird pollination were inferred from a larger**
271 **species-level phylogeny (Vargas et al. 2020).**
272

273 <u>Loss of floral scent is correlated with downregulation of floral terpene synthase genes.</u>

274 To analyze the genetic basis of floral scent evolution, we first chose a clade of closely
275 related species with variation in floral scent production and pollination systems. We chose the

276 clade containing *Costus lasius* due to the variation in floral scent production and the availability
277 of a high-quality genome. In this clade, *Costus lasius* (hummingbird-pollinated) produces on
278 average three floral scent compounds, whereas *Costus allenii* (bee-pollinated) produces 24 (one
279 of the highest in our dataset), and *Costus villosissimus* (bee-pollinated) produces seven
280 compounds (Figure 4A).

281 We first produced floral transcriptomes of all three species. After clustering transcripts
282 with over 95% similarity, we found 112998 transcripts present in *C. allenii*, 80208 in *C. lasius*,
283 and 96886 in *C. villosissimus*. Re-mapping rates were high in all three species: 99.2% in *C.*
284 *allenii*, 99% in *C. lasius*, and 99.1% in *C. villosissimus*. All three also had BUSCO
285 (Benchmarking Universal Single-Copy Orthologs) scores of more than 80% (Table S3),
286 suggesting high-quality transcriptomes with most single-copy orthologs from the Embryophyta
287 database present in all species.

288 Due to the dominance of terpenoid compounds in the floral scent, we focused on the
289 expression of terpene synthase genes (TPS). TPS genes are responsible for producing terpenoid
290 compounds. We found that 20 TPS genes were expressed in the floral transcriptome of *C. allenii*,
291 compared to five in *C. villosissimus*, and only one in *C. lasius*, when expression was defined as
292 one transcript per million (TPM) in at least one sample of the species. Furthermore, we found
293 that counts of TPS expression per TPS gene expressed were much higher in both *C. allenii* and
294 *C. villosissimus* when compared to *C. lasius* (Fig. 4B). In fact, in *C. allenii,* a TPS is the highest
295 expressed gene in the floral transcriptome, and TPSs are also the 4th and 6th most expressed genes
296 in the transcriptome, highlighting the extent of expression. On average, 12% of total expression
297 can be attributed to TPSs in *C. allenii*, 0.17% in *C. villosissimus* and 0.00014% in *C. lasius*,
298 highlighting the orders of magnitude differences between species.

299

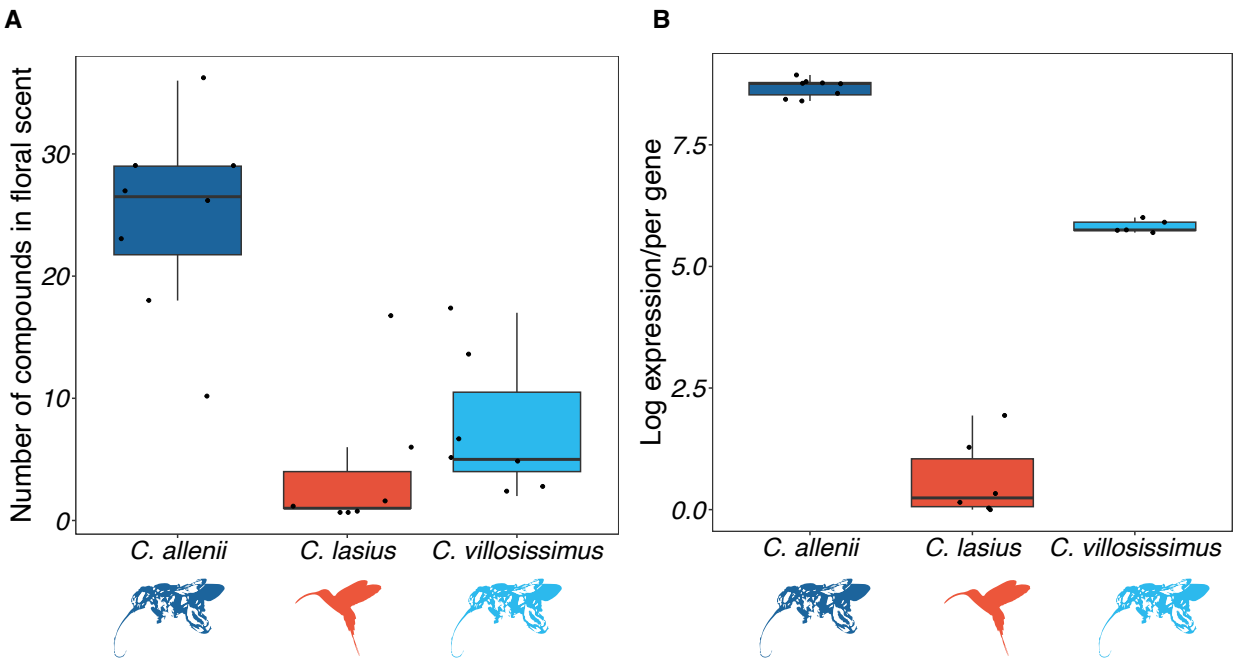


Figure 4. Differences in floral scent and gene expression between *C. allenii*, *C. lasius*, and *C. villosissimus*. Species differ significantly in A) number of floral scent components (ANOVA, $F_{2,19}$=20.9, p<0.001), and (B) expression of terpene synthase transcripts (ANOVA, $F_{2,16}$=508.4, p<0.001).

Floral scent loss is not driven by gene loss in *Costus*

While the transcriptome assemblies can give us insight into levels of TPS expression, it is hard to determine gene copy number due to the redundancies in the assemblies. For this, we need genome data to allow us to determine the number of TPS genes in each species. There are currently three publicly available *Costus* genomes: *C. bracteatus* (bee), *C. lasius (hummingbird)*, and *C. pulverulentus* (hummingbird) (Valderrama et al. 2022; Harenčár et al. 2023). *Costus lasius* and *C. pulverulentus* represent two independent transitions to hummingbird pollination (Kay and Grossenbacher 2022). To generate a dataset of two hummingbird- and two bee-pollinated species, we assembled an additional high-quality reference genome for bee-pollinated *C. allenii* (Table S4 for assembly statistics, Figure S3 to see alignment with *C. lasius* genome). We generated high-quality whole genome annotations for each species with high completeness and consistency (Table S5).

We annotated between 63-75% of each genome as repeat elements (Tables S6, S57, S8, S9). We found that, of the elements that could be classified, retrotransposable elements

319 represented the largest percentage of the genome. These are the most abundant form of repeat
320 elements found in angiosperm genomes. Many repeats were not classified into previously known
321 families, perhaps suggesting the existence of new repeat elements.

322 We annotated a total of 66 TPSs in *C. allenii*, 44 in *C. bracteatus*, 40 in *C. lasius*, and 44
323 in *C. pulverulentus*. All TPSs belonged to previously described TPS families TPS-a, TPS-b,
324 TPS-c, TPS-e/f, and TPS-g (Fig. 5, Fig. S4). Those TPSs responsible for floral scent biosynthesis
325 are primarily found in the TPS-a, TPS-b, and TPS-g families. TPS-a members tend to produce
326 monoterpenes, TPS-b members sesquiterpenes, and TPS-g members either monoterpenes or
327 sesquiterpenes. We identified lineage-specific duplications in both TPS-a and TPS-b, as well as
328 TPS-e/f in the bee-pollinated *C. allenii*. We did not find any evidence for pseudogenization or
329 gene loss in the two hummingbird-pollinated species included in the gene family analyses, *C.*
330 *lasius* and *C. pulverulentus*, with TPS genes containing complete protein domains and start
331 codons.

332 We used computational analysis of gene family evolution (CAFE5) to model gene gain
333 and loss across the species phylogeny for each TPS clade identified in Figure 4 (Mendes et al.
334 2020). We found that *C. allenii* has gains in all five clades (TPS-a, TPS-b, TPS-c, TPS-e/f, TPs-
335 g) with no evidence of losses, *C. bracteatus* shows no gains but losses in 4 clades, *C. lasius* and
336 *C. pulvulerentus* both show gains in 1 and losses in 3. There is no clear correlation between
337 pollination group and gene losses or gains with some evidence that *C. allenii* show enhanced
338 gene duplication. We also estimated gene duplications and losses for each species using a gene
339 tree-species tree reconciliation algorithm, reconcILS (Mishra et al. 2024). The algorithm
340 reconciles the gene tree and species tree considering not only duplication and loss events but also
341 incomplete lineage sorting (ILS). This is important for *Costus* due to the recent divergence times
342 between species (Vargas et al. 2020). Using this approach, we identified the greatest number of
343 duplication events for *C. allenii* (20), with 6 in *C. bracteatus*, and 4 in each hummingbird-
344 pollinated species (*C. lasius* and *C. pulvulerentus*). We found the fewest loss events in *C. allenii*
345 (2), with more losses in *C. pulvulerentus* (5), *C. bracteatus* (6), and *C. lasius* (11). We find
346 support for increased lineage-specific duplications in *C. allenii*, however, our gene family
347 analyses do not find evidence for increased gene losses in hummingbird-pollinated species.

348 We mapped the RNAseq reads both to the *C. allenii* genome and the *C. lasius* genome. In
349 *C. allenii*, we found a consistent percentage of successfully mapped reads attributed to TPSs as
350 with the transcriptome (12% transcriptome, 13% mapped to either genome). We found some
351 differences for both *C. lasius* and *C. villosissimus*. For the *C. villosissimus* transcriptome
352 mapping, we found 0.17% of reads were attributed to TPSs, in comparison to 0.009% when
353 mapped to either genome. Due to a lack of genome for *C. villosissimus*, it is unclear whether this
354 could be due to genetic changes or rearrangements not captured by the available genomes. For *C.*
355 *lasius*, we found only 0.00014% of reads mapped to the transcriptome were due to TPSs, in
356 contrast to 0.011% of the genome-mapped reads.

357 When *C. allenii* reads were mapped to the *C. allenii* genome, we found high levels of
358 multi-mapping. This is due to recent TPS duplications in *C. allenii* which are highly expressed.
359 To illustrate this, the highest gene expressed in *C. allenii* is *TPS49*, and the third most expressed
360 "gene" consists of reads that multimap to *TPS49* and *TPS52*. This clade is represented by only
361 one gene in each of the other species. Other recent lineage-specific duplications include *TPS2*,
362 *TPS3*, *TPS4*, and *TPS5* where the majority of reads multimap due to sequence similarity between
363 duplicates. Both of these clades have only one member in each of the other species with an
364 available genome sequence (Fig S3). This highlights the complexity of mapping to gene families
365 with recent duplications, especially when they are highly expressed, meaning that the typical
366 choice of ignoring multi-mapped reads will lose large percentages of mapped reads from these
367 duplications. We, therefore, decided to carry out our differential expression analysis using the *C.*
368 *lasius* genome.

369 By mapping all reads to the *C. lasius* genome, we identified 4790 genes differentially
370 expressed between *C. allenii* and *C. lasius*, 7678 between *C. allenii* and *C. villosissimus*, and
371 6191 between *C. lasius* and *C. villosissimus*. Of the 40 TPS genes found in the *C. lasius* genome,
372 15 were differentially expressed between *C. allenii* and *C. lasius*, with 13 of these upregulated in
373 *C. allenii*. This includes *TPS33*, orthologous to the *TPS49*/*52* clade, the most highly expressed
374 genes in *C. allenii*. We found 19 TPSs that were differentially expressed between *C. allenii* and
375 *C. villosissimus*, with 14 upregulated in *C. allenii* and 5 in *C. villosissimus*. Ten TPS genes were
376 differentially expressed between *C. lasius* and C. *villosissimus*, with 7 upregulated in *C.*
377 *villosissimus* and 3 in *C. lasius*. Overall, more TPSs were upregulated in *C. allenii* when

compared to either *C. lasius* or *C. villosissimus*. Our gene ontology analysis found little evidence for functional enrichment in differentially expressed genes between species (Table S10).

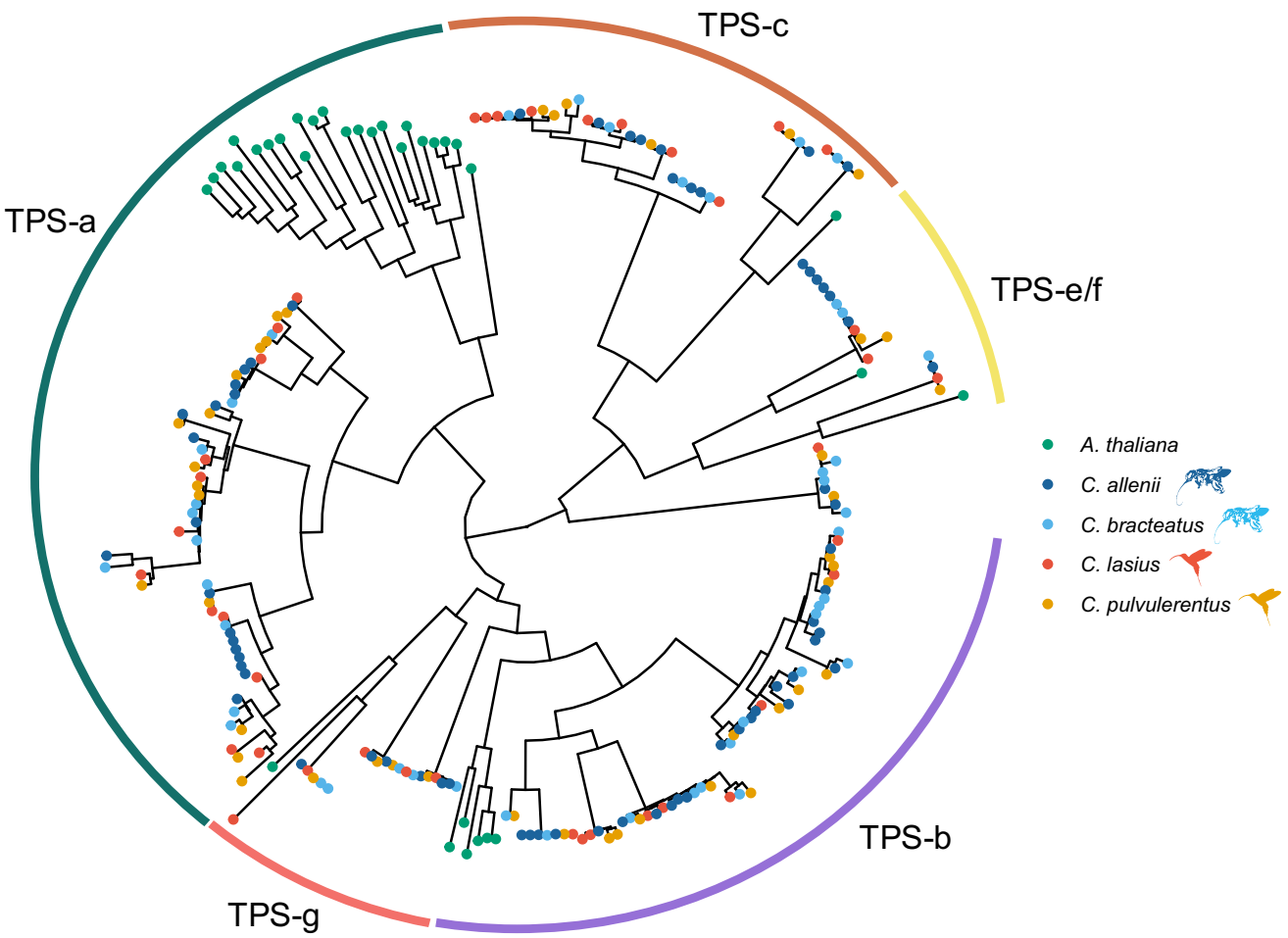


Figure 5. Phylogenetic analysis of terpene synthase gene families. We found the highest rates of gene gain and lowest rates of gene loss for *C. allenii* with the other three species exhibiting mixes of loss and gains across the gene tree. The phylogeny was constructed in IQ-TREE with 1000 bootstraps and model LG. TPS family classifications based on *Arabidopsis thaliana* are illustrated.

Discussion

388 Repeated evolutionary transitions provide unique opportunities to study the extent of
389 convergence both at the phenotypic and molecular level, shedding light on the predictability of
390 evolution. Transitions to hummingbird pollination are widespread in the Americas and are
391 predicted to be associated with loss of floral scent, however, this has not been studied in a
392 comparative framework across multiple independent transitions within the same genus. In this
393 study, we combine comparative chemical ecology with genomics and transcriptomics and
394 demonstrate that floral scent in *Costus* is dominated by terpenes and its evolution is highly
395 associated with pollinator type. We found that hummingbird-pollinated *Costus* species have lost
396 or reduced their floral scent across eight independent evolutionary origins, whereas bee-
397 pollinated species in the Americas exhibit either maintenance of floral scent or gain of a more
398 diverse scent profile. We find that this is related to downregulation of genes associated with
399 floral scent production (in this case terpene synthases) in floral tissues of hummingbird-
400 pollinated species but is not accompanied by loss of terpene synthase genes.

401 Bee-pollinated *Costus* species showed much greater variation in their floral scents than
402 hummingbird-pollinated species, with some having very diverse floral scents, yet others
403 producing far fewer compounds. Bee-pollinated species also show greater morphological
404 variation, possibly reflecting variation in orchid bee diversity (Kay and Grossenbacher 2022).
405 Furthermore, not all species in our dataset follow the expected patterns. For example, *Costus*
406 *arabicus* and *Costus vinosus* (bee-pollinated) produce little scent. It is unclear whether the
407 variation in floral scent between different bee-pollinated species is related to pollinator
408 preference, other abiotic or biotic factors, or historical contingency.

409 There are multiple hypotheses for why floral scent would be lost in hummingbird-
410 pollinated species. Firstly, it could be because there is simply no selection to maintain the
411 production due to poor sense of smell, or lack of preference shown for floral scent, by
412 hummingbirds (Goldsmith and Goldsmith 1982; Knudsen et al. 2004; Driver and Balakrishnan
413 2021). If this were the case, we might expect to find a correlation between length of time
414 following pollinator transition and extent of floral scent loss. However, our data shows that
415 downregulation of floral scent can occur on relatively quick evolutionary timescales. The Latin
416 American radiation of *Costus* is dated at 3 million years (95% CI, 1.50–4.87) and so all
417 transitions to hummingbird-pollination have occurred within this timeframe (Vargas et al. 2020).
418 *Costus wilsonii*, *C. lasius*, and *C. spiralis* are all hummingbird-pollinated species that are most

closely related to bee-pollinated species, representing more recent evolutionary shifts to hummingbird pollination. Yet all three species produce little to no floral scent, suggesting there can be quick breakdown of floral scent production following pollinator transitions. This observation is consistent with previous studies that have shown that floral scent often evolves rapidly in response to selection (Ramos and Schiestl 2019; Zu et al. 2020; Liu et al. 2024).

The rapid and repeated loss of scent in hummingbird-pollinated species suggests a shared strong selection pressure against floral scent production. One potential mechanism is the energetic cost of terpene production. A similar argument has been made in the case of eye regression in cavefish. If vision is energetically costly to maintain, then it should be selected against when no longer useful (Moran et al. 2015; Sifuentes-Romero et al. 2023). This is supported by evidence of the energetic costs of maintaining a visual system (Moran et al. 2015). The metabolic cost of terpene production is unknown but is thought to be minimal (Raguso 2016; Pichersky and Raguso 2018). In fact, it has been argued that the complex network of terpene metabolism, in which individual enzymes produce multiple products, makes terpene pathways more efficient, compared with a linear pathway with separate enzymes and regulators for each step (Lanier et al. 2023).

An alternative scenario to the energetic cost is that terpene production has ecological costs in hummingbird-pollinated species. For example, there may be strong selection against bee attraction (Castellanos et al. 2004; Bergamo et al. 2016). Many benefits to hummingbird, compared with bee, pollination have been described. Birds do not consume pollen, leading to less pollen loss, and birds can travel longer distances between plants and promote higher gene flow, which is beneficial to combat inbreeding depression (Thomson and Wilson 2008; Krauss et al. 2017; Dellinger et al. 2022). One caveat is that comparisons in pollination effectiveness between bees and hummingbirds have not been carried out in orchid bees. Both male and female orchid bees tend to forage over long distances, perhaps resulting in a more “bird-like” pollen transport (Janzen 1971; Janzen 1981; Wikelski et al. 2010; Opedal et al. 2017; Kay and Grossenbacher 2022; Gamba and Muchhala 2023). While pollinator efficiency is unknown in *Costus*, rapid floral scent loss is congruent with the hypothesis of “bee-avoidance” following establishment of hummingbird pollination.

448 A final piece of evidence to support bee-repellence is the large reduction in beta-ocimene
449 production in hummingbird-pollinated flowers relative to bee-pollinated flowers, as this
450 compound is a common insect pollinator attractant (Farré-Armengol et al. 2017). In *Mimulus*,
451 loss of monoterpenes, including beta-ocimene, maintains reproductive isolation between bee-
452 and hummingbird-pollinated species (Byers et al. 2014). Interestingly, we found that α- and β-
453 pinene were produced by both bee- and hummingbird-pollinated species. One potential reason is
454 that pinenes do not repel orchid bees but do repel other types of bees. Terpenoid-dominated
455 scents have been found to repel honey bees (Larue et al. 2016), and in particular α-pinene has a
456 repellent effect on honey bees (Fernandes et al. 2019). Furthermore, α-pinene has been found in
457 other hummingbird-pollinated species with a suggested role as a moth pollinator deterrent
458 (Bischoff et al. 2014). We speculate that bee-pollinated species emit a mix of compounds, some
459 of which potentially attract orchid bees, and some of which potentially repel other pollinators to
460 narrow attraction to orchid bees, specifically. Then when floral scent is reduced in hummingbird-
461 pollinated flowers, the compounds which are attractive to orchid bees are lost but not those
462 which are general bee repellants. Further experimental work is needed to test these hypotheses.

463 We found that the floral scents of *Costus* are dominated by terpene compounds. Although
464 few Zingiberales have described floral scents, our results are consistent with previous reports
465 from the genus *Hedychium* that also found terpene-dominated floral scents (Báez et al. 2011;
466 Zhou et al. 2022). Terpene diversity in a given organism is mainly determined by the terpene
467 synthase (TPS) gene family, which carries out crucial steps in terpene formation (Tholl 2006).
468 We did not find evidence that hummingbird-pollinated species with an available genome
469 sequence (*C. lasius* and C. *pulverulentus*) exhibited increased rates of TPS gene loss than the
470 bee-pollinated species. There are multiple reasons why trait loss would be associated with
471 changes in gene regulation rather than gene loss. While *C. lasius* represents a more recent
472 transition to hummingbird pollination than *C. pulvulerentus*, both transitions have occurred
473 within the last ~3 million years (95% CI, 1.50–4.87)as *Costus* radiated in Latin America (Vargas
474 et al. 2020). There may not have been enough time for pseudogenization to occur, a process
475 estimated to take 0.5-6 million years (Marshall et al. 1994; Esfeld et al. 2018). Potentially, the
476 genes will eventually be lost but are currently in the first phase, reduction in expression, which
477 according to the "step-wise" model to pseudogenization is then followed by relaxed selection
478 and then loss-of-function mutations (Graham et al. 2023).

479 Alternatively, we might not expect any TPS genes to be lost even in the long term.
480 Terpenes, and therefore TPSs, are important for many other roles in plants. Plants use terpenes to
481 deter herbivores, to attract parasitoids or predators of herbivores, and as anti-fungal and anti-
482 bacterial compounds to prevent disease (Gershenzon and Dudareva 2007). These pleiotropic
483 functions of TPSs could explain why we observe downregulation in floral tissues specifically
484 rather than any pseudogenization. One argument against this is that in other systems TPS
485 pseudogenization does occur following pollinator transitions. In *Mimulus*, loss of monoterpene
486 production (specifically beta-ocimene), is not due to expression differences but loss-of-function
487 mutations (Peng et al. 2017). It is unclear whether pseudogenization will occur at some point in
488 *Costus*.

489 Although we did not observe gene loss in hummingbird-pollinated lineages, we did
490 observe extensive lineage-specific duplications in the bee-pollinated species *C. allenii*. *Costus*
491 *allenii* was one of the species with highest scent diversity in our dataset, and we found multiple
492 TPS genes to exhibit several recent lineage-specific duplication events. We also found extremely
493 high TPS expression, accounting for approximately 12% of expression in floral tissue. The gene
494 with the highest expression in floral tissue was found to be a TPS-a family member, with high
495 expression also found also for several TPS-b members. These two families produce
496 sesquiterpenes and monoterpenes, respectively, both of which are found in the floral scent of *C.*
497 *allenii*. In *C. lasius* and *C. villosissimus*, we found reduced TPS expression compared to *C.*
498 *allenii*, as expected based off their reduced floral scent production. We also saw high levels of
499 phenotypic variation in *C. villosissimus*, and there is no publicly available genome, making
500 interpretation of the results more difficult. We speculate that *C. allenii* has higher floral scent
501 production than closely related *C. villosissimus* because *C. allenii* grows as isolated individuals
502 in deeply shaded, perennially wet forest, whereas *C. villosissimus* grows at forest edges and large
503 treefall gaps, and often has conspecifics nearby (Chen and Schemske 2015). Thus *C. allenii* may
504 need more scent to attract pollinators. While the overall trends are clear, we hesitate to over-
505 interpret the gene expression data as we have only sampled at one timepoint and it is possible
506 that there is floral scent production happening earlier that is not captured in our data.

507 We observe more consistency and predictability in the phenotypic patterns than the
508 underlying molecular mechanisms. In many systems, parallel changes in floral scent have
509 accompanied independent pollinator shifts, including the current study (Knudsen et al. 2004;

510 Byers et al. 2014). This can also be seen in other reproductive shifts, such as transitions to self-
511 fertilization (Wozniak et al. 2022). In some systems, the underlying molecular mechanisms of
512 floral scent loss are also convergent, having been shown to involve the same loci in multiple
513 independent evolutionary events. For example, loss of floral scent in different *Mimulus* lineages
514 is due to changes at the same loci, including frameshifts, deletions, and potentially post-
515 transcriptional regulation (Peng et al. 2017). A similar pattern was found in two different systems
516 that have lost benzaldehyde floral scent production due to either evolution of self-fertilization
517 (Sas et al. 2016), or transition to hummingbird pollination (Amrad et al. 2016). The same locus
518 contains loss-of-function mutations in both cases (Raguso 2016). However, this is not universally
519 true. In *Capsella*, the loss of beta-ocimene following transition to self-fertilization in one lineage
520 is due to changes in subcellular enzyme localization, but QTL results suggests another locus is
521 responsible for scent loss in a related species (Wozniak et al. 2022). In this study, we saw no
522 evidence for pseudogenization. Instead, we found evidence that gene expression differences are
523 responsible for scent differences between *Costus* species, with scent loss likely due to changes
524 in, potentially tissue-specific, gene expression.

525 In summary, we have shown that *Costus* species exhibit repeated loss of floral scent
526 following transitions from bee to hummingbird pollination, suggesting a shared strong selection
527 pressure favoring scent loss. We also find bee-pollinated species that have recently gained more
528 diverse scents, which, at least in *C. allenii*, is associated with lineage-specific terpene synthase
529 duplications. We show that scent loss is not associated with loss of terpene synthase genes but
530 rather appears to be due to gene downregulation. This demonstrates the capacity for rapid
531 metabolic changes in response to selection following pollinator transitions. This system provides
532 us with the exciting opportunity to identify the cis- or trans-regulatory changes driving gene
533 expression differences in multiple independent transitions to hummingbird pollination, allowing
534 us to study the extent of genetic convergence.

535 Materials and Methods

536 Floral scent sample collection

537 We collected 100 samples of 30 species of *Costus* between 2019 and 2023 (Table S1).
538 We sampled an average of 4 individuals per species for bee-pollinated species, and 3 for
539 hummingbird-pollinated species. We, therefore, expect that missing compounds due to under-

540 sampling will occur to a similar extent for both bee- and hummingbird-pollinated species. All
541 samples for *Costus glaucus*, *Costus laevis*, *Costus montanus*, *Costus stenophyllus*, and *Costus*
542 *wilsonii* were collected in Costa Rica (La Gamba, Las Cruces, Las Alturas, and Monteverde). All
543 other species were sampled in the UCSC greenhouses in Santa Cruz, California, USA from
544 plants composing the UCSC living *Costus* collection. For each day of sampling, both floral scent
545 samples and ambient samples were collected (with one exception: a *Costus scaber* sample
546 collected on April 8th 2019 was missing a control and so we used a control from the nearest date
547 available, March 7th). Scent was sampled from a single flower using dynamic headspace
548 sampling from approximately 9 am until 3 pm (6 hours). This time point was chosen as it is the
549 window of most pollinator activity during the lifespan of the flower (Kay and Schemske 2003).
550 Flowers from living plants were placed inside oven bags (Reynold's). A Spectrex PAS-500 air
551 pump with Tygon tubing was used to pull scented-air through a volatile trap for 6 hours at a flow
552 rate of 200 ml air / min. The volatile trap contained a filter of glass wool and 50 mg Porapak. We
553 conditioned scent traps by passing 5 mL of hexane before usage. Volatiles were eluted from the
554 trap using 400µl of hexane and stored at -20°C.

555 Chemical analysis

556 Samples were analyzed using Agilent model 5977A mass-selective detector with an
557 Agilent GC model 7890B. Samples were run on a HP-5 Ultra Inert column (Agilent, 30m x
558 0.25mm, 0.25 µm). For analysis, one microliter of each sample was injected in splitless mode
559 using the ALS 7694 autosampler with helium as the carrier gas (250°C injector temperature).
560 The program started at 40 °C for 3 minutes and then rose at 5 °C/min to 210°C. The temperature
561 was held at 210°C for 1 minute, before rising at 20°C/min to 300°C where it was held for 2.5
562 minutes, and 315°C for a final minute.

563 Compounds were identified by comparing mass spectra and retention indices with
564 reference libraries. The identity of some compounds was confirmed by comparison to authentic
565 reference standards (see Table S2). Compounds that were not found in at least two samples were
566 removed. A two-step process was used to filter datasets for floral scent components. First,
567 compounds were filtered to include only those which are found in five times higher amounts in
568 floral than ambient samples using the "filter_ambient_ratio" function from the *bouquet* package
569 (Powers et al. 2023). This resulted in a list of ambient compounds which were not considered

570 floral scent compounds. Second, using the raw data matrices, we subtracted the ambient
571 compound amounts from the corresponding floral scent amount collected on the same date. We
572 then removed the ambient compounds identified by the package *bouquet* from this subtracted
573 matrix to create the final data matrix of floral scent.

574 Statistical analyses of floral scent

575 To visualize divergence in floral scent among samples we used nonmetric
576 multidimensional scaling (NMDS) using the "metaMDS" function in *vegan* (Oksanen et al.
577 2020). To test for differences both between pollination groups, and between floral and ambient
578 samples, we carried out a PERMANOVA (permutational multivariate analysis of variance) using
579 the "adonis2" function in *vegan* (Bray–Curtis distance matrix, 1000 permutations). To identify
580 which groups were significantly different, we carried out post hoc pairwise testing using the
581 "pairwise.perm.MANOVA" function in the *RVAideMemoire* package with a Bonferroni
582 correction (Hervé 2021). We did this for both filtered and unfiltered datasets.

583 We tested for differences in compound number produced by bee-pollinated and
584 hummingbird-pollinated species using ANOVA with a nested model including species within
585 pollinator group. To test for differences in the diversity of compounds produced by bee-
586 pollinated and hummingbird-pollinated species, we used the package *chemodiv* (Petrén et al.
587 2023). This package considers the biochemical and structural diversity of the compounds present
588 in a sample to calculate the overall chemodiversity per sample. First, we used the "calcDiv"
589 function to calculate the functional Hill diversity index for each sample. Then we carried out an
590 ANOVA test for the diversity indices, using a nested model of species nested within pollinator
591 group to test for differences in chemodiversity. We also tested for differences in compound
592 production using ANOVA. In this case a proxy for compound amount was used: the sum of the
593 total area of floral scent peaks in the sample, or the total ion abundance in the sample. To test for
594 difference in compound frequency between bee- and hummingbird-pollinated samples, we
595 carried out chi-squared tests and adjusted the p-values using the "p.adjust" function in R.

596 Phylogenetic analyses of floral scent

597 To investigate the evolution of floral scent in a phylogenetic context we used a previously
598 published phylogeny (Vargas et al. 2020). First, we trimmed the phylogeny to include only those
599 species for which we have scent data using the "drop.tip" function in *ape* (Paradis and Schliep

2019). To test for correlation between pollination group and the phylogenetic covariance matrix, we used a two-block partial least squares analysis using the "two.b.pls" function in the *geomorph* package (Adams and Otárola-Castillo 2013; Adams and Collyer 2018). A significant result means that our power to test for differences between groups using phylogenetic simulation-based is weakened due to group aggregation on the phylogeny. We then used the "phylosig" function in the package *phytools* to test for phylogenetic signal in multiple traits (compound number, compound diversity, compound amount, (*E*)-beta-ocimene production, and (E)-beta-caryophyllene production) (Revell 2012). We also tested if floral scent data overall shows phylogenetic signal using the multivariate generalization of Bloomberg's K (K_{mult}) (Adams 2014). Under Brownian motion, K_{mult} is expected to be 1. If $K_{mult}>1$, higher phylogenetic signal is detected than the null expectation under Brownian motion. If $K_{mult}<1$, lower phylogenetic signal is detected than the null expectation under Brownian motion. To calculate the K_{mult} of *Costus* floral scent, we used the filtered dataset with 1000 iterations to test significance. We carried out the analysis using the "physignal" function in the *geomorph* package (Adams and Otárola-Castillo 2013). Due to the weak phylogenetic signal detected in both multivariate and univariate analyses, we used the "procD.pgls" function in the *RRPP* package to test for differences in traits between pollinator groups rather than a simulation-based ANOVA (Adams and Collyer 2018; Collyer and Adams 2018). This function uses a method of randomizing residuals in a permutation procedure to evaluate the significance of phenotypic differences between groups considering the phylogeny and has higher statistical power than methods using phylogenetic simulations (Adams and Collyer 2018). We set lambda at 4.14795e-05 as calculated for the univariate variables. We corrected for multiple testing using the "p.adjust" function in R.

To better understand how floral scent has evolved in the genus *Costus*, we carried out an ancestral state reconstruction of the number of compounds produced. The species-level phylogeny from (Vargas et al. 2020) contains 24 hummingbird-pollinated species, 25 bee-pollinated species, and 3 bee-pollinated African outgroups. Our subsampled phylogeny contains samples in proportion to the full phylogeny (13 hummingbird-pollinated species, 13 bee-pollinated species, and 1 bee-pollinated African outgroup). We note that there are limitations to this sub-sampled phylogeny but we find value in using it to infer general trends and not values at specific nodes. To do this, we used the function "ace" in the *ape* package for ancestral character

631 estimation (Paradis and Schliep 2019). We divided the compound number trait into twelve
632 categories, each of which were assigned a different shade of blue. We plotted both the compound
633 number for each phylogenetic tip label and internal node to illustrate how floral scent has
634 evolved in *Costus*. We overlaid the origins of hummingbird pollination as described in (Vargas
635 et al. 2020).

636 Sample collection for RNAseq

637 Live individuals of *Costus allenii* and *Costus villosissimus* were collected in Soberanía
638 National Park, Colón Province, Panama. Vouchers are deposited at University of Panama
639 (*Costus allenii*: 0132264, 0132265; *Costus villosissimus*: 0132270). Live individuals of *Costus*
640 *lasius* were collected in El Valle de Antón, Cocle Province, Panama. Voucher is deposited in
641 Michigan State University (MSC) herbarium (Kay 0321). For all three species, the individuals
642 used for RNAseq analyses were grown in greenhouses at UCSC. Freshly opened flowers of each
643 species were removed at approximately 9 am and flash frozen immediately in liquid nitrogen
644 before being stored at -80°C until extraction. Flowers of *Costus* open early in the morning and
645 fall off by mid-afternoon.

646 RNA extraction and sequencing

647 Frozen flowers were removed and dissected on dry ice. Tissue from the petals and labella
648 was broken down using a TissueLyser (Qiagen). RNA was then extracted using an RNeasy Mini
649 Kit (Qiagen) with Qiashredder spin columns (Qiagen) following standard protocols. Samples
650 were sent to Novogene for quality control, library preparation, and sequencing. Our final dataset
651 consisted of 19 samples: 8 *Costus allenii* (from 4 individual plants), 6 *Costus lasius* (from 2
652 individual plants), and 5 *Costus villosissimus* (from 1 individual plant). Samples were sequenced
653 on a NovaSeq 6000 using 150 bp paired end reads. This generated approximately 53 million
654 reads per library (mean=52.74 million, SD=12.01 million, N=19). The raw sequence data were
655 deposited in the SRA under the BioProject accession PRJNA1010186.

656 Transcriptome assembly

657 Before assembling transcriptomes, we trimmed the raw reads using TrimGalore! (Martin
658 2011). We then assembled transcriptomes for each species using Trinity version 2.5.1 with
659 default settings (Grabherr et al. 2011; Haas et al. 2013). To reduce redundancy, we clustered

660 contigs with a minimum sequence identity of 95% using CD-HIT version 4.6.8 (Fu et al. 2012).
661 We assessed transcriptome completeness and quality using BUSCO (v5.6.1)
662 (embryophyta_odb10 dataset) (Simão et al. 2015; Manni et al. 2021; Zdobnov et al. 2021). We
663 also checked assembly quality by mapping RNAseq reads back to the assembly using Bowtie
664 (v2.4.1) to calculate mapping rates (Langmead and Salzberg 2012).

665 Terpene synthase gene expression analysis

666 We extracted the longest isoforms for each Trinity "gene" using the Trinity script
667 "get_longest_isoform_seq_per_trinity_gene.pl" and translated these sequences. We then
668 searched for potential TPSs using hmmsearch two HMM profiles: Terpene_syth_C (PF03936)
669 and Terpene Synthase N-terminal domain (PF01397). We used the R package rhmmer to parse
670 the output from hmmsearch resulting in a list of putative TPS sequences for each species. To
671 reduce the TPSs to only those expressed in the dataset, we used a threshold of TPM>1
672 (transcripts per million) in at least one individual of a species (Eddy 2011; Arendsee 2017). To
673 calculate expression of each putative TPS we normalized counts using Gene length corrected
674 trimmed mean of M-values (GeTMM), an approach that improves inter- and intra-sample
675 comparisons (Smid et al. 2018). To do this, we first calculated FPKM (fragments per kilobase of
676 transcript per million fragments mapped) for each transcript and carried out TMM-normalization
677 using the package *edgeR* (Robinson et al. 2010). We then combined GeTMM values for
678 transcripts which were identified by Trinity to be the same gene. This does not normalize counts
679 between species but is used to determine whether genes are lowly or highly expressed within a
680 species.

681 *Costus allenii* genome assembly

682 We selected a wild-collected individual *C. allenii* from Pipeline Road, Soberanía
683 National Park, Colón Province, Panama for *de novo* genome assembly. A voucher for this
684 population of *C. allenii* is deposited at the University of Panama Herbarium (PMA:0132264).
685 This individual was grown to flowering in the UCSC greenhouses from a cutting, and we
686 harvested 1.5 g fresh tissue from leaf meristems and flower buds. We isolated cell nuclei
687 following Workman (Workman et al. 2019) and then used the Circulomics Nanobind Plant
688 Nuclei Big DNA kit to extract HMW DNA. We assessed DNA concentration with a Qubit
689 Fluorometer (Thermo Fisher Scientific), purity with 260/280 and 260/230 absorbance ratios from

690 a NanoDrop Spectrophotometer (Thermo Fisher Scientific) and ran a 7% agarose gel at 70 V for
691 over 8 h to assess DNA fragment integrity and size. HMW DNA was sent to the UC Davis
692 Genomics Center for additional cleaning, Pacific Biosciences HiFi library preparation, and
693 sequencing on one PacBio Revio SMRT cell. This produced 74.8 Gb of HiFi yield and 7.9 M
694 HiFi reads. We assembled the PacBio data into a primary and alternative assembly using
695 HiFiasm (v0.19.8)(Cheng et al. 2021). We then used RagTag (v2.1.0)(Alonge et al. 2022) to
696 scaffold *C. allenii* contigs using the closely related *C. lasius* as a reference. We used D-GENIES
697 (Cabanettes and Klopp 2018) with Minimap2 (Li 2018) to generate a dotplot comparing the two
698 assemblies both before and after scaffolding. The genome assembly quality was assessed using
699 BUSCO v 5.6.1 (Simão et al. 2015). Raw sequence data and the final genome assembly were
700 submitted to NCBI under BioProject PRJNA1091680.

701 Genome annotations

702 Of the three species for which we generated RNAseq data, an unannotated genome was
703 previously available for *C. lasius* (Harenčár et al. 2023). In this study we also generated a
704 genome assembly for *C.allenii*. In addition, previously published genomes are available for both
705 *C. bracteatus* and the more distantly-related *C. pulverulentus* (Valderrama et al. 2022; Harenčár
706 et al. 2023). We started by processing RNAseq data as evidence for annotations. Trimmed reads
707 from *C. lasius* were mapped using 2pass mapping in STAR to the *C. lasius* genome, and the *C.*
708 *pulverulentus* genomes (Martin 2011; Dobin et al. 2013). Bam files were concatenated for use as
709 described below. Trimmed reads from *C. allenii* were mapped using 2pass mapping in STAR to
710 both the *C. allenii* and the closely related *C. bracteatus* genomes, and bam files concatenated per
711 species (Martin 2011; Dobin et al. 2013).

712 We annotated the genomes using BRAKER3 (v3.0.7) which combines RNA-seq and
713 protein data in an automated pipeline (Lomsadze et al. 2005, 2014; Stanke et al. 2006, 2008;
714 Gotoh 2008; Iwata and Gotoh 2012; Buchfink et al. 2015; Hoff et al. 2016, 2019; Kovaka et al.
715 2019; Brůna et al. 2020, 2021; Pertea and Pertea 2020; Gabriel et al. 2023). This pipeline
716 requires a masked genome. We identified and annotated tandem repeats and transposable
717 elements using RepeatModeler (Flynn et al. 2020) (v.2.0.4 for *C. lasius* and v.2.0.5 for the other
718 genomes) and RepeatMasker v.4.1.5 (Tarailo-Graovac and Chen 2009). We then used
719 RepeatMasker with the outputs of RepeatModeler to generate softmasked genomes. The masked

720 genomes were then provided to BRAKER along with plant proteins downloaded from OrthoDB
721 (odb10_plants) (Zdobnov et al. 2021). For all species, BRAKER was run in ETP mode which
722 combines evidence from both proteins and RNAseq (Gabriel et al. 2021). BUSCO v 5.6.1
723 (Simão et al. 2015) and OMArk (Nevers et al. 2022) were used to evaluate the completeness and
724 consistency of the sets of gene models. During the project, the *C. lasius* genome was updated
725 (direction of arms on chromosome 1 were altered) and so we re-scaffolded chromosome 1 of *C.*
726 *allenii* and transferred our annotation using liftoff (Shumate and Salzberg 2021).

727 Terpene synthase gene family annotation

728 Following whole-genome annotation, we manually curated the gene models for terpene
729 synthase (TPS) genes in the annotation. To do this, we used bitacora (Vizueta et al. 2020), a
730 pipeline which curates an existing annotation and finds additional gene family members. In this
731 pipeline, BLASTP and hmmer are used to search for gene family members of interest within the
732 existing annotation (Altschul et al. 1990; Eddy 2011). Additional regions of the genome are then
733 searched using TBLASTN and are annotated using GeMoMa (Keilwagen et al. 2019). Hmmer is
734 used to validate the gene models. As input, we used two HMM profiles: Terpene_syth_C
735 (PF03936) and Terpene Synthase N-terminal domain (PF01397) downloaded from Pfam. In
736 addition, we provided previously annotated genes from *Arabidopsis thaliana* and *Oryza sativa*
737 (Chen et al. 2011; Yu et al. 2020; Jia et al. 2022). Gene models from bitacora and the genome-
738 wide annotation from BRAKER were compared and TPS genes were manually curated using
739 IGV (Robinson et al. 2011). We also searched for additional TPSs by carrying out exonerate
740 searches using the curated *Costus* gene models (Slater and Birney 2005). We checked all
741 annotated TPSs for complete protein domains using the NCBI conserved domain search
742 (Marchler-Bauer et al. 2015). We excluded proteins with less than 250 amino acids.

743 Gene family analysis

744 To classify these TPSs in previously described TPS families, we constructed a
745 phylogenetic tree based on amino acid sequences with TPSs from *Arabidopsis thaliana*. The
746 *Costus* TPS protein sequences, combined with those from *A. thaliana*, were aligned with
747 MAFFT v7.508 (-maxiterate 1000, using L-INS-I algorithm) (Katoh and Standley 2013). We
748 then trimmed the alignment using trimAl (-gt0.6, sites only included when present in 60% of
749 sequences) (Capella-Gutierrez et al. 2009). We used this alignment to construct a gene tree with

750 IQ-TREE and the ModelFinder function to determine the best-fit model (Nguyen et al. 2015;
751 Kalyaanamoorthy et al. 2017; Hoang et al. 2018). We rooted the tree using the *phytools* package
752 in R ("midpoint.root" function)(Revell 2012; R Core Team 2023). We plotted the resulting gene
753 trees using the following packages in R: *ape* (Paradis and Schliep 2019), *evobiR* (Blackmon and
754 Adams 2015), *ggnewscale* (Campitelli 2023), *ggstar* (Xu 2022), *ggtree* (Yu et al. 2017),
755 *ggtreeExtra* (Xu et al. 2021), and *tidytree* (Yu 2022).

756 To investigate patterns of gene family evolution we used computational analysis of gene
757 family evolution CAFE (v5.1) which uses a birth-death model of gene family evolution to model
758 gene gain and loss across a species tree (Mendes et al. 2020). The required input files are a
759 species tree which we adapted from (Vargas et al. 2020), and gene counts for each species in
760 each clade (as identified in Fig. 5).

761 We also took an alternative approach by using reconcILS to reconcile our gene tree and
762 species tree to estimate gene duplications and losses for each species (Mishra et al. 2024). The
763 algorithm considers not only duplication and loss events but also incomplete lineage sorting
764 (ILS), particularly relevant in this case due to the recent divergence times between *Costus*
765 species (Vargas et al. 2020). To create a gene tree, we aligned all *Costus* TPS nucleotide
766 sequences with MAFFT v7.508 (-maxiterate 1000, using L-INS-I algorithm) (Katoh and
767 Standley 2013). The alignment was trimmed using trimAl (-gt0.6, sites only included when
768 present in 60% of sequences) (Capella-Gutierrez et al. 2009) and used to construct a gene tree
769 with IQ-TREE and the ModelFinder function to determine the best-fit model (Nguyen et al.
770 2015; Kalyaanamoorthy et al. 2017; Hoang et al. 2018). We rooted the tree using the *phytools*
771 package in R ("midpoint.root" function)(Revell 2012; R Core Team 2023). The species tree was
772 adapted from (Vargas et al. 2020). We used the default parameter settings.

773 <u>Differential expression analyses</u>

774 First, we trimmed the reads using TrimGalore! (Martin 2011). We then mapped the reads
775 from all three species to both the *Costus allenii* and *Costus lasius* genome (Harenčár et al. 2023).
776 We ran 2pass mapping using STAR (Dobin et al. 2013). In general, we found higher levels of
777 multi-mapped *C. allenii* reads when mapped to *C. allenii*. To investigate this we used mmquant
778 (with merge, -m, set to 10) to count the reads, a tool which can account for multi-mapping reads
779 (Zytnicki 2017). The output from mmquant was read into R for further analysis. To calculate

780 percentage of gene expression due to TPS expression, we normalized reads by library size using
781 TMM (trimmed mean of M values) (Robinson and Oshlack 2010). Due to the consistency in TPS
782 expression as calculated by mapping to both genomes, we decided to use the *C. lasius* genome
783 assembly to identify genes differentially expressed between the three species to remove the
784 complication of multi-mapping. To do this, we first counted reads using featureCounts (Liao et
785 al. 2014) and again read the output into R. We filtered genes to include those with at least one
786 count per million (CPM) in at least one library and normalized reads by library size using TMM
787 (trimmed mean of M values). Differential gene expression was evaluated using the non-
788 parametric "noiseqbio" function from the *NOISeq* package (Tarazona et al. 2011; Tarazona et al.
789 2015). We used the default cut-off of q=0.95 in the "degenes" function which is equivalent to an
790 adjusted p-value of 0.05.

791 GO Analysis

792 We functionally annotated *C. alennii* genes using Blast2GO (Götz et al. 2008; Burge et
793 al. 2012). We imported these functional annotations into R to carry out enrichment analysis with
794 the package *topGO* (Alexa and Rahnenfuhrer 2022). We considered the background to be all
795 genes included in that specific differential expression analysis after filtering.

796 Plotting and data manipulation

797 Additional packages used for general plotting: *cowplot* (Wilke 2020), *ggplot2* (Wickham
798 2009), and *scico* (Pedersen and Crameri 2023). Packages used for data transformation and
799 manipulation include: *data.table* (Barrett et al. 2024), *dplyr* (Wickham et al. 2021), *knitr* (Xie
800 2023), *magrittr* (Bache and Wickham 2023), *readr* (Wickham et al. 2023), *reshape2* (Wickham
801 2007), and *tibble* (Müller and Wickham 2022). Analyses were carried out in R version 4.3.2 (R
802 Core Team 2023).

803 Data availability

804 Data and R scripts used for analysis are available from Open Science Framework:
805 https://osf.io/2ap4k/?view_only=6b8f5ff7d2ae499b871e1b1fb6fe6219. The raw sequence reads
806 and final genome assembly for *Costus allenii* are deposited in the SRA (BioProject

807 PRJNA1091680)

808 https://dataview.ncbi.nlm.nih.gov/object/PRJNA1091680?reviewer=p4jo1ioj3vqualqk1kvege7aq

809 l. The raw RNAseq data is deposited in the SRA (BioProject PRJNA1010186)

810 https://dataview.ncbi.nlm.nih.gov/object/PRJNA1010186?reviewer=fil9vpq6up5g3pj7e654rtokje

811 Acknowledgements

812 This work was funded by the National Science Foundation Dimensions of Biodiversity

813 grant awarded to K.M.K. (DEB-1737889) and S.R.R. (DEB-1737771), and the Jean H.

814 Langenheim Chair in Plant Ecology and Evolution held by K.M.K. We thank the Organization

815 for Tropical Studies, Estación Biológica Monteverde, and Estación Tropical La Gamba for

816 facilitating field work in Costa Rica and the Smithsonian Tropical Research Institute for

817 facilitating field work in Panama. All field research was conducted with appropriate research

818 permits in Costa Rica (M-P-SINAC-PNI-ACAT-026-2018, ACC-PI-0272018, MPC-SINAC-

819 PNI-ACLAP-020-2018, INV-ACOSA-076-18, MPC-SINAC-PNI-ACTo-020-18, CONAGEBIO

820 R-056-2019OT and R-058-2019-OT) and Panama (SE/AP-13-19, 09-99) We thank Julia

821 Harenčár for assistance with lab work, Pedro Juarez and Cecilia Girvin for assistance with field

822 sampling, and Alyssa Kaatmann, Aubrie Tait, Kate Uckele, Melina Sauerman Carrizosa, and

823 Selena Vengco for assistance with greenhouse sampling. We thank J. Velzy and S. Childress for

824 greenhouse care of the UCSC living *Costus* collection. We thank the Ramírez lab for helpful

825 discussions.

826

827 Author Contributions

828 Conceptualization, K.D., K.M.K., and S.R.R.; Investigation, K.D. and K.M.K.; Formal

829 Analysis, K.D.; Visualization, K.D.; Resources, K.M.K. and S.R.R.; Writing – Original Draft,

K.D.; Writing – Review and Editing, K.D., K.M.K., and S.R.R.; Supervision, K.M.K. and S.R.R.; Project Administration, K.M.K. and S.R.R.; Funding Acquisition, K.M.K. and S.R.R.

Declaration of Interests

The authors declare no competing interests.

References

Abrahamczyk S, Renner SS. 2015. The temporal build-up of hummingbird/plant mutualisms in North America and temperate South America. *BMC Evol. Biol.* 15:104.

Adams DC. 2014. A generalized K statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. *Syst. Biol.* 63:685–697.

Adams DC, Collyer ML. 2018. Phylogenetic ANOVA: Group-clade aggregation, biological challenges, and a refined permutation procedure. *Evolution* 72:1204–1215.

Adams DC, Otárola-Castillo E. 2013. geomorph: an r package for the collection and analysis of geometric morphometric shape data. *Methods Ecol. Evol.* 4:393–399.

Albert VA, Williams SE, Chase MW. 1992. Carnivorous Plants: Phylogeny and Structural Evolution. *Science* 257:1491–1495.

Alexa A, Rahnenfuhrer J. 2022. topGO: Enrichment Analysis for Gene Ontology.

Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 23:258.

Amrad A, Moser M, Mandel T, de Vries M, Schuurink RC, Freitas L, Kuhlemeier C. 2016. Gain and Loss of Floral Scent Production through Changes in Structural Genes during Pollinator-Mediated Speciation. *Curr. Biol.* 26:3303–3312.

Arendsee Z. 2017. rhmmer: utilities parsing “HMMER” results. Available from: https://CRAN.R-project.org/package=rhmmer

Bache SM, Wickham H. 2023. magrittr: A Forward-Pipe Operator for R. Available from: https://CRAN.R-project.org/package=magrittr

Báez D, Pino JA, Morales D. 2011. Floral Scent Composition in *Hedychium coronarium* J. Koenig Analyzed by SPME. *J. Essent. Oil Res.* 23:64–67.

Barreto E, Boehm MMA, Ogutcen E, Abrahamczyk S, Kessler M, Bascompte J, Dellinger AS, Bello C, Dehling DM, Duchenne F, et al. 2024. Macroevolution of the plant–hummingbird pollination system. *Biol. Rev.*:000.

Barrett T, Dowle M, Srinivasan A, Gorecki J, Chirico M, Hocking T. 2024. data.table: Extension of `data.frame`. Available from: https://r-datatable.com

Bergamo PJ, Rech AR, Brito VLG, Sazima M. 2016. Flower colour and visitation rates of *Costus arabicus* support the ‘bee avoidance’ hypothesis for red-reflecting hummingbird-pollinated flowers. *Funct. Ecol.* 30:710–720.

868 Bischoff M, Jürgens A, Campbell DR. 2014. Floral scent in natural hybrids of *Ipomopsis*
869 (Polemoniaceae) and their parental species. *Ann. Bot.* 113:533–544.

870 Blackmon H, Adams RA. 2015. EvobiR: Tools for comparative analyses and teaching
871 evolutionary biology.

872 Burge S, Kelly E, Lonsdale D, Mutowo-Muellenet P, McAnulla C, Mitchell A, Sangrador-Vegas
873 A, Yong S-Y, Mulder N, Hunter S. 2012. Manual GO annotation of predictive protein
874 signatures: the InterPro approach to GO curation. *Database J. Biol. Databases Curation*
875 2012:bar068.

876 Byers KJRP, Bradshaw HD, Riffell JA. 2014. Three floral volatiles contribute to differential
877 pollinator attraction in monkeyflowers (*Mimulus*). *J. Exp. Biol.* 217:614–623.

878 Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and
879 simple way. *PeerJ* 6:e4958.

880 Campitelli E. 2023. ggnewscale: Multiple Fill and Colour Scales in "ggplot2." Available from:
881 https://CRAN.R-project.org/package=ggnewscale

882 Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated
883 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.

884 Castellanos MC, Wilson P, Thomson JD. 2004. 'Anti-bee' and 'pro-bird' changes during the
885 evolution of hummingbird pollination in *Penstemon* flowers. *J. Evol. Biol.* 17:876–885.

886 Chen GF, Schemske DW. 2015. Ecological differentiation and local adaptation in two sister
887 species of Neotropical *Costus* (Costaceae). *Ecology* 96:440–449.

888 Chen L, DeVries AL, Cheng C-HC. 1997. Convergent evolution of antifreeze glycoproteins in
889 Antarctic notothenioid fish and Arctic cod. *Proc. Natl. Acad. Sci.* 94:3817–3822.

890 Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly
891 using phased assembly graphs with hifiasm. *Nat. Methods* 18:170–175.

892 Collyer ML, Adams DC. 2018. RRPP: An r package for fitting linear models to high-
893 dimensional data using residual randomization. *Methods Ecol. Evol.* 9:1772–1779.

894 Dellinger AS, Paun O, Baar J, Temsch EM, Fernández-Fernández D, Schönenberger J. 2022.
895 Population structure in Neotropical plants: Integrating pollination biology, topography
896 and climatic niches. *Mol. Ecol.* 31:2264–2280.

897 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
898 TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

899 Driver RJ. 2022. Evolution of olfactory receptors in birds. Available from:
900 https://thescholarship.ecu.edu/handle/10342/12266

901 Driver RJ, Balakrishnan CN. 2021. Highly Contiguous Genomes Improve the Understanding of
902 Avian Olfactory Receptor Repertoires. *Integr. Comp. Biol.* 61:1281–1290.

903 Eddy SR. 2011. Accelerated Profile HMM Searches. *PLOS Comput. Biol.* 7:e1002195.

904 Esfeld K, Berardi AE, Moser M, Bossolini E, Freitas L, Kuhlemeier C. 2018. Pseudogenization
905 and Resurrection of a Speciation Gene. *Curr. Biol.* 28:3776-3786.e7.

906 Farré-Armengol G, Filella I, Llusià J, Peñuelas J. 2017. β-Ocimene, a Key Floral and Foliar
907 Volatile Involved in Multiple Interactions between Plants and Other Organisms.
908 *Molecules* 22:1148.

909 Fernandes N de S, Silva FAN, Aragão FAS de, Zocolo GJ, Freitas BM. 2019. Volatile Organic
910 Compounds Role in Selective Pollinator Visits to Commercial Melon Types. *J. Agric.*
911 *Sci.* 11:p93.

912 Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation
913 sequencing data. *Bioinformatics* 28:3150–3152.

914 Gamba D, Muchhala N. 2023. Pollinator type strongly impacts gene flow within and among
915 plant populations for six Neotropical species. *Ecology* 104:e3845.

916 García-Robledo C, Erickson DL, Staines CL, Erwin TL, Kress WJ. 2013. Tropical Plant–
917 Herbivore Networks: Reconstructing Species Interactions Using DNA Barcodes. *PLOS*
918 *ONE* 8:e52967.

919 Gershenzon J, Dudareva N. 2007. The function of terpene natural products in the natural world.
920 *Nat. Chem. Biol.* 3:408–414.

921 Goldsmith KM, Goldsmith TH. 1982. Sense of Smell in the Black-Chinned Hummingbird. *The*
922 *Condor* 84:237.

923 Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M,
924 Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with
925 the Blast2GO suite. *Nucleic Acids Res.* 36:3420–3435.

926 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
927 Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-
928 Seq data without a reference genome. *Nat. Biotechnol.* 29:644–652.

929 Graham AM, Jamison JM, Bustos M, Cournoyer C, Michaels A, Presnell JS, Richter R, Crocker
930 DE, Fustukjian A, Hunter ME, et al. 2023. Reduction of Paraoxonase Expression
931 Followed by Inactivation across Independent Semiaquatic Mammals Suggests Stepwise
932 Path to Pseudogenization. *Mol. Biol. Evol.* 40:msad104.

933 Grant KA, Grant V. 1968. Hummingbirds and their flowers. New York: Columbia University
934 Press

935 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D,
936 Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq
937 using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8:1494–
938 1512.

939 Harenčár J, Vargas OM, Escalona M, Schemske DW, Kay KM. 2023. Genome assemblies and
940 comparison of two neotropical spiral gingers: *Costus pulverulentus* and *C. lasius*. *J.*
941 *Hered.*:esad018.

942 Hervé M. 2021. RVAideMemoire: Testing and Plotting Procedures for Biostatistics. R package
943 version 0.9-81. Available from: https://CRAN.R-project.org/package=RVAideMemoire

944 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the
945 Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35:518–522.

946 Janzen DH. 1971. Euglossine Bees as Long-Distance Pollinators of Tropical Plants. *Science*
947 171:203–205.

948 Janzen DH. 1981. Bee Arrival at Two Costa Rican Female Catasetum Orchid Inflorescences, and
949 a Hypothesis on Euglossine Population Structure. *Oikos* 36:177–183.

950 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder:
951 fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14:587–589.

952 Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
953 Improvements in Performance and Usability. *Mol. Biol. Evol.* 30:772–780.

954 Kay KM, Grossenbacher DL. 2022. Evolutionary convergence on hummingbird pollination in
955 Neotropical *Costus* provides insight into the causes of pollinator shifts. *New Phytol.*
956 236:1572–1583.

957 Kay KM, Schemske DW. 2003. Pollinator Assemblages and Visitation Rates for 11 Species of
958 Neotropical *Costus* (Costaceae). *Biotropica* 35:198–207.

959 Knudsen JT, Tollsten L, Groth I, Bergstöm G, Raguso RA. 2004. Trends in floral scent
960 chemistry in pollination syndromes: floral scent composition in hummingbird-pollinated
961 taxa. *Bot. J. Linn. Soc.* 146:191–199.

962 Krauss SL, Phillips RD, Karron JD, Johnson SD, Roberts DG, Hopper SD. 2017. Novel
963 Consequences of Bird Pollination for Plant Mating. *Trends Plant Sci.* 22:395–410.

964 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*
965 9:357–359.

966 Lanier ER, Andersen TB, Hamberger B. 2023. Plant terpene specialized metabolism: complex
967 networks or simple linear pathways? *Plant J.* 114:1178–1201.

968 Larue A-AC, Raguso RA, Junker RR. 2016. Experimental manipulation of floral scent bouquets
969 restructures flower–visitor interactions in the field. *J. Anim. Ecol.* 85:396–408.

970 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–
971 3100.

972 Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for
973 assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.* 30:923–930.

974 Liu JW, Milet-Pinheiro P, Gerlach G, Ayasse M, Nunes CEP, Alves-Dos-Santos I, Ramírez SR.
975 2024. Macroevolution of floral scent chemistry across radiations of male euglossine bee-
976 pollinated plants. *Evolution* 78:98–110.

977 Losos JB. 2011. Convergence, adaptation, and constraint. *Evolution* 65:1827–1840.

978 Maas PJM. 1972. Costoideae (Zingiberaceae). *Flora Neotropica* 8:1–139.

979 Maas PJM. 1977. Renealmia (Zingiberaceae - Zingiberoideae) Costoideae (Additions)
980 (Zingiberaceae). *Flora Neotropica* 18:1–218.

981 Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel and
982 Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for
983 Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* 38:4647–4654.

984 Marshall CR, Raff EC, Raff RA. 1994. Dollo's law and the death and resurrection of genes.
985 *Proc. Natl. Acad. Sci. U. S. A.* 91:12283–12287.

986 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.
987 *EMBnet.journal* 17:10–12.

988 Mendes FK, Vanderpool D, Fulton B, Hahn MW. 2020. CAFE 5 models variation in
989 evolutionary rates among gene families. *Bioinformatics* 36:5516–5518.

990 Mishra S, Smith ML, Hahn MW. 2024. reconcILS: A gene tree-species tree reconciliation
991 algorithm that allows for incomplete lineage sorting. :2023.11.03.565544. Available
992 from: https://www.biorxiv.org/content/10.1101/2023.11.03.565544v2

993 Moran D, Softley R, Warrant EJ. 2015. The energetic cost of vision and the evolution of eyeless
994 Mexican cavefish. *Sci. Adv.* 1:e1500363.

995 Müller K, Wickham H. 2022. tibble: Simple Data Frames.

996 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective
997 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.*
998 32:268–274.

Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara R, Simpson G, Solymos P, et al. 2020. vegan: Community Ecology Package. R package version 2.5-7. Available from: https://CRAN.R-project.org/package=vegan

Opedal ØH, Falahati-Anbaran M, Albertsen E, Armbruster WS, Pérez-Barrales R, Stenøien HK, Pélabon C. 2017. Euglossine bees mediate only limited long-distance gene flow in a tropical vine. *New Phytol.* 213:1898–1908.

Orr HA. 2005. The Probability of Parallel Evolution. *Evolution* 59:216–220.

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

Pedersen TL, Crameri F. 2023. scico: Colour Palettes Based on the Scientific Colour-Maps. Available from: https://CRAN.R-project.org/package=scico

Peng F, Byers KJRP, Bradshaw Jr. HD. 2017. Less is more: Independent loss-of-function OCIMENE SYNTHASE alleles parallel pollination syndrome diversification in monkeyflowers (*Mimulus*). *Am. J. Bot.* 104:1055–1059.

Petrén H, Köllner TG, Junker RR. 2023. Quantifying chemodiversity considering biochemical and structural properties of compounds with the R package chemodiv. *New Phytol.* 237:2478–2492.

Pichersky E, Raguso RA. 2018. Why do plants produce so many terpenoid compounds? *New Phytol.* 220:692–702.

Powers J, Eisen K, Campbell D, Raguso R. 2023. bouquet: Filter and Augment Biological GC-MS Data. Available from: https://github.com/jmpowers/bouquet

R Core Team. 2023. R: A language and environment for statistical computing. Available from: http://www.R-project.org

Raguso RA. 2016. Plant Evolution: Repeated Loss of Floral Scent — A Path of Least Resistance? *Curr. Biol.* 26:R1282–R1285.

Ramos SE, Schiestl FP. 2019. Rapid plant evolution driven by the interaction of pollination and herbivory. *Science* 364:193–196.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3:217–223.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:R25.

1032 Rosas-Guerrero V, Aguilar R, Martén-Rodríguez S, Ashworth L, Lopezaraiza-Mikel M, Bastida
1033 JM, Quesada M. 2014. A quantitative review of pollination syndromes: do floral traits
1034 predict effective pollinators? *Ecol. Lett.* 17:388–400.

1035 Sackton TB, Clark N. 2019. Convergent evolution in the genomics era: new insights and
1036 directions. *Philos. Trans. R. Soc. B Biol. Sci.* 374:20190102.

1037 Sas C, Müller F, Kappel C, Kent TV, Wright SI, Hilker M, Lenhard M. 2016. Repeated
1038 Inactivation of the First Committed Enzyme Underlies the Loss of Benzaldehyde
1039 Emission after the Selfing Transition in *Capsella*. *Curr. Biol.* 26:3313–3319.

1040 Schemske DW, Bradshaw HD. 1999. Pollinator preference and the evolution of floral traits in
1041 monkeyflowers (*Mimulus*). *Proc. Natl. Acad. Sci.* 96:11910–11915.

1042 Shumate A, Salzberg SL. 2021. Liftoff: accurate mapping of gene annotations. *Bioinformatics*
1043 37:1639–1643.

1044 Sifuentes-Romero I, Aviles AM, Carter JL, Chan-Pong A, Clarke A, Crotty P, Engstrom D,
1045 Meka P, Perez A, Perez R, et al. 2023. Trait Loss in Evolution: What Cavefish Have
1046 Taught Us about Mechanisms Underlying Eye Regression. *Integr. Comp. Biol.* 63:393–
1047 406.

1048 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO:
1049 assessing genome assembly and annotation completeness with single-copy orthologs.
1050 *Bioinformatics* 31:3210–3212.

1051 Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence
1052 comparison. *BMC Bioinformatics* 6:31.

1053 Smid M, Coebergh van den Braak RRJ, van de Werken HJG, van Riet J, van Galen A, de Weerd
1054 V, van der Vlugt-Daane M, Bril SI, Lalmahomed ZS, Kloosterman WP, et al. 2018. Gene
1055 length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data
1056 performs similarly in intersample analyses while improving intrasample comparisons.
1057 *BMC Bioinformatics* 19:236.

1058 Tarazona S, Furió-Tarí P, Turrà D, Pietro AD, Nueda MJ, Ferrer A, Conesa A. 2015. Data
1059 quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc
1060 package. *Nucleic Acids Res.* 43:e140.

1061 Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A. 2011. Differential expression in
1062 RNA-seq: a matter of depth. *Genome Res.* 21:2213–2223.

1063 Tholl D. 2006. Terpene synthases and the regulation, diversity and biological roles of terpene
1064 metabolism. *Curr. Opin. Plant Biol.* 9:297–304.

1065 Thomson JD, Wilson P. 2008. Explaining Evolutionary Shifts between Bee and Hummingbird
1066 Pollination: Convergence, Divergence, and Directionality. *Int. J. Plant Sci.* 169:23–38.

1067 Tripp EA, Manos PS. 2008. Is floral specialization an evolutionary dead-end? Pollination system
1068 transitions in Ruellia (Acanthaceae). *Evol. Int. J. Org. Evol.* 62:1712–1737.

1069 Tsuchimatsu T, Fujii S. 2022. The selfing syndrome and beyond: diverse evolutionary
1070 consequences of mating system transitions in plants. *Philos. Trans. R. Soc. B Biol. Sci.*
1071 377:20200510.

1072 Valderrama E, Landis JB, Skinner D, Maas PJM, Maas-van de Kramer H, André T, Grunder N,
1073 Sass C, Pinilla-Vargas M, Guan CJ, et al. 2022. The genetic mechanisms underlying the
1074 convergent evolution of pollination syndromes in the Neotropical radiation of *Costus* L.
1075 *Front. Plant Sci.* 13:874322.

1076 Vargas OM, Goldston B, Grossenbacher DL, Kay KM. 2020. Patterns of speciation are similar
1077 across mountainous and lowland regions for a Neotropical plant radiation (Costaceae:
1078 *Costus*). *Evolution* 74:2644–2661.

1079 Wessinger CA. 2024. How the switch to hummingbird pollination has greatly contributed to our
1080 understanding of evolutionary processes. *New Phytol.* 241:59–64.

1081 Wickham H. 2007. Reshaping Data with the reshape Package. *J. Stat. Softw.* 21:1–20.

1082 Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York
1083 Available from: http://ggplot2.org

1084 Wickham H, François R, Henry L, Müller K. 2021. dplyr: A Grammar of Data Manipulation. R
1085 package version 1.0.7. Available from: https://CRAN.R-project.org/package=dplyr

1086 Wickham H, Hester J, Bryan J. 2023. readr: Read Rectangular Text Data. Available from:
1087 https://CRAN.R-project.org/package=readr

1088 Wikelski M, Moxley J, Eaton-Mordas A, López-Uribe MM, Holland R, Moskowitz D, Roubik
1089 DW, Kays R. 2010. Large-Range Movements of Neotropical Orchid Bees Observed via
1090 Radio Telemetry. *PLOS ONE* 5:e10738.

1091 Wilke CO. 2020. cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2."
1092 Available from: https://CRAN.R-project.org/package=cowplot

1093 Workman R, Fedak R, Kilburn D, Hao S, Liu K, Timp W. 2019. High Molecular Weight DNA
1094 Extraction from Recalcitrant Plant Species for Third Generation Sequencing. Available
1095 from: https://www.protocols.io/view/high-molecular-weight-dna-extraction-from-
1096 recalcit-4vbgw2n

1097 Wozniak N, Sartori K, Kappel C, Zhao L, Erban A, Fehrle I, Jantzen F, Orsucci M, Rosa S,
1098 Lenhard M, et al. 2022. Convergence and molecular evolution of floral fragrance after
1099 independent transitions to self–fertilization. :2022.10.04.510758. Available from:
1100 https://www.biorxiv.org/content/10.1101/2022.10.04.510758v1

1101 Xie Y. 2023. knitr: A General-Purpose Package for Dynamic Report Generation in R. Available
1102 from: https://yihui.org/knitr/.

1103 Xu S. 2022. ggstar: Multiple Geometric Shape Point Layer for "ggplot2." Available from:
1104 https://CRAN.R-project.org/package=ggstar

1105 Xu S, Dai Z, Guo P, Fu X, Liu S, Zhou L, Tang W, Feng T, Chen M, Zhan L, et al. 2021.
1106 ggtreeExtra: Compact Visualization of Richly Annotated Phylogenetic Data. *Mol. Biol.*
1107 *Evol.* 38:4039–4042.

1108 Yost JM, Kay KM. 2009. The evolution of postpollination reproductive isolation in *Costus*. *Sex.*
1109 *Plant Reprod.* 22:247–255.

1110 Yu G. 2022. Data Integration, Manipulation and Visualization of Phylogenetic Trees. Chapman
1111 & Hall Available from: https://www.routledge.com/Data-Integration-Manipulation-and-
1112 Visualization-of-Phylogenetic-Trees/Yu/p/book/9781032233574

1113 Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an r package for visualization and
1114 annotation of phylogenetic trees with their covariates and other associated data. *Methods*
1115 *Ecol. Evol.* 8:28–36.

1116 Zdobnov EM, Kuznetsov D, Tegenfeldt F, Manni M, Berkeley M, Kriventseva EV. 2021.
1117 OrthoDB in 2020: evolutionary and functional annotations of orthologs. *Nucleic Acids*
1118 *Res.* 49:D389–D393.

1119 Zhou Y, Abbas F, He J, Yan F, Wang Q, Yu Y, Yu R, Fan Y. 2022. Floral volatile chemical
1120 diversity in *Hedychium* F1 hybrid population. *Ind. Crops Prod.* 184:115032.

1121 Zu P, Schiestl FP, Gervasi D, Li X, Runcie D, Guillaume F. 2020. Floral signals evolve in a
1122 predictable way under artificial and pollinator selection in *Brassica rapa*. *BMC Evol.*
1123 *Biol.* 20:127.

1124 Zytnicki M. 2017. mmquant: how to count multi-mapping reads? *BMC Bioinformatics* 18:411.

1125