Title

Diversification of Linanthus in western North America: Integrating genomics and chemical ecology to study the evolution of desert plants

Author

Anghel, Ioana

Publication Date

2024

UNIVERSITY OF CALIFORNIA

Los Angeles

Diversification of *Linanthus* in western North America: Integrating genomics and chemical ecology to study the evolution of desert plants

A dissertation submitted in partial satisfaction of the

requirements for the degree

Doctor of Philosophy in Biology

by

Ioana Gabriela Anghel

2024

ABSTRACT OF THE DISSERTATION

Diversification of *Linanthus* in western North America: Integrating genomics and chemical ecology to study the evolution of desert plants

By

Ioana Gabriela Anghel

Doctor of Philosophy in Biology

University of California, Los Angeles, 2024

Professor Felipe Zapata Hoyos, Chair

Speciation usually occurs in geographic isolation, with lineages accumulating differences through time without exchanging genes. Yet how have closely related taxa speciated with potential for gene flow? I address this question in *Linanthus*, a phenotypically diverse flowering plant genus that rapidly diversified in a small geographic space. I integrated phylogenomics, trait evolution, chemical ecology, and landscape genomics to investigate patterns of *Linanthus* diversification. In Chapter 1, I reconstructed the evolutionary history across all species of *Linanthus*, including an average of seven individuals per species with a third of samples co-occurring with congeners. I found that most species were monophyletic, despite potential gene flow. The perennial and annual night blooming clades had unresolved species relationships,

suggesting rapid speciation or cryptic diversity. Perenniality evolved once, night blooming three times, and flower color polymorphisms were likely ancestral. Classic allopatric speciation was likely not the only geographic mode of speciation across the genus, with young species overlapping in range. In Chapter 2, I investigated patterns of floral scent variation across populations and species of *Linanthus*. Scent is a complex and cryptic dimension of diversity that flowering plants use to ensure reproductive success through pollinator attraction. I found unusually high variation in scent profiles of *Linanthus* species. Within clades of *Linanthus*, species scent profiles were differentiated, suggesting fragrance is a mechanism of reproductive isolation in closely related species with range overlap. In Chapter 3, we assembled the reference genome of *Linanthus parryae*, which serves as a genomic resource for Chapter 1 and 4, and for understanding how biological variation originates and is maintained across *Linanthus*. Lastly, in Chapter 4, I used whole genome sequencing to assess the relative role of natural selection and genetic drift maintaining genetic variation across populations of *L. parryae*. I found evidence that both isolation by distance and by environment structure genetic divergence in this species. Precipitation and temperature variability had the highest contribution to genetic differentiation associated with climate, reflecting previous findings that fluctuating rainfall patterns maintain phenotypic variation. Together, this work provides integrative evidence to understanding how *Linanthus* species and their phenotypic diversity has originated and persists in nature.

The dissertation of Ioana Gabriela Anghel is approved.

Victoria Sork

Nathan Kraft

Kathleen Kay

Felipe Zapata Hoyos, Committee Chair

University of California, Los Angeles,

2024

Table of Contents

ABSTRACT OF THE DISSERTATION ..... **ii**

Table of Contents ..... **v**

Acknowledgements ..... **viii**

List of Figures ..... **xiii**

List of Tables ..... **xv**

Supplementary Materials ..... **xvi**

Curriculum Vitae ..... **xvii**

When the sand blossoms: phylogeny, trait evolution, and geography of speciation in *Linanthus* (Polemoniaceae) **1**

Abstract *1*

Introduction *2*

Materials and Methods *7*

Data collection and processing 7

DNA extraction, library preparation and sequencing 9

Assembly 11

Phylogenetic inference and species trees 13

Population structure 14

Ancestral state reconstruction 16

Geography of speciation 17

Results *19*

Data processing 19

Phylogenetic inference 20

Early diverging grade 21

Perennial clade 21

Coastal California and Baja California annuals clade 23

Desert annuals clade 23

Annual night blooming clade 24

Trait evolution and ancestral states 25

Geography of speciation 26

Discussion *26*

Phylogenetic relationships are well-resolved and most species are monophyletic 26

Perennials evolved from annuals 28

Night blooming evolved three times 29

Night blooming is rare and may be a mechanism of allochronic speciation 29

Color polymorphism may be ancestral 32

No predominant mode of speciation, but range overlap in sister species is indicative of parapatric speciation 36

Conclusion ... *38*
Acknowledgements ... 39
Author contributions ... 40
Data availability statement ... 40
Figures ... *41*
Appendix ... *50*
References ... *51*
***Linanthus* floral scent is highly variable yet differentiates species ... 62**
Abstract ... *62*
Introduction ... *63*
Methods ... *67*
Floral volatile collection ... 67
Gas-chromatography/mass spectrometry (GC/MS) analysis ... 68
Statistical analysis ... 69
Results ... *71*
Discussion ... *73*
Figures ... *78*
Figure 2.1 ... 78
Figure 2.2 ... 79
Figure 2.3 ... 80
Figure 2.4 ... 82
Figure S2.1 ... 83
References ... *83*
Reference genome of the color polymorphic desert annual plant sandblossoms, *Linanthus parryae* ... 88
Landscape genomics of *Linanthus parryae* reveals both isolation by distance and by environment and high genetic structure across its range ... 99
Abstract ... *99*
Introduction ... *100*
Methods ... *106*
Sampling ... 106
DNA Extraction and sequencing ... 106
Data filtering ... 106
Environmental data processing ... 108
Genetic structure ... 108
Isolation by distance versus isolation by environment ... 109
Results ... *110*
Genetic structure ... 110
Climatic variation across the landscape ... 111
Isolation by distance versus isolation by environment ... 112
Predicting adaptive genetic variation ... 112
Discussion ... *113*
Genetic structure ... 115

Isolation by environment 117
Isolation by distance 119
Caveats 120
Conclusion *120*
Acknowledgments *121*
Figures *122*
Figure 4.1. 122
Figure 4.2 123
Figure 4.3 125
Figure 4.4 125
Figure 4.5 126
Tables *126*
Table 4.1 126
Table 4.2 128
Table 4.3 129
Appendix *130*
Figure S4.1 130
References *130*

Acknowledgements

I am grateful to my advisor, Felipe Zapata, for investing his time and focus in helping me through the many challenges of creating, executing and writing a dissertation. Your support changed the course of my life by showing me that enthusiasm about a topic, and specifically plants, can sustain me through the ups and downs of pursuing the dream of getting a PhD. You are a fountain of different types of expertise, and I feel very lucky to have learned from you. Thank you for trusting me to be your first graduate student.

I am very grateful for all the mentors who supported me in my journey to applying to grad school. As a non-traditional graduate student who did not major in biology in my undergraduate studies, I was lucky to find mentors who encouraged me to pursue this path and who introduced me to opportunities to make it a reality. I am immensely grateful for Chelsea Specht's guidance, who supported me early in my botanical pursuit and gave me one of my first immersive opportunity to experience how research is done. I am grateful for Anna Larsen, who introduced me to the Jepson Herbarium Workshops and its wonderful universe of botany nerds, and who has been an enthusiastic supporter in my exploration of various botanical careers. I thank Bob Patterson, whose field botany class first catalyzed my passion for botany, and whose work in *Linanthus* inspired me to pursue working in this genus.

Thank you to my committee members, Kathleen Kay, Nathan Kraft, and Victoria Sork. Each of you has pushed me to become a better thinker and writer, and I am immensely grateful for your investment in my development as a scientist. I thank my collaborators, Isaac Lichter-Marck, Lydia Smith, Santiago Ramirez, Jasen Liu, Sarah Jacobs, Sonal Singhal, for your support in

teaching me methods and protocols, discussing ideas and concepts, helping with manuscript edits and so much more.

I am thankful to my lab mates, Sarah Jacobs, Sara Pedraza, Isaac Lichter-Marck, Claudia Henriquez, Lindsay Reedy, Christian Henry, Irene Liao, Sophie Winitsky, for creating a supportive community where we could discuss ideas, improve presentations and manuscripts, and be each other's emotional support system.

I thank my UCLA EEB Department graduate students for their support in having a similarly marking experience, especially Chris, Marcel, Gaurav, Maddi, and Marvin. I am grateful for my compatriot graduate student, Martha, for your support, life enthusiasm, and teaching me confidence by example. I especially thank Mary, my sister in suffering and celebration of all things PhD.

Thank you to my family. My partner in crime Chris, for joining me on this crazy ride of moving to a new city and for always believing I could when I could not. Your support is the reason why we have made it to the finish line. My parents, for immigrating when my sister and I were teenagers. Your courage in uprooting your lives for better opportunities for our family is the reason we can pursue our dreams. I am grateful and humbled by the great personal and career sacrifices you made that paved the way for me to work in a career that I love. Thank you, my baby, for making it clear that a dissertation has got nothing on being a mother, either in terms of importance, meaning and difficulty. You are my treasure.

There were so many others who touched this project along the way that I have likely forgotten to mention. As I take stock of all the people who have supported me through this

process, I am overwhelmed with gratitude and the feeling that we have formed a community that has changed me for the better in the way I think and act in the world.

Above all, I thank the plants and the process of evolution for allowing me to be witness to their wonder. I feel most alive and connected to what matters when I am surrounded by plants in their natural habitat. It has been a great privilege to be able to spend so much time in wild areas, in the places we evolved to live as an animal species. I pledge to help protect these places so that our future generations can find connection and meaning in these sacred natural places.

Chapter 1 is a version of the manuscript that has been accepted for publication in the *American Journal of Botany* with authors Ioana Anghel, Lydia L. Smith, Isaac H. Lichter-Marck, and Felipe Zapata.

Author contributions: FZ contributed to study design and to the editing and revision of the manuscript. LLS trained IA in lab work and prepared sequencing libraries. ILM assembled the target capture sequence data and advised on analyses.

Funding: Funding to IA was provided by the UCLA Ecology and Evolutionary Biology Department, the UCLA La Kretz Center and Stunt Ranch “Graduate Research Grant”, the Botanical Society of America “Bill Dahl Graduate Student Research Award”, the American Society of Plant Taxonomists “Graduate Student Research Grant”, the California Native Plant Society “Hardman Native Plant Research Award” and FZ’s start-up funding. ILM. was funded by NSF DBI #2209393.

Chapter 2 is in preparation for publication with authors Ioana Anghel, Felipe Zapata, Santiago Ramirez, and Jasen Liu.

Author contributions: SR and JL provided training on techniques, guidance on analyses and results interpretation. FZ provided input on study design, data interpretation and manuscript editing. **Funding:** Funding was provided by the UCLA Ecology and Evolutionary Biology Department, the UCLA La Kretz Center and Stunt Ranch "Graduate Research Grant", and Society of Systematic Biologists Graduate Student Research Award.

Chapter 3 was published in the *Journal of Heredity*, and reprinted here with permission from the American Genetic Association.

Anghel, I. G., Jacobs, S. J., Escalona, M., Marimuthu, M. P. A., Fairbairn, C. W., Beraut, E., Nguyen, O., Toffelmier, E., Shaffer, H. B., & Zapata, F. (2022). Reference genome of the color polymorphic desert annual plant sandblossoms, Linanthus parryae. *Journal of Heredity*, *113*, 712–721. https://doi.org/10.1093/jhered/esac052

Author contributions: SJJ contributed to fieldwork and manuscript editing. ME assembled the reference genome. MPAM did the high molecular weight extractions. CWF, EB, ON performed extractions, library prep and sequencing. ET and HBS contributed to manuscript review and comments. FZ helped design the study and edited the manuscript.

Funding: This work was done with the support of the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224].

Chapter 4 is a version of the manuscript in preparation for publication with the authors Ioana Anghel, Felipe Zapata, Sarah Jacobs, and Sonal Singhal.

Author contributions: FZ and SJ helped with study design and field work. All authors advised on analyses and results interpretation, and manuscript edits.

Funding: This work was done with the support of the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224].

List of Figures

Chapter 1: When the sand blossoms: phylogeny, trait evolution, and geography of speciation in *Linanthus* (Polemoniaceae)

Figure 1.1. *Linanthus* species floral diversity.

Figure 1.2. The phylogeny of *Linanthus* inferred using ddRAD data.

Figure 1.3. The phylogeny of *Linanthus* inferred using target capture data (Angisoperms353).

Figure 1.4. Perennial clade genetic structure plot.

Figure 1.5. Annual night bloomer clade genetic structure plot.

Figure 1.6. Phylogeny of *Linanthus* showing the distribution of three phenotypic traits: perenniality, night-blooming and corolla color polymorphism.

Figure 1.7. Stochastic character density maps for three phenotypic traits: perenniality, night-blooming and corolla color polymorphism.

Figure 1.8. Phylogenetic distance does not predict range overlap of pairs of *Linanthus* species, suggesting there is no predominant geographic mode of speciation.

Chapter 2: *Linanthus* floral scent is highly variable yet differentiates species

Figure 2.1. A non-metric multidimensional scaling (NMDS) of individual scent profiles for 13 species of *Linanthus* with multiple individuals per species.

Figure 2.2. Canonical analyses of principal coordinates (CAP) of scent variation constrained by species.

Figure 2.3. Canonical analysis of principal coordinates (CAP) of scent variation constrained by species, by clade.

Figure 2.4. Hierarchical clustering analysis of compound emissions visualized with a dendrogram of compounds by samples or species averages.

Figure S2.1. Phylogenetic relationships and major clades identified in Anghel et al. 2024.

Chapter 3: Reference genome of the color polymorphic desert annual plant sandblossoms, *Linanthus parryae*

Figure 3.1. *Linanthus parryae* morphology, habitat and range distribution.

Figure 3.2. Visual overview of the *Linanthus parryae* genome assembly metrics.

Chapter 4: Landscape genomics of *Linanthus parryae* reveals both isolation by distance and by environment and high genetic structure across its range

Figure 4.1. The flower color polymorphism in *Linanthus parryae*: individuals have either blue or white petals.

Figure 4.2. Map of sampling localities of the *L. parryae* individuals included in this study plotted against all occurrences of the species.

Figure 4.3. Spatial population structure of *L. parryae* with different values of genetic clusters.

Figure 4.4. GDM fitted functions, modeling the relationships between predictor variables (geographic distance and three environmental PCs) and the response variable (genetic distance).

Figure 4.5. (A) Spatial interpolation of genetic dissimilarity associated with environmental gradients predict areas of rapid genetic turnover (B) Principal Component Analysis biplot of the GDM.

Figure S4.1. Principal component analysis (PCA) of the linkage disequilibrium pruned SNP data set of L. parryae.

List of Tables

Chapter 2: *Linanthus* floral scent is highly variable yet differentiates species

Supplemental Table 2.1. Floral scent compounds by samples and by species.

Chapter 3: Reference genome of the color polymorphic desert annual plant sandblossoms, *Linanthus parryae*

Table 3.1. Assembly pipeline and software used.

Table 3.2. Sequencing and assembly statistics, and accession numbers.

Table 3.3. Classification of repeat elements generated from RepeatMasker.

Chapter 4: Landscape genomics of *Linanthus parryae* reveals both isolation by distance and by environment and high genetic structure across its range

Table 4.1. Samples included in this study, including sample identifier, location and flower color.

Table 4.2. Loadings of each WorldClim environmental layer for PC1-PC3.

Table 4.3. The Generalized Dissimilarity Model (GDM) incorporating all predictors explained 32.94% of the deviance in genetic turnover.

Supplementary Materials

Chapter 1: Available online at github.com/ioanaanghel/Linanthus_phylogeny.

Chapter 2: Ch2_Table1_FilteredCmpds_44cmpds_13sp_76spls.xlsx (File Format .xlsx)

Chapter 3: Available online from the Journal of Heredity:

https://academic.oup.com/jhered/article/113/6/712/6701609#supplementary-data

Curriculum Vitae

EDUCATION

Ph.D. Candidate, Ecology and Evolutionary Biology — 2020
University of California, Los Angeles

B.A. Politics, Psychology — 2007 – 2010
University of California, Santa Cruz

GRANTS AND FELLOWSHIPS

Collegium of University Teaching Fellow, University of California, Los Angeles	2023 – 2024
California Native Plant Society, LA/SMM Chapter Travel Award	2022
Vavra Fellowship, UCLA Ecology and Evolutionary Biology	2017 – 2022
UCLA Summer Mentored Research Fellowship	2021
Society of Systematic Biologists Graduate Student Research Award	2020
Botanical Society of America Bill Dahl Graduate Student Research Award	2020
American Society of Plant Taxonomists Graduate Student Research Grant	2020
California Conservation Genomics Project (CCGP), Role: Co-PI	2020
UCLA Ecology and Evolutionary Biology Departmental Award	2019 – 2023
UCLA La Kretz Center and Stunt Ranch Graduate Research Grant	2019, 2020
California Native Plant Society - The Hardman Native Plant Research Award	2019
Graduate Division Travel Grant, UCLA	2018

PUBLICATIONS

1. **Anghel, I.G.**, Smith L.L, Lichter-Marck, I.H., and Zapata, F. Phylogenomics of the North American Desert Radiation *Linanthus* (Polemoniaceae) Reveals Mixed Trait Lability and No Single Geographic Mode of Speciation (in press). doi.org/10.1101/2024.06.13.598867
2. Chanderbali, A.C, Dervinis, C., **Anghel, I.G.**, Soltis, D.E., Soltis, P.S., and Zapata F. (2023) Draft genome assemblies for two species of *Escallonia* (Escalloniales). BMC Genomic Data, 25(1). doi.org/10.1186/s12863-023-01186-7
3. Medeiros, C., Henry, C., Trueba, S., **Anghel, I.G.**, Dannet Diaz de Leon Guerrero, S., Pivovaroff, A., Fletcher, L., John, G., Lutz, J., Mendez Alonzo, R., and Sack, L. (2023) Predicting plant species climate niches on the basis of mechanistic traits. Functional Ecology, 37(11): 2786-2808. doi.org/10.1111/1365-2435.14422
4. **Anghel, I.G.**, Jacobs, S.J., Escalona, M., Marimuthu, M.P.A., Fairbairn, C.W., Beraut, E., Nguyen, O., Toffelmier, E., Shaffer, H.B., and Zapata, F. (2022) Reference genome of the color polymorphic desert annual plant sandblossoms, *Linanthus parryae*. Journal of Heredity, (113), 712–721. doi.org/10.1093/jhered/esac052
5. Jacobs, S.J., Grundler, M.C., Henriquez, C.L., Diaz, R., **Anghel, I.G.**, Lavandero, N., and Zapata, F. Insights into the phylogeny of the Escalloniales and other Campanulids using whole plastome data. (in manuscript)

PRESENTATIONS & PANELS

Botany Conference 2024, Grand Rapids, MI — 2024
"Phylogenomics of the North American Desert Radiation *Linanthus* (Polemoniaceae)"

EEB Department Research Day, UCLA, Los Angeles, CA — 2024
"The evolution of species and traits in *Linanthus*"

California Botanical Society, Botany Speaker Series — 2023
"Evolutionary relationships and phenotypic patterns in Linanthus"

California Native Plant Society, 2022 Conference, San Jose, CA — 2022
"Deciphering evolutionary relationships in *Linanthus"*

Guest Speaker for EEB103 course *Plant Diversity and Evolution*, UCLA, CA — 2022
"People and Plants: Our Common Evolution and History"

California Botanic Society, 28th Graduate Student Symposium, Virtual 2021
"Close relatives in sympatry: Is there evidence of gene flow between species of *Linanthus*?
Botany Conference 2018, Rochester, MN 2018
Poster "Mapping phenotypes: Patterns in the distribution of a polymorphic species"
Mildred E. Mathias Botanical Garden, UCLA, CA 2017
"Botany of the Holidays – Mistletoe"
Mildred E. Mathias Botanical Garden, UCLA, CA 2019
"Darwin's Birthday Walk"
UCLA Academic Advancement Program, UCLA, CA 2019
"Alternatives to PreMed: Research and Graduate Opportunities in STEM"

RESEARCH EXPERIENCE

University of California, Los Angeles – graduate student, Zapata Lab 2017-2024
University of California, Berkeley – student researcher, Specht Lab 2016 – 2017
California Academy of Science – research volunteer, Nagalingum Lab 2016
San Francisco State University – research volunteer, Parker Lab 2016

TEACHING & MENTORING

Collegium of University Teaching Fellow, University of California, Los Angeles 2023 – 2024
- Developed and taught a seminar titled *Plants and Humans, an Ancient Connection*

Teaching Assistant, University of California, Los Angeles 2018 – 2023
- *Introduction to Ecology and Behavior Laboratory* 2023
- *Evolution* 2021, 2020
- *Field Biology Quarter* 2019
- *Plant Diversity and Evolution* 2020, 2018
- *Conservation Biology* 2019, 2018
- *Biology of Superheroes* 2019

Laboratory & Field Mentoring, University of California, Los Angeles 2017-2024
Near-Peer Mentor for Preparing Leaders and Nurturing Tomorrow's Scientists (PLANTS) program 2024

SERVICE

Departmental Awards Committee Member, UCLA EEB Department 2023-2024
Graduate Student Representative Board Member, Botanical Society of America 2021-2023
Research Award Application Reviewer, UCLA La Kretz Center/Stunt Ranch 2022-2023
Booth Volunteer, Exploring Your Universe, UCLA 2023, 2024
Reviewer, Madroño Journal 2022
Professional Development Committee Member, UCLA EEB Department 2020-2022
Co-organizer 28th Graduate Student Symposium, California Botanical Society 2021
Student Advisor, California Native Plant Society 2019-2020
PlantingScience Liaison, Botanical Society of America 2019
Graduate Student Mental Health Working Group, UCLA EEB Department 2019

PROFESSIONAL EXPERIENCE

Conservation and Land Management Intern, Chicago Botanical Garden 2017
Biological Science Technician (Plants), US Forest Service, Flagstaff, AZ 2016

Chapter 1

When the sand blossoms: phylogeny, trait evolution, and geography of speciation in *Linanthus* (Polemoniaceae)

Abstract

Understanding how plants have successfully diversified in novel environments is a central question in evolutionary biology. *Linanthus* (Polemoniaceae) occurs in arid desert and mediterranean areas of Western North America and exhibits extensive floral trait variation, multiple color polymorphisms, differences in blooming time, and variation in life history strategies. Here, we reconstruct the evolutionary history of this group. We generated restriction-site associated (ddRAD) sequences for 180 individuals and target capture (TC) sequences for 63 individuals, with complete species sampling. Using maximum likelihood and pseudo-coalescent approaches, we inferred phylogenies of *Linanthus* and used these phylogenies to model the evolution of phenotypic traits and investigate the geographic speciation history of this genus. More recent relationships are consistent and well supported with both ddRAD and TC data. Most species are monophyletic despite extensive local sympatry and range overlap, suggesting strong isolating barriers. The non-monophyly of the night-blooming and perennial species is possibly due to rapid speciation or issues with current species delimitation. Perenniality likely evolved from annuality, a rare shift in angiosperms. Night blooming evolved three times independently. Flower color polymorphism is an evolutionarily labile trait and is likely ancestral. No single geographic mode of speciation characterizes the

radiation but most species overlap in range, suggesting they evolved in parapatry. Our results illustrate the complexity of phylogenetic inference for recent radiations, even with multiple sources of genomic data and extensive sampling. This analysis provides a foundation to understand aridity adaptations, such as evolution of flower color polymorphisms, night blooming, and perenniality, as well as speciation mechanisms.

Introduction

Desert biomes make up 17% of Earth's land surface (Millennium Ecosystem Assessment 2005), and appear lifeless for most of the year when lack of precipitation and high temperatures hamper plant survival and growth. However, the desert comes alive in an explosion of color after it rains, when dormant seeds germinate and flowers form thick brushstrokes of pigment. This uncommon and brief outburst of life is made possible by the annual angiosperms, which await suitable environmental cues to germinate in mass. In California, deserts make up 38% of the area (Mooney and Zavaleta 2016), with annuals making up a large part of the species diversity at 52% of total plant species in these deserts (Calflora 2023). This overrepresentation of annuals in the desert and their varied adaptations make them ideal organisms to study the patterns and processes of diversification in this harsh environment.

The deserts of California and their diversity of annual species seem to be young on a geologic time scale, emerging less than 2 million years ago (Thorne 1986). Recent meta-analyses of California plant diversity support that plants have recently diversified in this geographic area.

For instance, Kraft et al. (2010) showed that the California desert regions have a high concentration of young species with restricted geographic ranges, estimating that Mojave Desert plants originated since the late Miocene. Thornhill et al. (2017) found a significant concentration of short phylogenetic branches (i.e., recent diversification events) restricted to the California deserts. However, how deserts facilitate the rapid diversification of plant species remains poorly understood. Habitat heterogeneity, large ranges with isolated populations, and a broad diversity of adaptations to xeric landscapes have all been proposed as potential drivers of high rates of speciation in desert species (Stebbins 1952). The evolution of the annual habit seems to be associated with unstable environments with dry conditions and unpredictable rainfall (Friedman 2020), which is typical in deserts. Annual plants have a fast rate of evolution which is correlated with their short life cycle, isolated populations, and variable environment (Smith and Donoghue 2008; Smith and Beaulieu 2009). Deserts may promote the evolution of new annual plant species (Stebbins 1952) and provide a system to study the largely unexplored correlation between harsh environments and plant diversification (Stebbins 1952, but see Hernández-Hernández et al. 2014; Singhal et al. 2021; Lichter-Marck and Baldwin 2022).

There are few phylogenetic studies of California desert annuals. Most published studies do not include all members of the focal clades and almost never include multiple individuals per species, limiting understanding of the patterns and processes of intra and interspecific diversification. Additionally, most studies have only used a handful of loci, largely resulting in poorly resolved phylogenies (e.g., Spencer and Porter 1997; Moore and Jansen 2006; Evans et

al. 2009; Porter et al. 2010; Cacho et al. 2014; Walden et al. 2014; Azani et al. 2019; Vasile et al. 2020). More recently, a limited number of studies have used genomic approaches and more extensive taxon sampling shedding light on species-level relationships, patterns of within-species genetic variation, and potential drivers of diversification (Simpson et al. 2017; Mabry and Simpson 2018; Lichter-Marck et al. 2020; Pearman et al. 2021; Rose et al. 2021; Singhal et al. 2021). Given that California deserts are young hotspots of biodiversity (Kraft et al. 2010), well-sampled groups are needed to assess the monophyly of lineages, elucidate species relationships, and untangle complex evolutionary patterns, typical of recent radiations. Thorough phylogenetic reconstructions are especially worthwhile in geographic regions where species overlap extensively in their geographic range with ample opportunities for gene flow. Such robust phylogenetic studies could inform our understanding of the evolution of traits, adaptations to extreme environments, and whether aridity can be a stimulus to evolution (Stebbins 1952).

The genus *Linanthus* Benth. (Polemoniaceae) is an ideal system to study species diversification in the heterogeneous arid environments of southwestern North America. Half of the currently recognized species co-occur in a geologic transition zone characterized by exceptional plant endemism and environmental variation in Southern California (Kraft et al. 2010). Nineteen out of 25 species overlap in geographic range, and at least fourteen species pairs co-flower and co-occur at a local scale (pers. obs.). Reproducing in a span of a few weeks in the spring, sympatric *Linanthus* species likely have strong reproductive isolation mechanisms to maintain their genetic

and phenotypic integrity. *Linanthus* species also display extensive interspecific diversity in habit, blooming time, flower color, and floral scent. These traits attract a diverse suite of pollinators, including beetles, moths, butterflies, hoverflies, long tongue flies, and bees (Chess et al. 2008; Rose and Sytsma 2021; pers. obs.), which likely facilitate the reproductive differentiation of species (Fig. 1.1). In addition, seven *Linanthus* species have extremely restricted geographic ranges and are ranked as rare, threatened, or endangered in California, including *L. bellus* (A.Gray) Greene, *L. bernardinus* N.S. Fraga & D. Bell, *L. concinnus* Milliken, *L. jaegeri* (Munz) J.M. Porter & L.A. Johnson, *L. killipii* H. Mason, *L. maculatus* (Parish) Milliken, and *L. orcuttii* (Parry) Jeps. (California Native Plant Society). The extensive sympatry of species, their diversity of pollinator attraction strategies, and their asymmetric geographic ranges point to a complex speciation history that may include speciation with gene flow, micro-allopatry, parapatric speciation, and ecological isolation. However, these phenomena have not been examined in detail.

Linanthus species also show extensive intraspecific variability in petal color and nectary markings, with individual plants taking on different phenotypes in the same population or in disparate areas of the species range. Eleven of the 26 species are petal color polymorphic, with both white and colored individuals across their range (Fig. 1.1 I&J, O&P, R&S, T&U, V&X). Seven of these color polymorphic species exhibit within-population color variation, with both colors present in the same population (pers. obs.). One notable example is *L. parryae* (A. Gray) Greene, the purple and white polymorphic species at the center of the classic evolutionary debate

investigating whether natural selection or genetic drift maintains intraspecific polymorphisms. Different color morphs seem to fare better in wet or dry years, a potential adaptation to the variable desert environment (Epling and Dobzhansky 1942; Wright 1943; Schemske and Bierzychudek 2007). Another potential strategy of several *Linanthus* species to survive in arid environments is night anthesis, where flowers are only open at night and closed during the day during the flowering season. Heat can have a suite of negative fitness repercussions on flowers, including reduced pollen fertility and nectar evaporation (Borghi et al 2019). Flowers opening during the cooler nights may be a strategy to avoid the adverse effects of higher temperatures. It is not known whether these desert adaptations have facilitated the diversification of *Linanthus*.

Linanthus, as currently recognized, includes 26 species (Porter and Johnson 2000; Porter and Patterson 2015) that have likely diversified in the Miocene, and is sister to a clade that includes the genera *Leptosiphon* Benth. and *Phlox* L. (Bell and Patterson 2000). Previous phylogenetic work for 15 species using the nuclear ribosomal internal transcribed spacer (ITS) recovered low support for most relationships across the group (Bell et al. 1999). Another ITS and matK phylogeny that sampled 17 species did not provide a better resolution into the phylogenetic relationships between species (Bell and Patterson 2000). A recent study using 14 nuclear loci for 22 species (with two samples per species) inferred a well-resolved phylogeny and found that eight species were not monophyletic (Landis 2016). However, this study lacked complete

taxonomic and broad geographic sampling per species. A robust phylogeny is needed to provide a backbone for tackling inter and intraspecies evolutionary questions in this group.

In this study, we present a phylogeny of *Linanthus* with complete species sampling and broad intraspecific sampling, using two types of genomic data, double-digest restriction-site associated DNA (ddRAD) sequencing (Peterson et al. 2012) and target capture (TC) of 353 nuclear angiosperm specific genes (Johnson et al. 2019). Our ddRAD data further include multiple samples from 17 species that co-occur locally with a congener to examine the potential for gene flow between sympatric individuals from different species. We use the resulting phylogenies to examine the monophyly of species, reconstruct evolutionary relationships, and investigate patterns of floral evolution and life history shifts. We also use the ddRAD data to study the population structure within selected clades to better understand existing species delimitations for several species. Lastly, we explore the geography of speciation across *Linanthus* to determine the most likely speciation mode in this remarkable desert radiation.

Materials and Methods

Data collection and processing

We included representative samples from all currently recognized species of *Linanthus* (Moran 1977; Porter and Johnson 2000; Porter and Patterson 2015) and outgroups from *Leptosiphon, L. breviculus* (A. Gray) J.M. Porter & L.A. Johnson, *L. chrysanthus* J.M. Porter & R. Patt, *L.*

lemmonnii (A. Gray) J.M. Porter & L.A. Johnson, and *L. nuttallii* J.M. Porter & L.A. Johnson, as well as *Phlox stansburyi* (Torr.) A. Heller. To minimize potential bias introduced by misidentifications we only included names at the species level, because many of the infraspecific taxa can only be reliably identified by experts with the aid of high power magnification. We collected 83 individuals in the field across California and Nevada and sampled 125 individuals from eight herbaria across the western United States (Appendix S1.1, see Supplemental Data with this article). We stored the field-collected tissue in silica gel until storage at -20C or in liquid nitrogen until storage at -80C in the laboratory at UCLA. Some *Linanthus* species are minute; therefore, we carefully dissected all collections to ensure we selected tissue from only one individual. In total, we included samples from 50 individuals that co-occurred with a congeneric species, either observed by us in the field or when herbarium specimens came from the exact same location. We included sympatric individuals for most sympatric species pairs. Seventeen of the species sampled co-occurred with a congener, with a total of 24 combinations of species pairs. See Appendix S1.1 for complete sample metadata.

We collected genomic data using two approaches, double-digest restriction-site associated sequencing (ddRADseq) and target capture (TC). RAD sequencing generates data that has been used to successfully address questions in both population genomics and phylogenetic relationships in closely related species (Rubin et al. 2012; Eaton et al. 2016, Jacobs et al. 2021). We used a double-digest RAD approach which uses two-enzymes to better recover the same fragments of DNA across samples and reduce sequencing costs (Peterson et al. 2012). In total,

we sampled 192 individuals across the 26 currently recognized species and four outgroups in *Phlox* and *Leptosiphon*, at an average of seven individuals sampled per species, to represent the species' breadth of geographic range and phenotypic diversity. The number of individuals sampled per species ranged from 1 to 17, proportional to the species geographic range size (Appendix S1.2, S1.3).

For TC, we use the Angiosperms353 bait set (Johnson et al. 2019). This approach can sequence both exons and their flanking regions providing informative phylogenetic data at various phylogenetic scales (Larridon et al. 2020; Slimp et al. 2021). We sampled 63 individuals across all species and four *Leptosiphon* outgroups. We dropped one species (*Linanthus uncialis* (Brandegee) Moran) because DNA extraction did not generate enough DNA. We sampled a range of 1 - 5 individuals per species, with an average of 2.5 individuals sampled per species of *Linanthus.*

DNA extraction, library preparation and sequencing

We extracted genomic DNA using a modified CTAB technique (Doyle and Doyle 1987; Cullings 1992) that includes an additional incubation step to remove pectin. We prepared sequencing libraries at the Evolutionary Genomics Laboratory of the Museum of Vertebrate Zoology at UC Berkeley.

ddRAD—For the RAD-seq libraries, we followed a modified version the 3RAD protocol (Bayona-Vásquez et al. 2019) with adapters and indexing oligos provided by the Glenn lab at the University of Georgia (https://baddna.uga.edu/). We normalized a total of 192 samples to 125 ng of DNA in 10 µL volume and used this as input for a combined digestion and ligation reaction. We used XbaI andEcoRI-HF restriction enzymes to digest the genomic DNA, and the third enzyme (NheI-HF) served to digest adapter dimer produced during the ligation stages of the reaction. Adapter-ligated samples were purified using Solid-phase reversible immobilization (SPRI) beads (Rohland and Reich 2012; Jolivet and Foley 2015). We then amplified the adapter-ligated libraries with indexing oligos (Glenn lab, University of Georgia) and the KAPA HiFi PCR Kit (Roche, Indianapolis, Indiana, USA) using 16 cycles of PCR, following this with another SPRI bead cleaning. We quantified the libraries, pooled them in equimolar amounts, and then size selected fragments at a length of 375-525 base pairs (bp) using a Pippin Prep (Sage Science, Beverly, Massachusetts, USA) at the Functional Genomics Laboratory at UC Berkeley. We quantified the final size-selected library pool with the Qubit® Fluorometer with dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and checked for quality using a Bioanalyzer DNA1000 kit (Agilent Technologies, Santa Clara, California, USA). We sequenced libraries on one lane of Illumina NovaSeq SP 150PE at Vincent J. Coates Genomics Sequencing Lab (GSL) at UC Berkeley (Berkeley, California, USA) for a total of 115 Gb of data at 20x coverage.

Target capture—We fragmented DNA from 63 samples, ranging from 175–1100 ng to a target size of 350 bp using a qSonica sonicator (Newton, Connecticut, USA). We sonicated samples

with high-quality DNA for 9 minutes at 40% amplitude with a 15s on/15s off pulse. Samples that had a variety of fragment sizes were sonicated for three minutes, while those that were already highly fragmented were not sonicated at all. We cleaned the sonicated DNA and size selected with a double-sided SPRIbead cleaning using 0.525x for right-side selection and 0.675x for left-side selection. We prepared uniquely dual-indexed libraries for Illumina sequencing using the KAPA HyperPrep Kit (Roche, Indianapolis, Indiana, USA) using one-quarter reagent volumes, with a custom-designed dual indexing oligo set developed at the Functional Genomics Laboratory at UC Berkeley. For capture hybridization we enriched 4 µg pools containing equal mass of 7-9 samples each at 62°C for forty hours using the Angiosperms353 gene set (Johnson et al. 2019) ordered from Arbor Biosciences ("Angiosperms 353 v1", Catalog #308108.v5) and their myBaits Target Capture v.5 reagents and protocol (http://www.arborbiosci.com/mybaits-manual, Arbor Bioscience, Ann Arbor, Michigan, USA). We amplified enriched products with KAPA HiFi 2X HotStart ReadyMix PCR Kit for 10-12 cycles. We checked the resulting libraries for quality with an Agilent Bioanalyzer DNA1000 assay (Agilent Technologies, Santa Clara, California, USA) and quantified them with the Qubit Fluorometer with dsDNA High Sensitivity Assay Kit. We sequenced 150 paired-end reads at QB3 Genomics on a partial lane of an Illumina NovaSeq S4 for a total of 80 Gb of data.

Assembly

ddRAD— The sequencing facility demultiplexed the raw sequence reads and quality-checked them with FASTQC v0.11.9 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

We used ipyrad to assemble the ddRAD reads using the reference genome for *Linanthus parryae* (Eaton and Overcast 2020; Anghel et al. 2022). We trimmed the variable length adapter sequences using cutadapt (Martin 2011). We filtered low quality base calls with a phred Qscore of 33 and allowed up to five low quality bases per read. We set the minimum length of the reads after trimming to 35 bp, and a stricter filter for adapters at two. We set the clustering threshold within samples to 94% and between samples to 90%. We set the minimum number of samples per locus to four (see full parameters in Appendix S1.4). After filtering, we used 180 samples in the final assembly.

Target capture— Raw demultiplexed sequence reads were trimmed and quality filtered using Trimmomatic version 0.39 (Bolger et al. 2014). Specifically, we used a sliding window of 5 base pairs, cutting when the average base quality score fell below 20, removed trailing and leading low quality base pairs when their quality was below 20, and dropped reads whose length was below 36. We assembled sequences for using HybPiper verson 1.3 (Johnson et al. 2016) with the mega353 target file (McLay et al. 2021) using the -bwa option to match reads to loc.. We assembled flanking regions using 'intronerate.py' in HybPiper and identified paralogs using the 'paralog_investigator.py' script. The output of HybPiper was exons, supercontigs, paralogs and introns. Assembly of multiple contigs containing >75% of the length of the reference protein were identified as paralogs. We removed sequences found to be paralogous and used supercontigs for further analysis because they contain both exons and flanking regions, which can provide information on both the conserved gene regions and the more variable non-coding

regions. We used MAFFT version 7.505 to align supercontigs (Katoh et al. 2009). Trimal was used to clean alignments using a gap threshold of 0.9 (Capella-Gutierrez et al. 2009). We assessed the occupancy of samples in supercontig alignments with the 'phyluce_align_get_only_loci_with_min_taxa' function in PHYLUCE (Faircloth 2016) and kept only supercontigs found in all samples for further analyses.

Phylogenetic inference and species trees

ddRAD—We inferred phylogenetic trees using an alignment of concatenated filtered ddRAD loci with 98,538 variant sites. We used a maximum likelihood analysis approach in IQ-TREE with the GTR model accounting for ascertainment bias (Minh et al. 2020). The ascertainment bias model is used for single-nucleotide polymorphisms (SNPs) and other types of data that do not contain constant sites, to prevent overestimation of branch lengths (Lewis 2001). We assessed support with 1000 ultrafast bootstrap replicates. We reconstructed phylogenetic trees with a minimum of 4, 9, 18, and 36 samples per locus to assess the role of missing data in recovering topologies. Within- and between-species relationships were consistent between these data sets. Results of analyses with the data set using a minimum of 4 samples per locus are shown here. We inferred species trees with SVDQuartets (Chifman and Kubatko 2014) in PAUP* (Swofford 2002). To obtain branch lengths under the multispecies coalescent, we used the qAge command (Peng et al. 2022). As for the concatenated analyses, we estimated species trees with matrices including a minimum of 4, 9, 18, and 36 samples per locus. The species tree recovered with the matrix using a minimum of 4 samples per locus is presented in the results.

We also reassembled the data to include one sample per species to generate a species tree to use for downstream analyses where only one tip represents the species. To do this, we chose the sample within a species with the highest number of recovered loci. If the species was not monophyletic (see Results), we chose the sample with the highest number of loci that belonged to the clade including most samples for such species. We built this tree in IQ-TREE with the GTR model accounting for ascertainment bias and assessed support with 1000 ultrafast bootstrap replicates.

Target capture—We inferred gene trees using the most conservative data set of 219 supercontigs with 100% occupancy using IQ-TREE with 1000 bootstrap replicates and the GTR + F default model of substitution (Minh et al. 2020). Then, we used newick utilities to collapse poorly supported gene tree nodes (ML bootstrap support < 0.2) into polytomies (Junier and Zdobnov 2010). We used the gene trees as input for ASTRAL-III v5.7.8 (Zhang et al. 2018) to infer a species tree for all 63 taxa as well as for the 24 species by assigning taxa to species a priori. We also estimated a phylogenetic tree with a concatenated matrix of the loci with 100% occupancy for the 63 taxa in IQ-TREE.

Population structure

To assess population structure within certain clades where the taxonomic assignment pointed to non-monophyletic species (see Results), we used rmaverick v1.0.5 (Verity and Nichols 2016).

We used only the RAD data for these analyses, with an average of seven samples per species. To prepare files for input into rmaverick, we processed the ipryrad Structure file following methods outlined at https://github.com/zapata-lab/ms_rhizophora/blob/main/analysis_code (Aburto-Oropeza et al. 2021). We processed the vcf file output from ipyrad using vcftools (Danecek et al. 2011). We removed non-biallelic sites and filtered genotypes called below a certain threshold across all individuals (--max-missing 0.25, 0.50, 0.75) for all the population structure analyses. We did this to compare the effect of missing data on the genetic structure output. We also kept only the center SNP from each locus to avoid effects of linked loci.

In rmaverick, we ran the MCMC sampling every 100 steps, with 10% burn-in, 1000 sampling iterations, 20 rungs, and chose a K range from 1 to n + 1 where n is the number of taxonomic assignments in the focal clade. We ran these analyses for 0.25, 0.50, 0.75 missing data sets, and with and without the admixture model. Comparing the data sets using 0.25, 0.50, 0.75 missing data, we found little difference in the structure cluster assignments or values of K, so we present the 0.50 missing data sets here. Because rmaverick is more accurate than other population clustering programs at estimating the number of subpopulations, we chose to report only the value of K with the highest evidence shown by the largest posterior probability (Verity and Nichols 2016). In all cases, the model without admixture was a better fit to the data.

Ancestral state reconstruction

To model the evolution of phenotypic traits, we coded all species for annuality/perenniality, day/night blooming, and lack/presence of corolla lobe anthocyanin pigment polymorphism (Appendix S1.1). We gathered this phenotypic data using the Jepson eflora, the Flora of North America, the monograph of *Linanthus,* and personal observations (Patterson and Porter 2021; Patterson and Porter in prep.; Danforth 1945). While corolla lobe polymorphic species have white, yellow, pink, lavender, purple, and/or peach forms, we focused only on the anthocyanin pigments. Therefore, we coded species with white andyellow corolla lobes as lacking anthocyanin pigments, and the species with pink, lavender, and purple corolla lobes as having anthocyanin pigments (Tanaka et al. 2008). All species with anthocyanin pigments were polymorphic, with populations or individuals within populations with white corollas.

Prior to inferring ancestral states, we calibrated the phylogeny built with ddRAD data and one sample per species to relative time using Penalized Likelihood with the 'chronos' function in the R package *phytools* (Revell 2012). Given uncertainty in the timing of the radiation, we applied an arbitrary age of 1 to the root and estimated relative divergence times under a discrete clock model with ten distinct rate categories. We then fitted a Markov-k (Mk) model for discrete character evolution (Lewis 2001). To do this, we used the R package *phytools* v1.9-23 and the function 'fitMk' (Revell 2012). We took an agnostic approach in choosing the model of evolution for the perenniality and night blooming traits. For these two traits, using the Akaike Information Criterion (AIC), we selected the best model of evolution between equal rates, all rates different,

and irreversible rates where loss of a trait is possible while its gain is not (Appendix S1.5). We used the irreversible model for perenniality and night blooming because it had the lowest AIC scores (Appendix S1.5). For anthocyanin absence or polymorphism, we chose the all rates different model because of the likely different evolutionary rates between loss and gain of flower pigmentation (Rausher 2008), though we report stochastic character density maps with alternate models of evolution (Appendix S1.6). We simulated stochastic character maps onto the phylogenetic tree and 100 simulations (Huelsenbeck et al. 2003). Using the 'densityMap' function, we visualized the stochastic mapping posterior probability density as a color gradient along the branches of the tree. Then, using the 'density' function, we calculated the relative distribution of state changes from the stochastic mapping across the tree. To account for potential non-identifiability of rates of character evolution and diversification, we additionally inferred ancestral states for each character using a BiSSE model in the R package *diversitree* (Maddison et al. 2007, Fitzjohn 2012) and tested for state dependent speciation using Analysis of Variance (ANOVA).

Geography of speciation

To explore the geographic speciation history of *Linanthus*, we used data on species geographic ranges and phylogenetic divergences (Barraclough and Vogler 2000). Specifically, we estimated the relationship between phylogenetic distance and geographic range overlap between species pairs to evaluate evidence for a predominant geographic (allopatric vs. sympatric) mode of speciation in *Linanthus*. Under this approach, if allopatric speciation is the dominant process,

geographic range overlap between young species pairs should increase from ca. 0% to random association as species pairs become more divergent over time. By contrast, if sympatric speciation is the dominant process, geographic range overlap should be ca. 100% between young species pairs but decrease over time among older pairs due to post-speciation geographic range changes (Losos and Glor 2003; Fitzpatrick and Turelli 2006; Skeels and Cardillo 2019). To determine species geographic ranges, we downloaded species location data from the Southwestern Environmental Information Network (SEINet 2022) and the California Consortium of Herbaria (CCH 2022), using only records backed by herbarium collections. We filtered the data to exclude records without precise GPS coordinates (fewer than two decimals precision) and clear outliers in terms of known species distributions. We used these records to create range maps for each species and estimate geographic range overlaps between species pairs using the R package *hypervolume* (Blonder et al. 2014) and the approach implemented at https://github.com/eliotmiller/ebirdr/blob/master/R/hypervolumeOverlaps.R (Miller et al. 2019). Because hypervolumes are geometric shapes with complex geometries, including the presence of holes (Blonder 2016), they can describe species geographic ranges more faithfully beyond simple ellipsoids or convex hulls. Areas of the species range perimeter with no occurrence points can be excluded and ranges can have disjunctions (Blonder et al. 2014). Consequently, this method enables estimates of range overlap in heterogeneous environments with potential for small-scale allopatry. We calculated the Sørenson similarity index for each species pair and recorded the values in a similarity matrix with values of 0 representing no range overlap and values of 1 representing complete overlap. We used the phylogeny built with

ddRAD data and one sample per species to calculate the phylogenetic distance between all species pairs using the 'cophenetic' function in the R package *ape* (Paradis et al. 2004). We coded species pairs belonging to the same clade as "within clade comparisons" and to different clades as "between clade comparisons". To test whether phylogenetic distance is a predictor of range overlap, we fitted a zero inflated beta regression model using the R package *gamlss* (Cribari-Neto and Zeileis 2010, Rigby and Stasinopoulos 2005). Beta regressions are well-suited for our response variable, range overlap, which is bounded between 0 and 1 (Ferrari and Cribari-Neto 2004). We ran three regressions using all of the data, comparisons of species in the same clade, and of species belonging to different clades.

Results

Data processing

The ddRAD sequencing yielded an average of 2,440,318 reads per sample. The assembly with 180 samples in a minimum of four individuals yielded 36,861,279 bp, 3,131,821 SNPs, and a total of 165,943 loci with 95% missing data. An average of 7646 loci per sample were retained. The assembly using one sample per species yielded 4,047,394 bp, 459,834 SNPs, 17,751 loci in a minimum of four individuals, and an average of 4053 loci per sample. For the TC data, we selected a total of 219 contigs with 100% taxon occupancy for downstream analyses.

Phylogenetic inference

The ddRAD and the TC data sets both resulted in well-resolved phylogenies. In the concatenated ddRAD maximum likelihood phylogeny, most nodes had a bootstrap value above 95, with only two nodes at 73 and 76 (Fig. 1.2). With an average of seven samples per species, 17species resolved as monophyletic, while eight were nonmonophyletic (Fig. 1.2). The TC phylogeny inferred using IQtree (Fig. 1.3A) recovered nearly all of the same species relationships as in the phylogeny using ddRAD (Fig. 1.2). There were two exceptions in the congruence of the ddRAD and TC phylogenies. One is in a clade recovered with ddRAD data that included *L. bellus*, *L. concinnus*, *L. dianthiflorus* (Benth.) Greene, *L. orcuttii*, and *L. uncialis*. The TC data resolves *L. dianthiflorus* and *L. concinnus*, and *L. bellus* and *L. orcuttii* as a grade, but no samples of *L. uncialis* were included in the TC analysis. The other incongruence was in a clade recovered with ddRAD data that included *L. demissus* (A. Gray) Greene, *L. bernardinus*, *L. killipii*, and *L. parryae.* Target capture data supports *L. bernardinus*, *L. killipii*, and *L. parryae* as more closely related to the annual night blooming clade than to *L. demissus*. The ddRAD and TC data represent two independent sources of data and some incongruence between the evolutionary history of these two types of genetic data might be expected due to incomplete lineage sorting (Appendix S1.11). Given that our results were generally consistent across data sets, we are informally referring to several common clades and one grade as follows.

Early diverging grade

This grade includes the following species previously described as *Gilia* section Giliastrum: *L. inyoensis* (I.M. Johnst.) J.M. Porter & L.A. Johnson, *L. campanulatus* (A. Gray) J.M. Porter & L.A. Johnson, *L. filiformis* (Parry Ex. A. Gray) J.M. Porter & L.A. Johnson, and *L. maculatus* (Parish) Milliken (Grant 1959, Bell et al 1999). While these species did not form a monophyletic group, the grade relationships are well-supported and identical in the ddRAD and TC phylogenies (Figs. 2, 3). The relationships were also consistent with previous studies with limited sampling (Bell et al. 1999, Bell and Patterson 2000). All five species in this grade were recovered as monophyletic (Figs. 2, 3).

Perennial clade

Both the ddRAD and TC analyses support the monophyly of a group including the perennial species *L. caespitosus* (Nutt.) J.M. Porter & L.A. Johnson, *L. californicus* (Hook. & Arn.) J.M. Porter & L.A. Johnson, *L. glaber* (R. Patt & Yoder-Will) J.M. Porter & L.A. Johnson, *L. jaegeri, L. pungens* (Torr.) J.M. Porter & L.A. Johnson, *L. veatchii* (Parry Ex. Greene) J.M. Porter & L.A. Johnson, and *L. watsonii* (A.Gray) Wherry (Figs. 2, 3). These species were previously recognized as the genus *Leptodactylon* Hook. & Arn. (Rydberg 1906). Although a single origin of the perennial clade was also supported in a recent study (Landis 2016), it was inconclusive in earlier phylogenetic analyses (Bell and Patterson 2000). The perennial clade included two nested subclades. One subclade included *L. caespitosus, L. glaber,* and *L. watsonii*, all of which occur outside of California in the Western United States (in Arizona, Colorado, Idaho, Montana,

Nebraska, Nevada, Utah, Wyoming). The other subclade included *L. californicus, L. jaegeri, L. pungens*, and *L. veatchii*, all of which are restricted to California or Baja California, with the exception of the more widespread *L. pungens,* which is found throughout western North America.

Most species in the perennial clade were not recovered as monophyletic (Figs. 2, 3). Although *L. caespitosus* and *L. glaber* were recovered as monophyletic (however, note the small sample size), both species were nested within a paraphyletic *L. watsonii*. In the other subclade, both *L. veatchii* and *L. californicus* were recovered as monophyletic, but they were nested within a clade that included the paraphyletic species *L. pungens* and *L. jaegeri*.

The rmaverick population structure analysis for the perennial clade showed the highest posterior probability for three distinct genetic clusters (Verity and Nichols 2016; Fig. 1.4). One cluster included all the samples of the species *L. watsonii* and *L. caespitosus*, which occur outside of California, another cluster included all the samples of the Baja California endemic *L. veatchii* and two samples of *L. pungens* from Southern California and Baja California, and a third cluster included all samples of the species *L. pungens, L. californicus,* and *L. jaegeri* (Fig. 1.4). The population structure results are consistent with the phylogenetic results showing a lack of genetic cohesiveness of the currently recognized taxonomic species in this clade.

Coastal California and Baja California annuals clade

The ddRAD phylogeny recovered a monophyletic group including *L. bellus, L. concinnus, L. dianthiflorus, L. orcuttii,* and *L. uncialis*, all of which occur in the coastal or higher elevation areas of Southern California and Baja California (Figs. 2,3). A recent phylogenetic study also recovered this clade (Landis 2016). While the monophyly of this group was not fully supported in the TC phylogeny, all species within this clade were closely related and formed a grade (Fig. 1.3). The ddRAD analysis with multiple samples per species recovered all species as monophyletic, though we could only include one sample for *L. concinnus* and *L. uncialis* (Fig. 1.2).

Desert annuals clade

In the ddRAD phylogeny, *L. bernardinus, L. demissus, L. killipii,* and *L. parryae* formed a clade sister to the annual night-blooming clade (see below) (Fig. 1.2). The TC species tree did not support the monophyly of the desert annual clade (Fig. 1.3). Instead, this tree recovered *L. demissus* as sister to a clade with two subclades, one including the remaining species in the desert annual clade (*L. bernardinus, L. killipii,* and *L. parryae*) and a subclade including the annual night-blooming species (see below) (Fig. 1.3). The broad sampling in the ddRAD phylogeny showed that most species in the desert annual clade were recovered as monophyletic, with the exception of *L. bernardinus* and *L. killipii*, both of which formed a clade with intermixed samples.

Annual night blooming clade

The annual night blooming clade included the species *L. arenicola* (M.E. Jones) Jeps. & L.H. Bailey, *L. bigelovii* (A. Gray) Greene, *L. dichotomus* Benth., *L. jonesii* (A. Gray) Greene, *L. maricopensis* J.M. Porter & R. Patt, and *L. viscainensis* Moran. All species in this clade are night bloomers, with the exception of *L. dichotomus* where populations of *L. dichotomus* subsp. *meridianus* flower during the day. This group was monophyletic in both the ddRAD and TC phylogenetic analyses (Figs 2, 3). This corroborates the section *Linanthus* described by Grant which included *L. bigelovii, L. dichotomus* and *L. jonesii* (Grant 1959). *Linanthus viscainensis* was later added to the section *Linanthus* based on its morphological similarities to *L. arenicola* (Moran 1977). This clade was also recovered in previous analyses with the addition of one accession of *L. filiformis* (Landis 2016).

Most species in this clade were highly paraphyletic or polyphyletic (Figs. 2, 3). *Linanthus arenicola* was the only monophyletic species. Six out of 11 samples of *L. jonesii* formed a clade, but it was nested within a more inclusive clade with several intermixed samples of *L. dichotomus* and *L. bigelovii*. Eight out of 11 *L. bigelovii* samples formed a clade, but it was nested in a more inclusive clade with several samples of *L. dichotomus* and *L. jonesii*. The only sample of *L. viscainensis* included here was nested within a more inclusive clade including samples of *L. bigelovii, L. dichotomus,* and *L. jonesii*. Thirteen out of 17 *L. dichotomus* samples and two samples of *L. bigelovii* formed a clade.

The population structure analysis in the annual night blooming clade showed the highest posterior probability for four distinct genetic clusters (Fig. 1.5). One cluster included the majority of samples identified as *L. bigelovii* (8/11), some samples identified as *L. jonesii* (5/11), and one sample identified as *L. dichotomus* (1/17). A second cluster included the majority of samples identified as *L. jonesii* (6/11), some samples identified as *L. dichotomus* (3/17), and one sample identified as *L. bigelovii* (1/11). A third cluster included only samples identified as *L. dichotomus* (11/17). A fourth cluster included some samples identified as *L. bigelovi* (2/11) and some samples identified as *L. dichotomus* (2/17) (Fig. 1.5). We did not detect any geographic signal characterizing these genetic clusters.

Trait evolution and ancestral states

Our traits do not seem to affect diversification rates. Perenniality, night blooming, and petal color polymorphism do not have an effect on diversification rates (P = 0.52, P = 0.51, P = 1, respectively) with AIC values supporting the state independent diversification model in each trait. Mapping the distribution of phenotypic traits onto the ddRAD phylogeny shows that perenniality is clustered in one clade, night blooming appears in the perennial clade and in the annual night blooming clade, and petal color polymorphisms are dispersed across the phylogeny in three of the major clades (Fig. 1.6). Stochastic character mapping showed that perenniality evolved once with no reversals to annuality (Fig. 1.7A). Night blooming evolved three times, once in the annual night blooming clade, and twice in the perennial clade where all but two species have night blooming populations (Fig. 1.7B, Appendix S1.7). Corolla lobe anthocyanin

pigment polymorphism may be ancestral in *Linanthus*, with several reversals to unpigmented corolla lobes (Fig. 1.7C). The posterior probability distribution for the number of changes shows that gain of anthocyanin polymorphism may have occurred twice and loss of anthocyanins ten times (Appendix S1.7).

Geography of speciation

The majority of species pairs showed < 50% overlap in geographic ranges regardless of phylogenetic distance (Fig. 1.8). Nonetheless, species range overlap varied from 0 to 80%. Between species pairs, there was no effect of phylogenetic distance on geographic range overlap (P = 0.93; Fig. 1.8). Young species pairs often had non-overlapping geographic ranges, however species comparisons within clades did not show a relationship between phylogenetic distance and range overlap (*P* = 0.90). Between species belonging to different clades, phylogenetic distance was also not a good predictor of geographic range overlap (P = 0.29). Together, these results suggest that a single geographic mode of speciation (allopatric or sympatric) does not predominate among *Linanthus* species.

Discussion

Phylogenetic relationships are well-resolved and most species are monophyletic

Our phylogenetic analyses generated well-resolved phylogenies that included all of the currently recognized species in *Linanthus*. The TC phylogeny (Fig. 1.3A) recovered nearly all of the same species relationships as in our concatenated ddRAD phylogeny (Fig. 1.2), and they both were

congruent with results from previous analyses with limited taxon and genetic sampling (Landis 2016). The species tree inferred with SVDQuartets from ddRAD data (Appendix S1.8) showed inconclusive species relationships due to very short branches, but this method is sensitive to large amounts of missing data common in RAD datasets (Nute et al. 2018). The species tree inferred with ASTRAL from TC data also recovered monophyly of most species, and most recent species relationships were congruent with the ddRAD data species trees (Appendix S1.9, S1.10). Notably, our results showed that most species are monophyletic even when we included multiple samples from different species co-occurring in sympatry with ample opportunities for interspecific gene flow. This suggests that reproductive isolating barriers have likely evolved for most species in this recent desert radiation. The lack of monophyly of some perennial and annual night blooming species may be due to interspecific gene flow, incomplete lineage sorting, or that current species limits are problematic. Recently, Porter and Patterson (2015) proposed a suite of infraspecific taxa within the perennial and night-blooming clades, reflecting the ongoing discovery of cryptic diversity in these groups. While we did not explicitly test these taxonomic hypotheses due to the difficulty of obtaining reliable identifications at subspecific rank for all herbarium specimens included in our study, we suspect that the non-monophyly of species witnessed in our study may support the elevation of some of these taxa to specific rank. Further work with more detailed study of herbarium material coupled with deeper taxon and genomic sampling and field experiments could be useful for rigorously testing these various hypotheses.

Perennials evolved from annuals

The species in the perennial clade all share several synapomorphies including the subshrub habit with a woody base, leaves that grow tightly in fascicles with sharp-tipped and filiform lobes or entire filiform leaves, and salverform corolla. These species occur in mountainous regions, consistent with the finding that alpine environments favor long lived species that can persist the colder winters (Billings and Mooney 1968; Ricklefs and Renner 1994).

Shifts from annual to perennial habit are considered rare across angiosperms, with perenniality thought to be the ancestral state in flowering plants and within most families and genera (Friedman 2020). However, recent phylogenetic comparative work has shown examples of transitions from annual to perennial habit (Tank and Olmstead 2008; Soltis et al. 2013). In Polemoniaceae, transitions between annuality to perenniality in both directions are common including in *Phlox* and *Leptosiphon*, which are two closely related genera of *Linanthus* (Barrett et al. 1996). Both the ddRAD and the TC phylogenies presented here support the evolution of the perennial species from an annual ancestor, and no reversals to annuality (Fig. 1.7A). Thus, the transition to perenniality is unidirectional from annual ancestors. The evolution of perenniality may have implications for diversification and evolution in *Linanthus*. Perennial plants generally show a slower rate of molecular evolution than annual plants, which has been attributed to longer generation times (Andreasen and Baldwin 2001) and larger population sizes (Bousquet et al. 1992). Our BiSSE analysis did not support an association between perenniality and

diversification, but the BiSSE model requires a larger data set than was used here to produce statistically robust results (Davis et al. 2013).

Night blooming evolved three times

We found that night blooming has only evolved three times in *Linanthus* (Fig. 1.7B, Appendix S1.7) with one reversal to day blooming in *L. dichotomus*. Night blooming has evolved multiple times across angiosperms but it is relatively rare (Silberbauer-Gottsberger and Gottsberger 1975; Grant 1983). In Polemoniaceae, hawkmoth pollination seems to have evolved in three genera, *Linanthus*, *Ipomopsis* Michx. and *Phlox*, but only *Linanthus* has flowers that open exclusively at night (Grant 1983). This suggests further specialization for hawkmoth pollination in *Linanthus* because there are no other pollinators available to visit the flowers at night.

Night blooming is rare and may be a mechanism of allochronic speciation

Night blooming species often occur in low densities and are spread across the landscape. This demographic pattern of night bloomers is associated with dry habitats, where daytime anthesis might lead to excessive water loss through transpiration (Stebbins 1970). Populations of the annual night blooming *Linanthus* species show these demographic patterns. If night blooming is indeed advantageous from a hydraulic perspective in *Linanthus*, the question is why more species have not evolved this trait given their widespread distribution in the hot, dry deserts. Detailed physiological and genomic studies could shed light on the biological mechanisms

underpinning this phenotype. Alternatively, it is plausible that shifts in flowering time have evolved as prezygotic isolating barriers and are linked to a more complex phenotype involved in pollination and possibly allochronic speciation (Taylor and Friesen 2017). Night blooming flowers are usually white, pale yellow or pale pink with strong scents emitted at night, and with long corolla tubes (Grant 1983; Knudsen and Tollsten 1993). These traits are present in all the annual night blooming *Linanthus*, with variation within and between species, making it a great system to study the genetic basis of this complex phenotype likely involved in speciation. For instance, this system can be used to study which suite of traits is specialized to moth pollination and which attracts a wider variety of pollinators. Future work can also investigate how the loss of petal pigment and specialization on night pollinators may be leading the night bloomers to a potential evolutionary dead-end (Tripp and Manos 2008).

Our results show that populations of species in the night blooming clade co-occur in sympatry with populations from species in other clades. Shift in blooming time could work as a temporal isolating barrier when no geographic barriers limit interspecific gene flow. Because night blooming species co-occur with species that have diverged at different times (Figs. 2, 8), it is unclear if night blooming is the cause of reproductive isolation or evolved later in secondary sympatry as a reinforcement mechanism (Taylor and Friesen 2017).

Selfing may be a mechanism of reproductive isolation between closely related species—

Outside of night anthesis, the *Linanthus* species in the annual night-blooming clade are united morphologically by their white or yellow corolla color, short or non-existing pedicel, and a cylindrical or urn shaped calyx, with the membrane separating sepals that are wider than the lobes. The stamens are included in the corolla tube, with the pistil at or below the stamens. This combination of traits is often found in flowers that can self-pollinate (Ushimaru and Nakata 2002). Indeed, most *Linanthus* species in the annual night blooming clade can set seed without the corolla emerging out of the calyx (pers. obs.). Self-compatibility is common in other hawkmoth pollinated flowers (Grant 1983). For instance, *L. arenicola* and some populations of *L. jonesii* and *L. bigelovii* have very short floral tubes, making hawkmoth pollination less likely. Therefore, it is possible that some populations of these species are autogamous. Further study is needed to determine whether these populations are facultative selfers, growing smaller autogamous flowers in years when conditions are not favorable. The gradient from selfing to outcrossing via hawkmoth pollination in this night blooming clade provides the opportunity to investigate different strategies for reproductive isolation in closely related species. Similar trends in selfing features are present in *Leptosiphon,* a closely related genus of *Linanthus* (Goodwillie 1999, Goodwillie and Stiller 2001, Goodwillie and Ness 2005). Selfing in *Leptosiphon* may have evolved to ensure reproductive success in environments with inconsistent pollinator visitation and with range expansion to drier habitats that decouple pollinator emergence from flowering time (Goodwillie 1997). Whether similar mechanisms operate and evolve independently in *Linanthus* is unknown.

Scent may act as a reproductive isolating mechanism, especially in night blooming species— Night blooming species use scent to attract pollinators because visual cues are less effective at night (Raguso and Willis 2002). Many night blooming flowers emit a heavy sweet musky scent, which is an olfactory attractor of noctuid moths (Raguso et al., 2003). *Linanthus dichotomus*, the showiest species in the night blooming clade, emits scents associated with moth attraction, and different chemical profiles between the day and night blooming subspecies (Chess et al. 2008). Analyzing the chemical profiles of *Linanthus* species will improve our understanding of the evolution of the hawkmoth pollination syndrome. Further chemical ecology studies are needed to explore how floral scent works as a potential isolating mechanism in *Linanthus*.

Color polymorphism may be ancestral

Flower anthocyanin polymorphisms in *Linanthus* may have evolved early in the history of the genus, in the clade that includes most species in the genus except for *L. inyoensis*, *L. filiformis* and *L. campanulatus* (Fig. 1.7C). Although these three early diverging species do not exhibit color polymorphisms, several other genera in Polemoniaceae have color-polymorphic species (Schemske and Bierzychudek 2007). Anthocyanin pigmentation likely evolved in the ancestor of *Linanthus* or even earlier in the history of Polemoniaceae (Landis et al. 2018), but the evolution of anthocyanin-based polymorphism has not been studied in this family. A noteworthy feature of the anthocyanin polymorphisms in *Linanthus* is that no species are pigmented and monomorphic (i.e., there are no species with only pink or purple petals), and all pigmented species also have a non-pigmented morph (i.e., a species with pink petals also has individuals

with white petals). The pigmented monomorphic trait may be a hidden state in the evolutionary history of *Linanthus* but it is not possible to infer from our analyses whether pigmented only species existed in the genus. Despite this, the prevalence of polymorphisms in the genus indicates that polymorphisms are a shared derived trait in at least some of the *Linanthus* lineages. It is expected that polymorphisms are rarely retained across speciation events, because the genetic variation responsible for the polymorphism as well as the disruptive selective pressure maintaining the polymorphism must persist through time (Jamie and Meier 2020). While polymorphisms might be a precursor to speciation, polymorphisms that persist through a speciation event must be a result of different forces than those driving speciation (Gray and McKinnon 2007). Small scale habitat differences or temporal fluctuations are possible disruptive forces that do not directly lead to reproductive isolation, and have been credited with maintaining the color polymorphism in *Linanthus parryae* (Schemske and Bierzychudek 2001, 2007). The prevalence of color polymorphisms across the *Linanthus* radiation and the potential for its mechanism to be maintained through speciation events makes this genus an ideal subject to investigate the dynamics between speciation and polymorphisms.

Flower pigmentation evolutionary transitions are common across the angiosperm phylogeny (Rausher 2008, Smith and Goldberg 2015) and in Polemoniaceae (Landis et al. 2018), but the macroevolution of color polymorphisms remains understudied. There are few documented examples of multiple closely related species sharing flower color polymorphisms with a few exceptions in *Antirrhinum* L. (Lamiaceae) and *Protea* L. (Proteacea, Carlson and Holsinger 2015,

Ellis and Field 2016). In California occurring Polemoniaceae, 36% of the species exhibit flower color polymorphism (Schemske and Bierzychudek 2007), in *Ipomopsis, Leptosiphon, Linanthus* and *Phlox*. In *Linanthus*, 44% of the species are polymorphic across populations and 28% within populations (Patterson and Porter 2021, pers. obs.). The pink, lavender, and purple color variation in *Linanthus* are likely anthocyanin based pigments (Tanaka et al. 2008). The inability to synthesize anthocyanin pigments has been suggested as a mechanism for producing the white corolla individuals (Warren and Mackenzie 2001), but this mechanism remains to be tested in *Linanthus*. The loss of the color polymorphism seems to have occurred several times in the history of *Linanthus* (Appendix S1.7), supporting that color is likely more easily lost than gained (Rausher 2008).

Many of the species in *Linanthus* exhibit continuous color variation. For example, the iconic *Linanthus parryae* has been coded as dimorphic in the classic papers investigating the evolution of such polymorphism (Epling and Dobzhansky 1942; Wright 1943; Schemske and Bierzychudek 2007). However, observations in the field across the flowering season indicate that individuals range in intensity of flower color, both within and between populations. In addition, the color of the purple morphs appears to fade with heat or time since anthesis (pers. obs.). These characteristics point to a potentially more complex mechanism for the genetics and plasticity of flower color than previously suspected in *Linanthus parryae*, and in the other *Linanthus* polymorphic species.

While anthocyanin-produced pigment influences pollinator attraction, it also has survival functions, including deterring herbivores and responding to abiotic stresses such as UV exposure and drought (Strauss and Whittall 2006). Across angiosperms, flower color polymorphisms are most common in heterogeneous environments, like the Mediterranean or high elevation biomes (Sapir et al. 2021). Experimental studies have shown that in white and pink/purple flower polymorphic species, the pink/purple morphs show greater tolerance to drought and heat stress (Warren and Mackenzie 2001; Coberly and Rausher 2003; Vaidya et al. 2018; Dittmar and Schemske 2023). *Linanthus parryae* pollinator visitation did not differ between two color morphs, but reproductive success was higher in white flowered individuals in wetter years, and in purple flowered individuals in dry years (Schemske and Bierzychudek 2001). Color morphs differences correlated to rainfall abundance point to a potential pleiotropic effect of flower color and environmental factors in *L. parryae* (Schemske and Bierzychudek 2007) but can also point to potential linked genes producing these seemingly unrelated phenotypes (Rausher 2008). The white and pink/purple polymorphism commonly found in *Linanthus* may be an important adaptation that allowed plants to tolerate the variable environmental conditions in xeric areas of Western North America and might have been a precursor to its diversification in this geographic area with high UV radiation, drought and heat. The prevalence and the potential ancestral origin of color polymorphisms in *Linanthus* opens promising future avenues for studying the mechanisms that maintain polymorphisms within species and across speciation events.

No predominant mode of speciation, but range overlap in sister species is indicative of parapatric speciation

We did not find a relationship between phylogenetic distance and species range overlap, thus we did not detect a prevalent geographic mode of speciation in *Linanthus*. Though current species distributions are not necessarily reflective of the species range at the time of speciation (Losos and Glor 2003), some signature of this pattern is likely present in the current geography of closely related species (Barraclough and Vogler 2000). Allopatric speciation has been considered the dominant mode of speciation (Mayr 1959, Coyne and Orr 2004). If this is the case, we would expect geographic range overlap to increase over time with sympatry arising secondarily after populations have accumulated enough differences and become reproductively isolated. However, we did not detect signals of such a pattern (Fig. 1.8). Recent studies have suggested that plant sympatric speciation is more prevalent than other geographic modes of speciation (Skeels and Cardillo 2019). Under this scenario, we would expect geographic range overlap to decrease due to post-speciation geographic range changes (Losos and Glor 2003). Yet, our findings for *Linanthus* were not consistent with this pattern (Fig. 1.8). Taken together, our results indicate that multiple speciation mechanisms are at play in the *Linanthus* radiation and not a single mode predominates. For instance, we found that some young species pairs show considerable geographic range overlap, while others overlap minimally. This difference suggests that in some cases speciation could have happened in parapatry with opportunities for homogenizing gene flow and in other cases, speciation likely happened in isolation in allopatry. Further geographic sampling at the landscape scale combined with detailed studies in regions of

geographic overlap using simulations and population genomic approaches will be essential to confidently discern speciation modes.

Recent studies of California plants suggest that the patterns of geographic range overlap, geographic range asymmetry, and time since divergence between species pairs were not consistent with allopatry as the dominant mode of geographic speciation (Anacker and Strauss 2014, Grossenbacher et al. 2014, Christie and Strauss 2018). However, these studies estimated range overlap using a different approach to the one employed here. We used overlap of hypervolumes that account for density of points and holes in the distribution of species occurrences, and holes in overlap from calculations of sympatry (Blonder et al. 2014). This method is sensitive to overlap at a finer scale, potentially excluding areas of overlap where the potential for gene flow between species pairs is low. Fine scale geographic partitioning or "micro-allopatry" may be common in California native plant species (Anacker and Strauss 2014, Grossenbacher et al. 2014). These previous studies used the overlap between polygons formed by occurrence points, and the difference in range overlap calculation methods could explain the overall discrepancies between our findings and the ones reported for those studies.

The expected pattern of increasing overlap with increasing phylogenetic distance under allopatric speciation may be apparent in a genus that diversified across a more homogenous environment, with fewer barriers to range expansion and contraction. However, this pattern may not emerge in *Linanthus* even if allopatric speciation is common because some species

pairs never experience secondary sympatry (Fig. 1.8). The lack of geographic overlap could result from specialization to certain habitats or pollinators, the inability to expand geographic ranges, or the existence of a heterogeneous environment with barriers to dispersal. These are patterns we commonly see in the ecology and geography of *Linanthus*. Testing these hypotheses will lay the groundwork for future studies exploring speciation mechanisms in *Linanthus* across the harsh North American deserts.

Conclusion

In this study, we present a comprehensive phylogenetic study of the genus *Linanthus*, a diverse radiation of mostly annual plants from the biodiverse deserts of Southwestern North America. Our phylogenies are the first to date to include complete species sampling and extensive intraspecific sampling. This approach allowed us to explore the monophyly of species and species relationships with increased rigor. Most species resolve as monophyletic despite rampant local sympatry and range overlap, suggesting the presence of strong isolating ecological barriers. The species within the annual night blooming clade and the perennial clade are not monophyletic, and closer taxonomic and population level studies are needed to untangle the evolutionary patterns of those species. Although we do not detect a strong signal for a predominant geographic mode of speciation, most species show some overlap in geographic range regardless of time since divergence. This suggests that some species could have evolved in parapatry, likely in the face of gene flow, while others likely evolved in isolation and never attained secondary sympatry. The strategies that *Linanthus* species have evolved to

deal with desert living, including flower color polymorphisms, facultative selfing, night blooming, and annuality, make it a rich system to study how plants can adapt to a drier world.

Acknowledgements

The authors thank Jade Corpus-Sapida and Sarah Payne for helping with DNA extractions. We are grateful to the curators of ASU, BRY, CAS, RSA, SBBG, SD, UCLA, and UNR herbaria for permission to sample from herbarium specimens, especially to Mare Nazaire, who provided a bulk of the herbarium samples used in this study and Elizabeth Makings, who provided a key determination. We thank Robert Patterson for inspiring discussions about *Linanthus* taxonomy and evolution, and Mark Porter for taxonomic context and training on the pectinase DNA extraction protocols. We are grateful to Claudia Henriquez for DNA extraction and library preparation training and support. We thank Sonal Singhal for discussion about alternate analyses when we seemed to hit a dead-end, and Isaac Overcast for help troubleshooting our ipyrad assemblies. We thank Duncan Bell, Maisie Borg, and Steven Serkanic for providing supplemental samples for this project. We are very appreciative to Kathleen Kay, Victoria Sork, and Nathan Kraft for providing comments that strengthened the arguments of this manuscript. We thank two anonymous reviewers for their valuable comments that improved the quality of the paper. Funding for this work was generously provided to I.G.A. by the Ecology and Evolutionary Biology Department at UCLA, the UCLA La Kretz Center and Stunt Ranch “Graduate Research Grant”, the Botanical Society of America “Bill Dahl Graduate Student Research Award”, the American Society of Plant Taxonomists “Graduate Student Research Grant”, the California

Native Plant Society “Hardman Native Plant Research Award” and F.Z.’s start-up funding. I.L.M. was funded by NSF DBI #2209393. This work is part of I.G.A.’s Ph.D. dissertation at UCLA.

Author contributions

I.G.A and F.Z. designed the study. I.G.A conducted the research, analyzed the data and wrote the manuscript. L.L.S and I.G.A prepared sequencing libraries. I.L.M. assembled the target capture sequence data and advised on analyses. All authors contributed to the editing and revision of the manuscript.

Data availability statement

All sequences generated for this work will be available on NCBI Sequence Read Archive (SRA) under the BioProject PRJNA1142084. Appendices, alignments, trees, character matrix and scripts will be available on github.com/ioanaanghel/Linanthus_phylogeny.

Figures

Figure 1.1. *Linanthus* encompasses extensive floral diversity with many polymorphic species in color and floral markings. Species greatly differ in corolla tube depth and in pollinators observed to visit flowers. Photographs were taken by I.G.A. (A) *L. arenicola.* (B) *L. bellus.* (C) *L. bernardinus.* (D) *L. bigelovii.* (E) *L. californicus.* (F) *L. campanulatus.* (G) *L. concinnus.* (H) *L. demissus.* (I,J) *L. dianthiflorus.* (K) *L. dichotomus.* (L) *L. filiformis.* (M) *L. inyoensis.* (N) *L. jonesii.* (O,P) *L. killipii*. (Q) *L. maculatus*. (R,S) *L. orcuttii*. (T,U) *L. parryae*. (V,X) *L. pungens.*

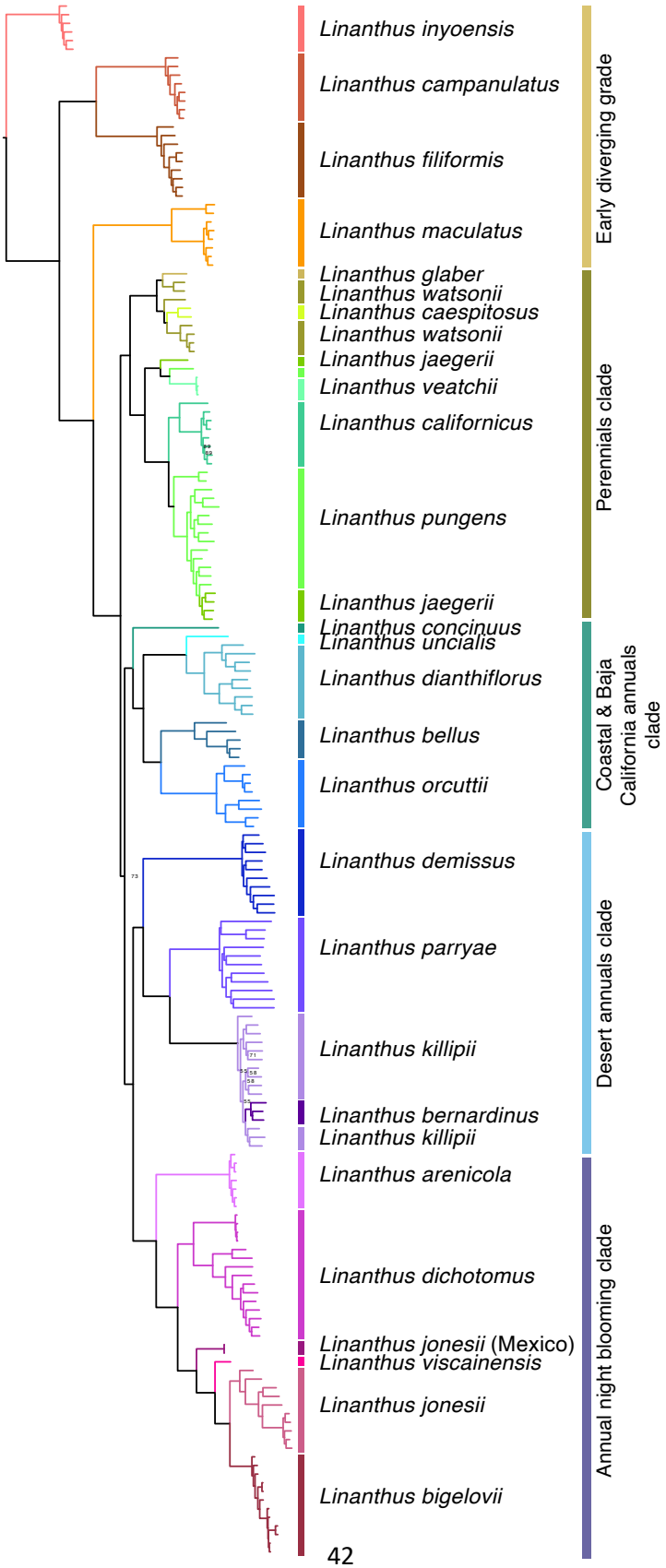
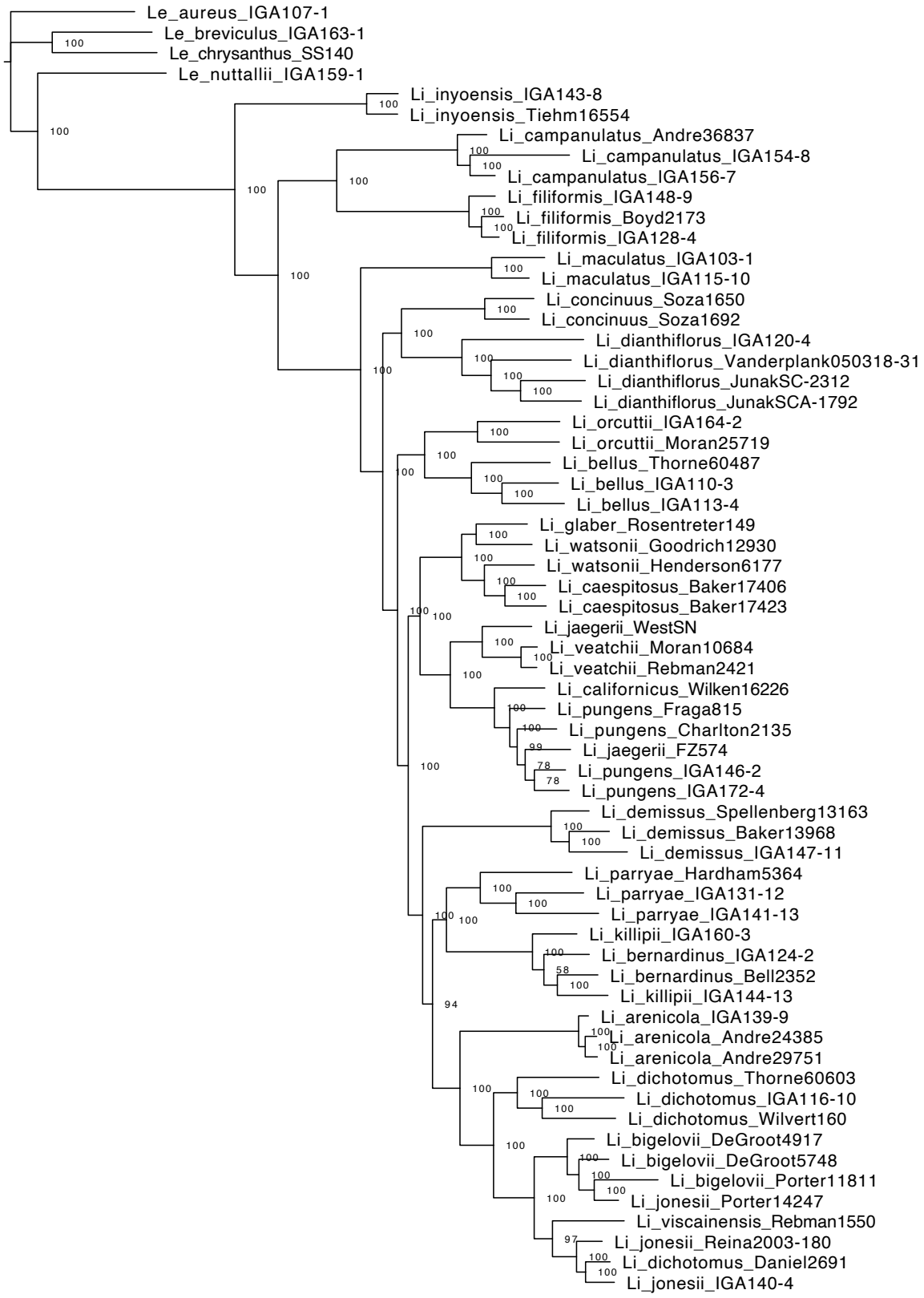

Linanthus inyoensis
Linanthus campanulatus
Linanthus filiformis
Linanthus maculatus
Linanthus glaber
Linanthus watsonii
Linanthus caespitosus
Linanthus watsonii
Linanthus jaegerii
Linanthus veatchii
Linanthus californicus
Linanthus pungens
Linanthus jaegerii
Linanthus concinuus
Linanthus uncialis
Linanthus dianthiflorus
Linanthus bellus
Linanthus orcuttii
Linanthus demissus
Linanthus parryae
Linanthus killipii
Linanthus bernardinus
Linanthus killipii
Linanthus arenicola
Linanthus dichotomus
Linanthus jonesii (Mexico)
Linanthus viscainensis
Linanthus jonesii
Linanthus bigelovii
Early diverging grade
Perennials clade
Coastal & Baja California annuals clade
Desert annuals clade
Annual night blooming clade

Figure 1.2. The phylogeny of *Linanthus* inferred using ddRAD data is well resolved and highly supported. A maximum likelihood phylogenetic tree inferred in IQtree including 180 samples for all species of *Linanthus* using a concatenated matrix of ddRAD data and a minimum of 4 samples per locus. Values at nodes represent bootstrap support below 90. The average of 7 samples per species included shows that most of the species are monophyletic. The clades recovered share common morphological features, habitats, or habits.


Le_aureus_IGA107-1
Le_breviculus_IGA163-1
Le_chrysanthus_SS140
Le_nuttallii_IGA159-1
Li_inyoensis_IGA143-8
Li_inyoensis_Tiehm16554
Li_campanulatus_Andre36837
Li_campanulatus_IGA154-8
Li_campanulatus_IGA156-7
Li_filiformis_IGA148-9
Li_filiformis_Boyd2173
Li_filiformis_IGA128-4
Li_maculatus_IGA103-1
Li_maculatus_IGA115-10
Li_concinuus_Soza1650
Li_concinuus_Soza1692
Li_dianthiflorus_IGA120-4
Li_dianthiflorus_Vanderplank050318-31
Li_dianthiflorus_JunakSC-2312
Li_dianthiflorus_JunakSCA-1792
Li_orcuttii_IGA164-2
Li_orcuttii_Moran25719
Li_bellus_Thorne60487
Li_bellus_IGA110-3
Li_bellus_IGA113-4
Li_glaber_Rosentreter149
Li_watsonii_Goodrich12930
Li_watsonii_Henderson6177
Li_caespitosus_Baker17406
Li_caespitosus_Baker17423
Li_jaegerii_WestSN
Li_veatchii_Moran10684
Li_veatchii_Rebman2421
Li_californicus_Wilken16226
Li_pungens_Fraga815
Li_pungens_Charlton2135
Li_jaegerii_FZ574
Li_pungens_IGA146-2
Li_pungens_IGA172-4
Li_demissus_Spellenberg13163
Li_demissus_Baker13968
Li_demissus_IGA147-11
Li_parryae_Hardham5364
Li_parryae_IGA131-12
Li_parryae_IGA141-13
Li_killipii_IGA160-3
Li_bernardinus_IGA124-2
Li_bernardinus_Bell2352
Li_killipii_IGA144-13
Li_arenicola_IGA139-9
Li_arenicola_Andre24385
Li_arenicola_Andre29751
Li_dichotomus_Thorne60603
Li_dichotomus_IGA116-10
Li_dichotomus_Wilvert160
Li_bigelovii_DeGroot4917
Li_bigelovii_DeGroot5748
Li_bigelovii_Porter11811
Li_jonesii_Porter14247
Li_viscainensis_Rebman1550
Li_jonesii_Reina2003-180
Li_dichotomus_Daniel2691
Li_jonesii_IGA140-4

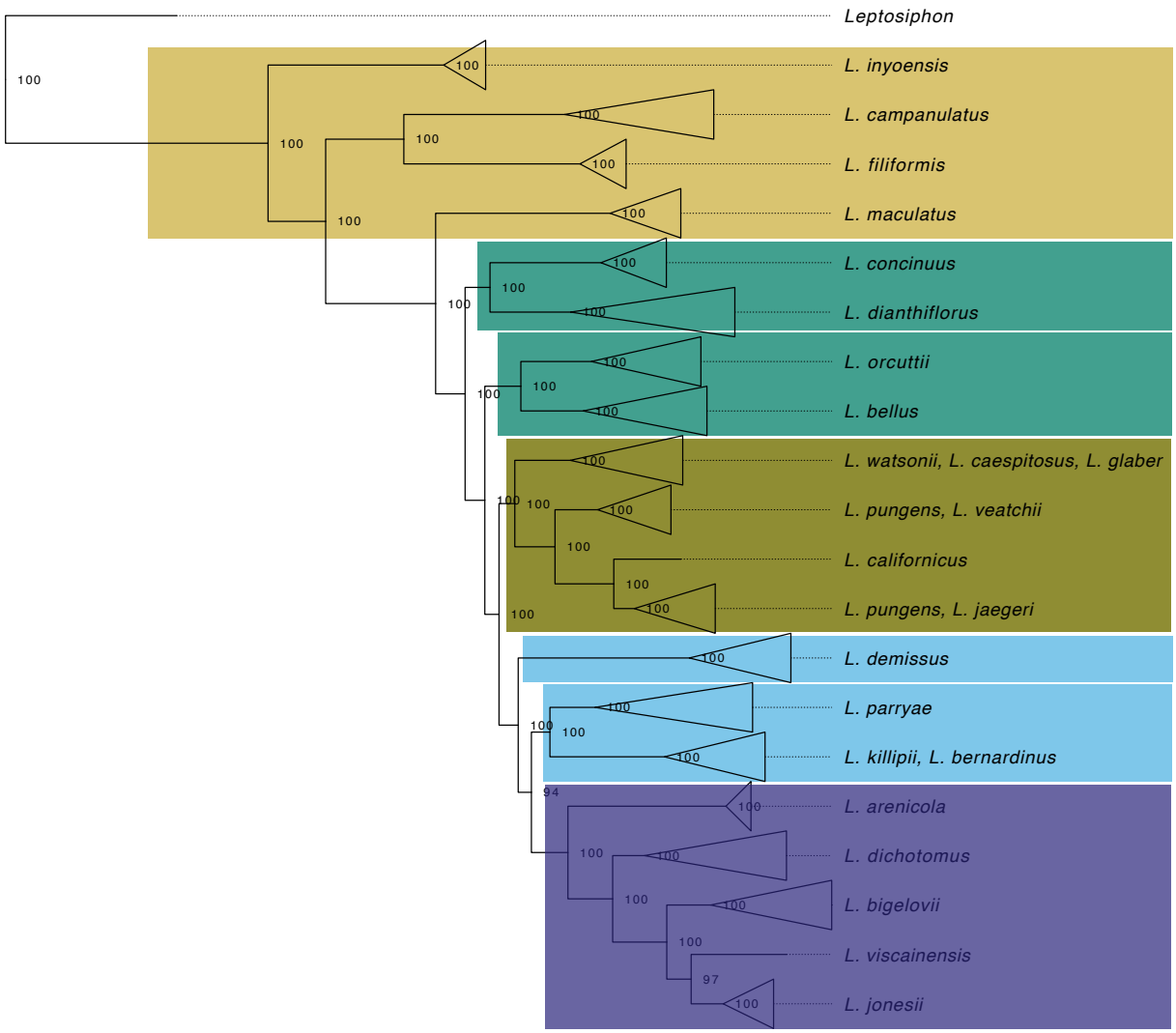


Figure 1.3. The phylogeny of *Linanthus* inferred using target capture data (Angisoperms353) data is well resolved and highly supported. This phylogeny is highly consistent with the phylogeny inferred using ddRAD data (see Fig. 1.2). A. Maximum likelihood phylogenetic tree inferred in IQtree including 63 samples across 22 species of *Linanthus* and 4 outgroups using a matrix of 219 concatenated Target Capture (Angiosperms353) loci with 100% occupancy. Values at nodes represent bootstrap support. Most species are recovered as monophyletic. B. The same phylogenetic tree as in A with collapsed species.

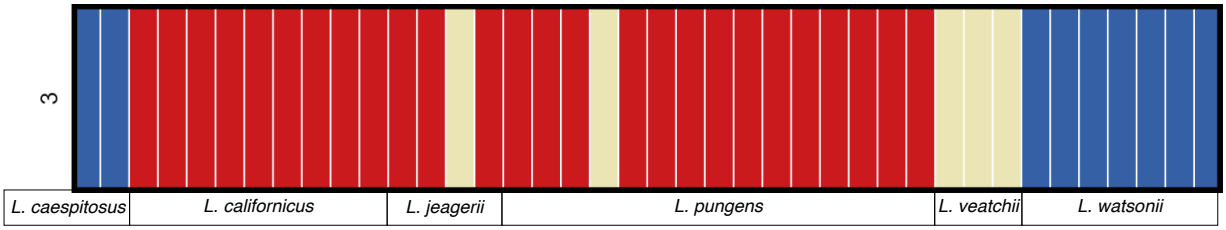


Figure 1.4. Genetic clusters did not match species taxonomy in the perennial clade. K=3 had the highest posterior probability, meaning that three genetic clusters were recovered for six taxonomic species. The red cluster included all samples identified as *L. californicus* and most samples identified as *L. pungens* and as *L. jaegeri*. The blue clade included all samples identified as *L. caespitosus* and *L. watsonii*. The tan group included all samples identified as *L. veatchii*, one sample identified as *L. jaegeri*, and one sample identified as *L. pungens*.

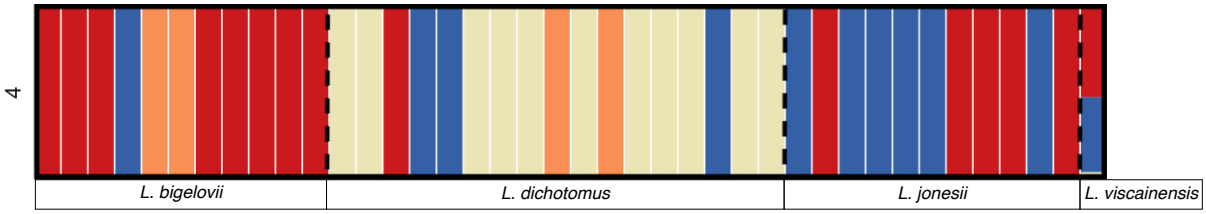


Figure 1.5. Genetic and taxonomic groups in the annual night bloomers showed little congruence. K=4 had the highest posterior probability, meaning that four genetic clusters were recovered for four taxonomic species. The red cluster included a majority of samples identified as *L. bigelovii*. The blue cluster included a majority of samples identified as *L. jonesii.* The tan cluster included only samples identified as *L. dichotomus*. The orange cluster included two samples identified as *L. bigelovi* and two samples identified as *L. dichotomus*. We excluded *L. arenicola* from this analysis because it was monophyletic.

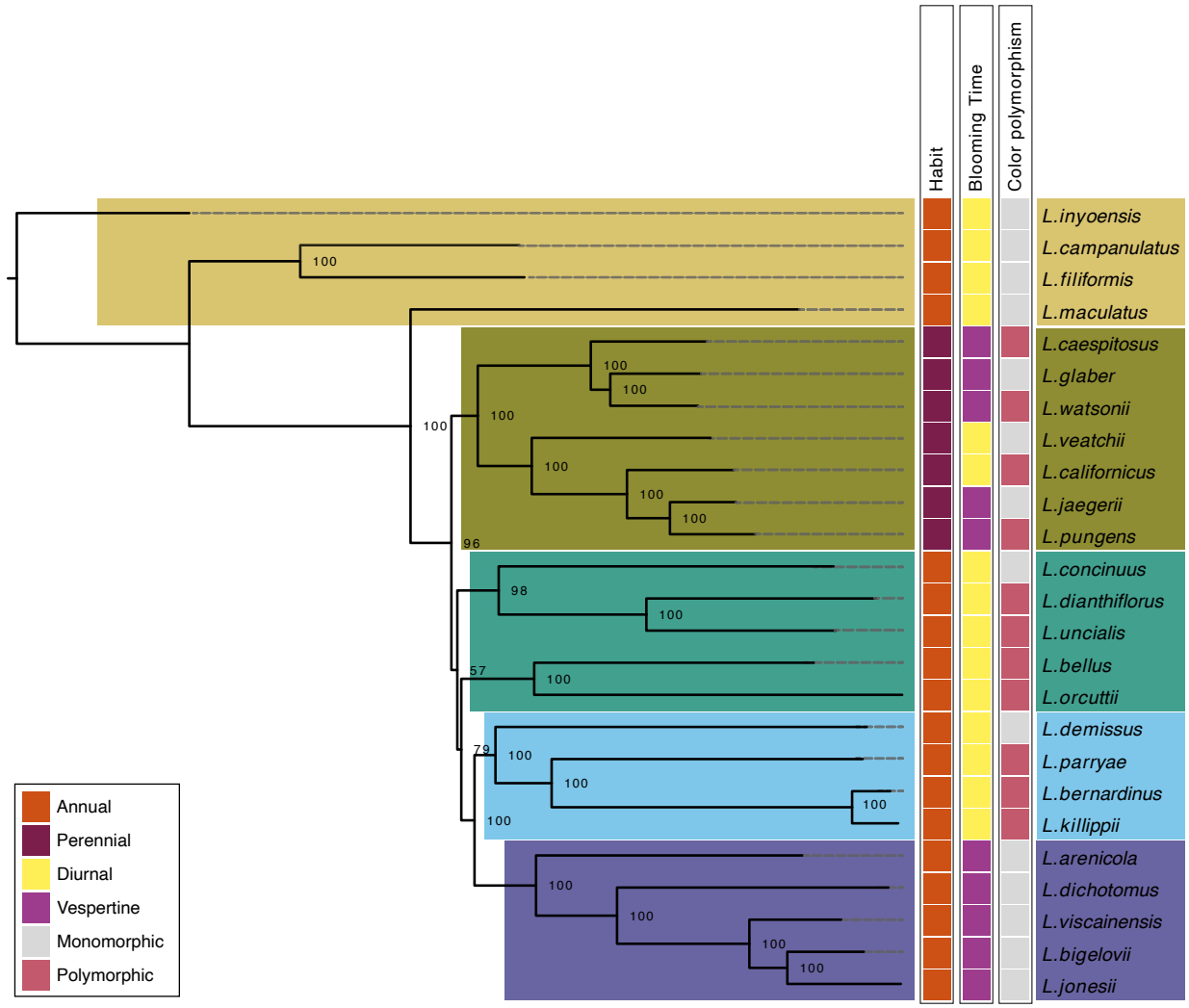


Figure 1.6. Phylogeny of *Linanthus* showing the distribution of three phenotypic traits. Left, *Linanthus* phylogeny using ddRAD data and one individual per species. Values represent bootstrap support. Right, diagram showing the character states distribution across species. Colors grouping taxa represent the clades we defined in Figure 2.

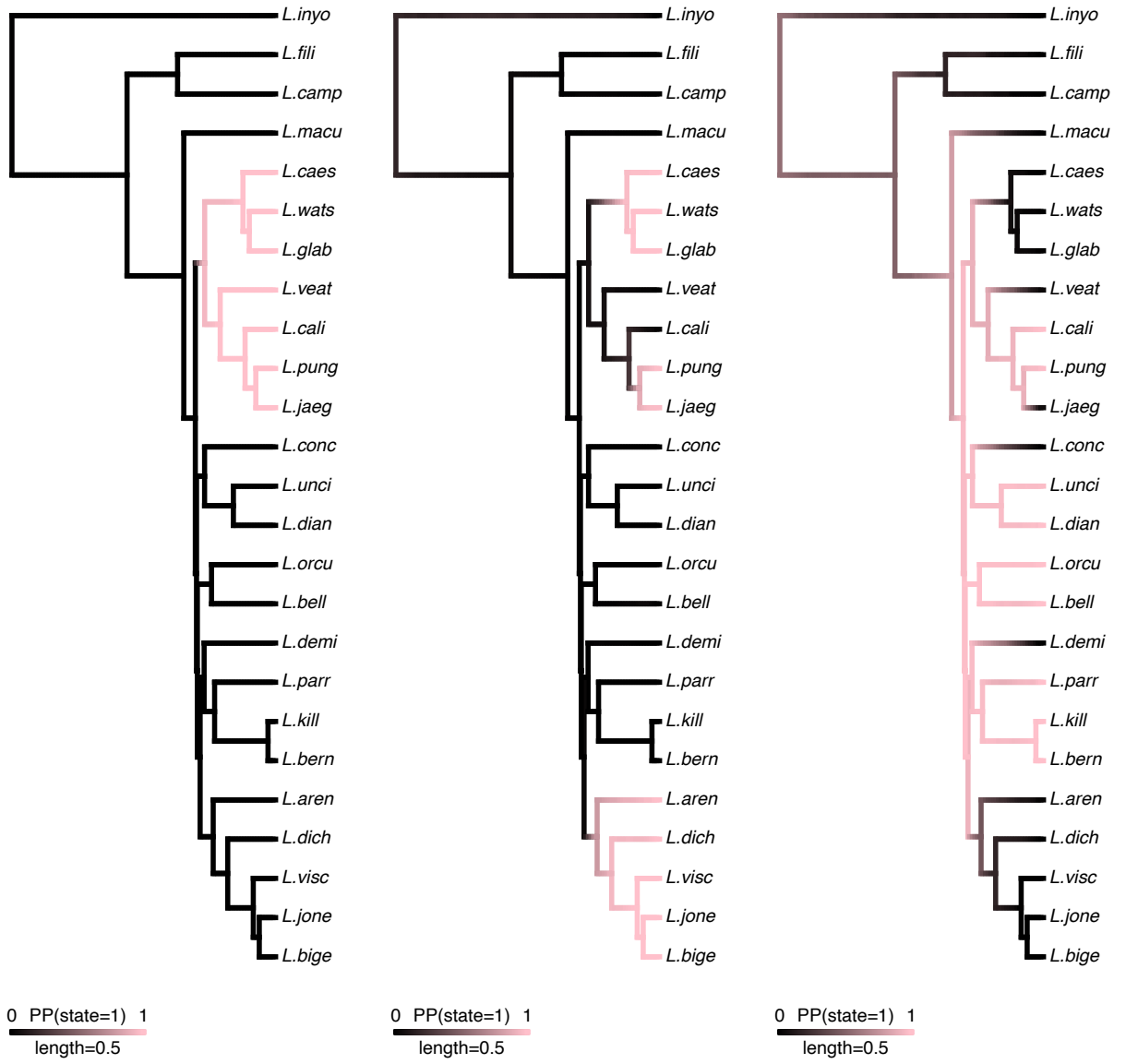


Figure 1.7. Across the *Linanthus* radiation, some phenotypic traits show little lability while others are highly labile. Each panel shows stochastic character density maps for different traits. A. Perenniality evolved once (1 = perenniality). B. Night blooming evolved three times, once in an annual clade, and twice in the perennial clade (1 = night blooming). C. Corolla lobe anthocyanin pigment polymorphisms may be ancestral in *Linanthus*, with several reversals to unpigmented corolla lobes. The most dominant state across *Linanthus*' evolutionary history is polymorphic, represented by the dominance of pink color along the branches in the phylogeny.

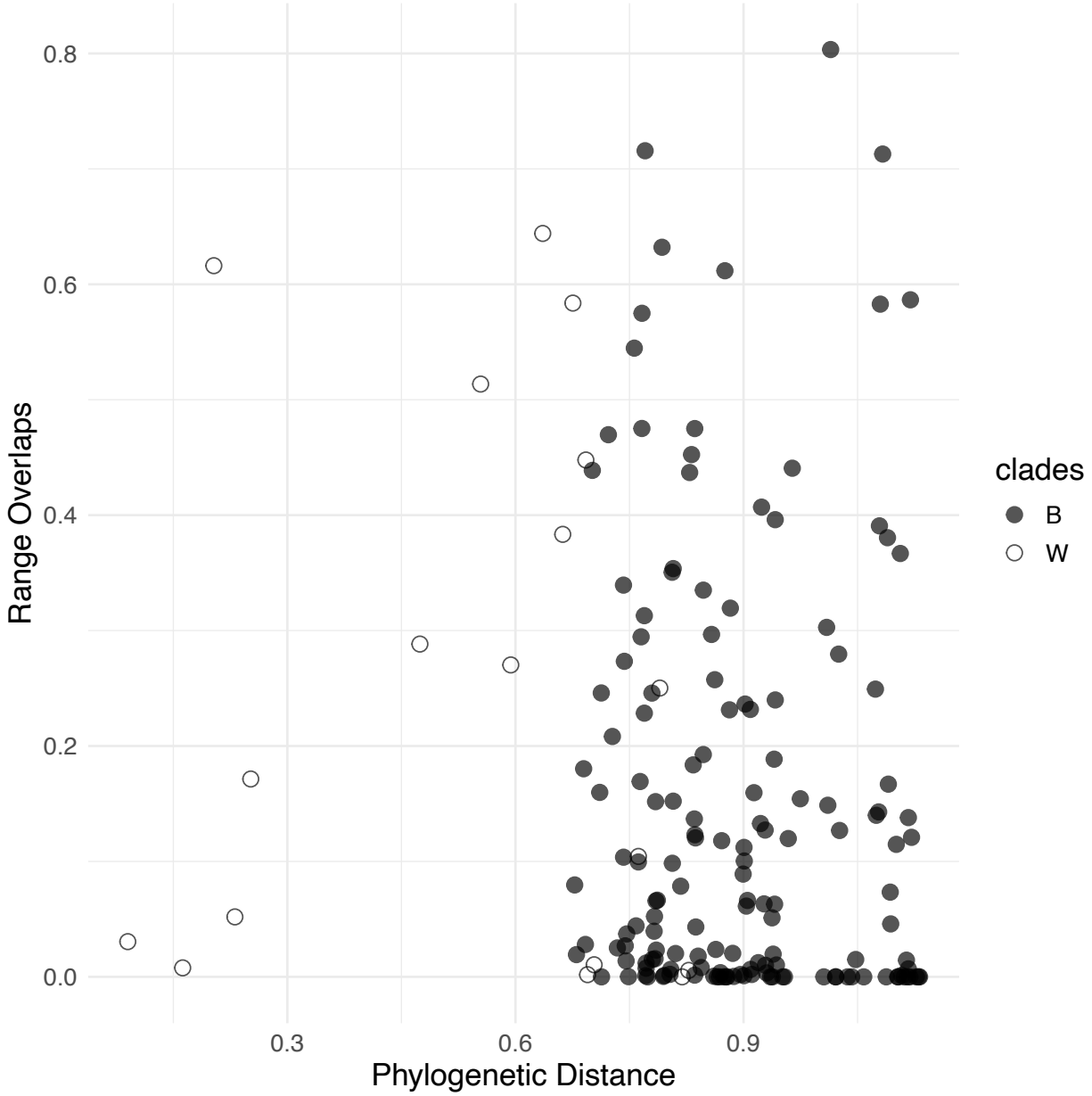


Figure 1.8. Phylogenetic distance does not predict range overlap of pairs of *Linanthus* species, suggesting there is no predominant geographic mode of speciation. There is no effect of phylogenetic distance on range overlap between species either within clades (W) or between clades (B). Pairwise phylogenetic distance is based on the ddRAD phylogeny using one individual per species, and geographic range overlap is estimated as the overlap in hypervolumes corresponding to geographic ranges for each species.

Appendix

Available at github.com/ioanaanghel/Linanthus_phylogeny.

Appendix S1.4. ipyrad parameters for assembly of RAD data.

Appendix S1.5. AIC scores for models of evolution to use in stochastic character mapping.

Appendix S1.6. Posterior probability density trees of anthocyanin polymorphism evolutionary history with two alternate models of evolution: equal rates and irreversible models.

Appendix S1.7. A distribution of trait changes across 1000 stochastic character mapped trees for perenniality, night blooming, and corolla anthocyanin polymorphism.

Appendix S1.8. RAD species trees. A. Species tree with ddRAD data generated with SVDQuartets. B. The same species tree with branch lengths showed inconclusive species relationships due to very short branches. Values represent branch lengths.

Appendix S1.9. TC species trees. A. Phylogenetic tree built with 63 samples across 22 species of Linanthus and 4 outgroups with 219 loci recovered via Target Capture (Angiosperms353) with 100% occupancy generated in ASTRAL. Values represent concordance factors. Most species are recovered as monophyletic. B. Species tree built with 219 loci recovered via Target Capture (Angiosperms353) with 100% occupancy generated in ASTRAL. Values represent concordance factors.

Appendix S1.10. Tanglegram of RAD and TC phylogenies with samples in common. The two data sets produce highly congruent phylogenies, with almost identical recent species relationships.

Appendix S1.11. The species tree calibrated using the divergence time of 20.8-42.5 MYA (Landis et al. 2018), for a range of lambda values ranging from zero, representing a free rate model, to one, representing a clock-like model. All of the models show that much of the diversification occurred since 20 MYA.

Appendix S1.12. A maximum likelihood phylogenetic tree inferred in IQtree including 180 samples for all species of *Linanthus* using a concatenated matrix of ddRAD data and a minimum of 4 samples per locus. Values at nodes represent bootstrap support. The average of 7 samples per species included shows that most of the species are monophyletic. This is the same

phylogenetic tree as in Figure 2, but it includes sample names.

References

Aburto-Oropeza, O., C. M. Burelo-Ramos, E. Ezcurra, P. Ezcurra, C. L. Henriquez, S. E. Vanderplank, and F. Zapata. 2021. Relict inland mangrove ecosystem reveals Last Interglacial sea levels. Proceedings of the National Academy of Sciences 118:7–11.

Anacker, B. L. and S. Y. Strauss. 2014. The geography and ecology of plant speciation: range overlap and niche divergence in sister species. Proceedings of the Royal Society B: Biological Sciences 281:1–9.

Andreasen, K., and Baldwin, B. G. (2001). Unequal evolutionary rates between annual and perennial lineages of checker mallows (*Sidalcea*, Malvaceae): Evidence from 18S-26S rDNA internal and external transcribed spacers. Molecular Biology and Evolution, 18(6), 936–944.

Anghel, I. G., S. J. Jacobs, M. Escalona, M. P. A. Marimuthu, C. W. Fairbairn, E. Beraut, O. Nguyen, E. Toffelmier, H. B. Shaffer, and F. Zapata. 2022. Reference genome of the color polymorphic desert annual plant sandblossoms, *Linanthus parryae*. Journal of Heredity 712–721.

Axelrod, D. I. 1972. Edaphic aridity as a factor in angiosperm evolution. American Naturalist 106:311–320.

Azani, N., A. Bruneau, M. F. Wojciechowski, and S. Zarre. 2019. Miocene climate change as a driving force for multiple origins of annual species in *Astragalus* (Fabaceae, Papilionoideae). Molecular Phylogenetics and Evolution 137:210–221.

Barraclough T. G. and A. P. Vogler. 2000. Detecting the Geographical Pattern of Speciation from Species-Level Phylogenies. American Naturalist. 155(4):419–434.

Barrett, S. C. H., L. D. Harder, and A. C. Worley. 1996. The comparative biology of pollination and mating in flowering plants. Philosophical Transactions: Biological Sciences 351:1271–1280.

Bayona-Vásquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., Bentley, K. E., Finger, J. W., Louha, S., Troendle, N., Diaz-Jaimes, P., Mauricio, R., and Faircloth, B. C. (2019). Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). PeerJ 7:e7724.

Bell, C. D. and R. Patterson. 2000. Molecular phylogeny and biogeography of *Linanthus* (Polemoniaceae). American Journal of Botany 87:1857–1870.

Bell, C. D., R. Patterson, and L. A. Hamilton. 1999. Sectional integrity in *Linanthus* (Polemoniaceae): A molecular phylogeny of section Dianthoides. Systematic Botany 24:632–644.

Billings, W. D. and H. A. Mooney. 1968. The ecology of arctic and alpine plants. Biological Reviews 43(4):481–529.

Blonder, B. 2016. Do hypervolumes have holes? American Naturalist 187(4): E93–E105.

Blonder, B., Lamanna, C., Violle, C., and B. J. Enquist. 2014. The n-dimensional hypervolume. Global Ecology and Biogeography, 23(5):595–609.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Borghi, M., Perez de Souza, L., Yoshida, T., and A. R. Fernie. 2019. Flowers and climate change: a metabolic perspective. New Phytologist 224(4): 1425–1441.

Bousquet J., S.H. Strauss, A.H. Doerksen, and R.A. Price. 1992. Extensive variation in evolutionary rate of rbcL gene sequences. Proceedings of the National Academy of Sciences. 89(16): 7844–7848.

Cacho, I. N., A. Millie Burrell, A. E. Pepper, and S. Y. Strauss. 2014. Novel nuclear markers inform the systematics and the evolution of serpentine use in *Streptanthus* and allies (Thelypodieae, Brassicaceae). Molecular Phylogenetics and Evolution 72:71–81.

California Native Plant Society, Rare Plant Program. 2023. Rare Plant Inventory (online edition, v9.5). Website https://www.rareplants.cnps.org [accessed 26 April 2023].

Capella-Gutierrez, S., Silla-Martinez, J. M., and T. Gabaldon. 2009. trimAl: a tool for automated trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972–1973. Carlson, J. E. and K. E. Holsinger. 2015. Extrapolating from local ecological processes to genus-wide patterns in colour polymorphism in South African *Protea*. Proceedings of the Royal Society B: Biological Sciences 282(1806).

Chess, S. K. R., Raguso, R. A., and G. LeBuhn. 2008. Geographic divergence in floral morphology and scent in *Linanthus dichotomus* (Polemoniaceae). American Journal of Botany, 95(12), 1652–1659.

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

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

Coberly, L. C. and M. D. Rausher. 2003. Analysis of a chalcone synthase mutant in *Ipomoea purpurea* reveals a novel function for flavonoids: Amelioration of heat stress. Molecular Ecology. 12(5): 1113–1124.

Consortium of California Herbaria (CCH). 2022. Data provided by the participants of the Consortium of California Herbaria. ucjeps.berkeley.edu/consortium (last accessed 12 October 2022).

Coyne, J. A. and H. A Orr. 2004. Speciation. Sinauer, Sunderland

Cribari-Neto, F., & Zeileis, A. 2010. Beta Regression in R. Journal of Statistical Software 34(2): 1–24.

Cullings, K. W. 1992. Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. Molecular Ecology 1:233–240.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Danforth, F. W. 1945. The genus *Linanthus*: a concept of the genus and a taxonomic study of its species. Stanford University, Masters Thesis.

Davis, M. P., Midford, P. E., and W. Maddison. 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. BMC Evolutionary Biology 13(38).

Dittmar, E. L. and D. W. Schemske. 2023. Temporal variation in selection influences microgeographic local adaptation. American Naturalist 202(4): 471–485.

Doyle, J. J. and J. L. Doyle. 1987. A rapid DNA Isolation procedure from small quantities of fresh leaf tissue. Phytochemical Bulletin 19:11–15.

Eaton, D. A. R., Spriggs, E. L., Park, B., and M. J. Donoghue. 2016. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Systematic Biology 66(3): 399–412.

Eaton, D. A. R. and I. Overcast. 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics 36:2592–2594.

Ellis, T. J. and D. L. Field. 2016. Repeated gains in yellow and anthocyanin pigmentation in flower colour transitions in the *Antirrhineae.* Annals of Botany 117(7):1133–1140.

Epling, C. and T. Dobzhansky. 1942. Genetics of natural populations. VI. Microgeographic races in *Linanthus parryae*. Genetics 27:317-322.

Evans, M. E. K., Smith, S. A., Flynn, R. S., and M.J. Donoghue. 2009. Climate, niche evolution, and diversification of the “bird-cage” evening primroses (Oenothera, sections Anogra and Kleinia). American Naturalist 173(2): 225–240.

Faircloth, B. C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32:786–788.

Ferrari, S. L. P and F. Cribari-Neto. 2004. Beta Regression for Modelling Rates and Proportions. Journal of Applied Statistics 31(7): 799–815.

FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution 3.6: 1084-1092.

Fitzpatrick, B. M. and M. Turelli. 2006. The geography of mammalian speciation: mixed signals from phylogenies and range maps. Evolution 60(3):601–615.

Friedman, J. 2020. The evolution of annual and perennial plant life histories: Ecological correlates and genetic mechanisms. Annual Review of Ecology, Evolution, and Systematics 51:461–481.

Goodwillie, C. 1997. Self-incompatibility and mating system evolution in *Linanthus* section Leptosiphon. University of Washington, Seattle, WA, Ph.D. dissertation.

Goodwillie, C. 1999. Multiple origins of self-compatibility in *Linanthus* section Leptosiphon (Polemoniaceae): Phylogenetic evidence from internal transcribed sequence data. Evolution 53(5): 1387–1395.

Goodwillie, C. and J. W. Stiller. 2001. Evidence for polyphyly in a species of *Linanthus* (Polemoniaceae): Convergent evolution in self-fertilizing taxa. Systematic Botany 26(2):273–282.

Goodwillie, C. and J. M. Ness. 2005. Correlated evolution in floral morphology and the timing of self-compatibility in *Leptosiphon jepsonii* (Polemoniaceae). International Journal of Plant Sciences 166(5): 741–751.

Grant, V. 1959. Natural history of the Phlox family. Vol. 1. Systematic botany. Martinus Nijhoff, The Hague, The Netherlands.

Grant, V. 1983. The systematic and geographical distribution of hawkmoth flowers in the temperate North American flora. Botanical Gazette 144:439–449.

Gray, S. M. and J. S. McKinnon. 2007. Linking color polymorphism maintenance and speciation. Trends in Ecology and Evolution 22(2): 71–79.

Grossenbacher, D. L., Veloz, S. D., and J. P. Sexton. 2014. Niche and range size patterns suggest that speciation begins in small, ecologically diverged populations in North American Monkeyflowers (Mimulus spp.). Evolution 68(5):1270–1280.

Hernández-Hernández, T., Brown, J. W., Schlumpberger, B. O., Eguiarte, L. E., and S. Magallón. 2014. Beyond aridification: Multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. New Phytologist 202(4):1382–1397.

Huelsenbeck, J. P., Nielsen, R., and J. P. Bollback. 2003. Stochastic mapping of morphological characters. Systematic Biology 52(2): 131–158.

Jacobs, S.J., Grundler, M.C., Henriquez, C.L. and Zapata F. 2021. An integrative genomic and phenomic analysis to investigate the nature of plant species in *Escallonia* (Escalloniaceae). Scientific Reports 11: 24013.

Jamie, G. A. and J. I. Meier. 2020. The persistence of polymorphisms across species radiations. Trends in Ecology and Evolution 35(9): 795–808.

Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. C. Zerega, and N. J. Wickett. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Applications in Plant Science 4:1–7.

Johnson, M. G., L. Pokorny, S. Dodsworth, L. R. Botigué, R. S. Cowan, A. Devault, W. L. Eiserhardt et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. Systematic Biology 68:594–606.

Jolivet, P. and J. W. Foley. 2015. Solutions for purifying nucleic acids by solid-phase reversible immobilization (SPRI). Ludmer Centre for Neuroinformatics and Mental Health, 2.

Junier, T. and E. M. Zdobnov. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics 26:1669–1670.

Katoh, K., G. Asimenos, and H. Toh. 2009. Multiple Alignment of DNA Sequences with MAFFT. Pp. 39–64 *in* D. Posada, ed. Bioinformatics for DNA Sequence Analysis. Totowa, NJ: Humana Press.

Knudsen, J. T. and L. Tollsten. 1993. Trends in floral scent chemistry in pollination syndromes : floral scent composition in moth-pollinated taxa. Botanical Journal of the Linnean Society 113(3):263–284.

Kraft, N. J. B., B. G. Baldwin, and D. D. Ackerly. 2010. Range size, taxon age and hotspots of neoendemism in the California flora. Diversity and Distributions 16:403–413.

Landis, J.B. 2016. Evolutionary development of flower variation in Polemoniaceae : Linking cellular phenotypes, genetics, and pollinator shifts. University of Florida, PhD Dissertation.

Landis, J. B., Bell, C. D., Hernandez, M., Zenil-Ferguson, R., McCarthy, E. W., Soltis, D. E., and P. S Soltis. 2018. Evolution of floral traits and impact of reproductive mode on diversification in the phlox family (Polemoniaceae). Molecular Phylogenetics and Evolution 127: 878–890.

Larridon, I., Villaverde, T., Zuntini, A. R., Pokorny, L., Brewer, G. E., Epitawalage, N., Fairlie, et al. 2020. Tackling rapid radiations with targeted sequencing. Frontiers in Plant Science 10: 1–17.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Systematic Biology 50(6):913–925.

Lichter-Marck, I. H., W. A. Freyman, C. M. Siniscalchi, J. R. Mandel, A. Castro-Castro, G. Johnson, and B. G. Baldwin. 2020. Phylogenomics of *Perityleae* (Compositae) provides new insights into morphological and chromosomal evolution of the rock daisies. Journal of Systematics and Evolution 58:853–880.

Lichter-Marck, I. H. and B. G. Baldwin. 2023. Edaphic specialization onto bare, rocky outcrops as a factor in the evolution of desert angiosperms. Proceedings of the National Academy of Sciences 120(6): e2214729120.

Losos, J. B. and R. E. Glor. 2003. Phylogenetic comparative methods and the geography of speciation. Trends in Ecology and Evolution 18(5): 220–227.

Mabry, M. E. and M. G. Simpson. 2018. Evaluating the Monophyly and Biogeography of *Cryptantha* (Boraginaceae). Systematic Botany 43:53–76.

Maddison W. P., Midford P. E., and S.P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. Systematic Biology 56: 701-710.

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

Mayr, E. 1959. Isolation as an evolutionary factor. Proceedings of the American Philosophical Society, 103(2), 221-230.

McLay, T.G., Birch, J.L., Gunn, B.F., Ning, W., Tate, J.A., Nauheimer, L., Joyce, et al. 2021. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. Applications in Plant Sciences 9(7).

Millennium Ecosystem Assessment. 2005. Ecosystems and human well-being: Desertification synthesis. Washington, DC.

Miller, E. T., Leighton, G. M., Freeman, B. G., Lees, A. C., and R. A. Ligon 2019. Ecological and geographical overlap drive plumage evolution and mimicry in woodpeckers. Nature Communications 10(1).

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

Mooney, H. and E. Zavaleta. 2016. Ecosystems of California. Univ of California Press.

Moore, M. J. and R. K. Jansen. 2006. Molecular evidence for the age, origin, and evolutionary history of the American desert plant genus *Tiquilia* (Boraginaceae). Molecular Phylogenetics and Evolution 39:668–687.

Moran, R. 1977. New or renovated Polemoniaceae from Baja California, Mexico (*Ipomopsis, Linanthus, Navarretia*). Madroño 24:141–159.

Nute, M., Chou, J., Molloy, E. K., and T. Warnow. 2018. The performance of coalescent-based species tree estimation methods under models of missing data. BMC Genomics, 19(Suppl 5):1–22.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Patterson, R. W. and J. M. Porter. In prep. *Linanthus*. For: Flora of North America Editorial Committee, eds. 1993+. Flora of North America North of Mexico. 25+ vols. New York and Oxford. Vol. 15.

Patterson, R. W. and J. M. Porter. 2021. *Linanthus*, in Jepson Flora Project (eds.) Jepson eFlora, Revision 9, https://ucjeps.berkeley.edu/eflora/eflora_display.php?tid=8871, accessed on April 15, 2023.

Pearman, P. B., T. S. Alioto, J. R. P. Trotta, and J. T. Columbus. 2021. Genotyping-by-sequencing resolves relationships in Polygonaceae tribe *Eriogoneae*. Taxon 70:826–841.

Peng, J., D. L. Swofford, and L. Kubatko. 2022. Estimation of speciation times under the multispecies coalescent. Bioinformatics 38:5182–5190.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7.

Porter, J. M., L. A. Johnson, and D. Wilken. 2010. Phylogenetic systematics of Ipomopsis (Polemoniaceae): Relationships and divergence times estimated from chloroplast and nuclear DNA sequences. Systematic Botany 35:181–200.

Porter, J. M. and R. W. Patterson. 2015. A fistful of Polemoniaceae : New names and combinations. Aliso 32.

Porter, M. and L. A. Johnson. 2000. A phylogenetic classification of Polemoniaceae. Aliso 19:55–91.

Rausher, M. D. 2008. Evolutionary transitions in floral color. International Journal of Plant Sciences 169(1): 7–21.

Raguso, R. A. and M. A. Willis. 2002. Synergy between visual and olfactory cues in nectar feeding by naive hawkmoths, *Manduca sexta*. Animal Behaviour 64:685–695.

Raguso, R. A., Levin, R. A., Foose, S. E., Holmberg, M. W., and L. A. McDade. 2003. Fragrance chemistry, nocturnal rhythms and pollination "syndromes" in *Nicotiana*. Phytochemistry 63(3): 265–284.

Revell, L. J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223.

Ricklefs R.E. and S.S. Renner. 1994. Species richness within families of flowering plants. Evolution 48(5):1619–36.

Rigby, R. A. and D. M. Stasinopoulos. 2005. Generalized Additive Models for Location, Scale and Shape. Journal of the Royal Statistical Society Series C: Applied Statistics 54(3):507–554.

Rohland, N. and D. Reich. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Research 22(5):939-46.

Rose, J. P., R. Kriebel, L. Kahan, A. DiNicola, J. G. González-Gallegos, F. Celep, E. M. Lemmon, et al. 2021. Sage insights into the phylogeny of *Salvia*: Dealing with sources of discordance within and across genomes. Frontiers in Plant Science 12:1–14.

Rose, J. P. and K. J. Sytsma. 2021. Complex interactions underlie the correlated evolution of floral traits and their association with pollinators in a clade with diverse pollination systems. Evolution 75(6):1431–1449.

Rubin B.E.R., Ree R.H., and C.S. Moreau. 2012. Inferring phylogenies from RAD sequence data. PLoS One 7:e33394.

Rydberg, P. 1906. Studies on the Rocky Mountain flora. XVI. Bulletin of the Torrey Botanical Club. 33: 149.

Sapir, Y., Gallagher, M. K., and E. Senden. 2021. What maintains flower colour variation within populations? Trends in Ecology and Evolution 36(6):507–519.

Schemske, D. W. and P. Bierzychudek. 2001. Perspective: Evolution of flower color in the desert annual *Linanthus parryae*: Wright revisited. Evolution 55(7): 1269–1282.

Schemske, D. W. and P. Bierzychudek. 2007. Spatial differentiation for flower color in the desert annual *Linanthus parryae*: Was Wright right? Evolution 61:2528–2543.

SEINet. 2022. Southwest Environmental Information Network. http//: swbiodiversity.org/seinet/index.php. (last accessed 12 October 2022)

Silberbauer-Gottsberger, I. and G. Gottsberger. 1975. Über sphingophile Angiospermen Brasiliens / About Sphingophilous Angiosperms in Brazil. Plant Systematics and Evolution 123:157–184.

Simpson, M. G., C. Matt Guilliams, K. E. Hasenstab-Lehman, M. E. Mabry, and L. Ripma. 2017. Phylogeny of the popcorn flowers: Use of genome skimming to evaluate monophyly and interrelationships in subtribe Amsinckiinae (Boraginaceae). Taxon 66:1406–1420.

Singhal, S., Roddy, A. B., DiVittorio, C., Sanchez-Amaya, A., Henriquez, C. L., Brodersen, C. R., Fehlberg, S., and F. Zapata. 2021. Diversification, disparification, and hybridization in the desert shrubs Encelia. New Phytologist 230(3): 1228-1241.

Skeels, A. and M. Cardillo. 2019. Reconstructing the geography of speciation from contemporary biodiversity data. American Naturalist 193(2): 240–255.

Slimp, M., Williams, L. D., Hale, H., and M. G. Johnson. 2021. On the potential of Angiosperms353 for population genomic studies. Applications in Plant Sciences 9(7): 1–13.

Smith, S. D. and E. E. Goldberg. 2015. Tempo and mode of flower color evolution. American Journal of Botany 102(7): 1014–1025.

Smith, S. A. and J. M. Beaulieu. 2009. Life history influences rates of climatic niche evolution in flowering plants. Proceedings of the Royal Society B: Biological Sciences 276:4345–4352.

Smith, S. A. and M. J. Donoghue. 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322:86–89.

Soltis, D. E., M. E. Mort, M. Latvis, E. V. Mavrodiev, B. C. O'meara, P. S. Soltis, J. G. Burleigh, and R. R. De Casas. 2013. Phylogenetic relationships and character evolution analysis of Saxifragales using a supermatrix approach. American Journal of Botany 100:916–929.

Spencer, C. and J. M. Porter. 1997. Evolutionary Diversification and Adaptation to Novel Environments in *Navarretia*. Systematic Botany 22:649–668.

Stebbins, G. L. 1952. Aridity as a stimulus to plant evolution. American Naturalist 86:33–44.

Stebbins, G. L. 1970. Adaptive radiation of reproductive characteristics in Angiosperms, I: Pollination mechanisms. Annual Reviews of Ecology and Systematics 1:307–326.

Strauss, S. Y. and J. B. Whittall. 2006. Non-pollinator agents of selection on floral traits. Pages 120–138 in L. D. Harder and S.C.H. Barrett, eds. Ecology and evolution of flowers. Oxford University Press, Oxford.

Swofford, D. L. 2002. PAUP. Phylogenetic analysis using parsimony (and other methods). Version 5. Sinauer Associates, Sunderland, Massachusetts.

Tanaka, Y., Sasaki, N., and A. Ohmiya. 2008. Biosynthesis of plant pigments: Anthocyanins, betalains and carotenoids. Plant Journal 54(4):733–749.

Tank, D. C. and R. G. Olmstead. 2008. From annuals to perennials: Phylogeny of subtribe Castillejinae (Orobanchaceae). American Journal of Botany 95:608–625.

Taylor R. S. and V. L. Friesen. 2017. The role of allochrony in speciation. Molecular Ecology. 26(13):3330-3342.

Thorne, R. F. 1986. A historical sketch of the vegetation of the Mojave and Colorado Deserts of the American Southwest. Annals of the Missouri Botanical Garden 73:642–651.

Thornhill, A. H., B. G. Baldwin, W. A. Freyman, S. Nosratinia, M. M. Kling, N. Morueta-Holme, T. P. Madsen, D. D. Ackerly, and B. D. Mishler. 2017. Spatial phylogenetics of the native California flora. BMC Biology 15:1–18.

Tripp, E. A. and P. S. Manos. 2008. Is floral specialization an evolutionary dead-end? Pollination system transitions in *Ruellia* (Acanthaceae). Evolution 62(7): 1712–1737.

Ushimaru, A. and K. Nakata. 2002. The evolution of flower allometry in selfing species. Evolutionary Ecology Research 8(4):1217–1227.

Vaidya, P., McDurmon, A., Mattoon, E., Keefe, M., Carley, L., Lee, C. R., Bingham, R. and J. T. Anderson. 2018. Ecological causes and consequences of flower color polymorphism in a self-pollinating plant (*Boechera stricta*). New Phytologist 218(1): 380–392.

Vasile, M. A., J. Jeiter, M. Weigend, and F. Luebert. 2020. Phylogeny and historical biogeography of Hydrophyllaceae and Namaceae, with a special reference to *Phacelia* and *Wigandia*. Systematics and Biodiversity 18:757–770.

Verity, R. and R. A. Nichols. 2016. Estimating the number of subpopulations (K) in structured populations. Genetics 203(4):1827–1839.

Walden, G. K., L. M. Garrison, G. S. Spicer, F. W. Cipriano, and R. W. Patterson. 2014. Phylogenies and chromosome evolution of *Phacelia* (Boraginaceae: Hydrophylloideae) Inferred from nuclear ribosomal and chloroplast sequence data. Madroño 61:16–47.

Warren, J. and S. Mackenzie. 2001. Why are all colour combinations not equally represented as flower-colour polymorphisms? New Phytologist 151(1):237–241.

Wright, S. 1943. An analysis of local variability of flower color in *Linanthus parryae*. Genetics 28:139–156.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:15–30.

Chapter 2

Linanthus floral scent is highly variable yet differentiates species

Abstract

Floral scent is a complex trait, with species emitting compounds that are both phylogenetically conserved and species specific. Most angiosperm species produce tens of volatile organic compounds (VOCs) in their floral scent bouquet, with certain compounds involved in biotic communication while others may be an artifact of non-selective forces. In species that depend on pollinators for fertilization, floral scent can be crucial for reproduction. Understanding the variation of floral scent across closely related species can help us understand how the fragrance phenotype contributes to reproductive isolation and reproductive success. In this study, we conducted a large-scale investigation of floral scent variation across 13 species of *Linanthus* (Polemoniaceae). More than half of the species in the genus coexist locally, potentially with scent causing reproductive isolation for sympatric species. *Linanthus* species are visited by a wide variety of insects, including flies, butterflies, bee-flies, bees, and beetles, and this pollinator variation may reflect the scent variation across the genus. We found that species emit a wide variety of floral compounds, ranging from 3 to 38 compounds per species. Species showed significant differences in scent profile, though there was high intraspecific variation. In constrained ordination visualizations, scent profile was visually differentiated by species only within sub-generic clades. Hierarchical clustering with species emissions averages showed that the species fragrance dendrogram was incongruent with the phylogenetic relationships in

Linanthus. Closely related species having a very distinct scent profile may indicate that scent is used as a reproductive isolation mechanism between closely related species with range overlap.

Introduction

Plants guide pollinators to their flowers through multiple sensory channels, including color, UV markings, shape, size, and fragrance (Raguso 2008a). While studies focused on visual characteristics of flowers, such as color and shape, have dominated pollination biology, odor-mediated pollination is emerging as one of the most important dimensions of floral signaling (Raguso 2008a). Scent plumes can direct a pollinator spatially, attracting it from a distance, guiding it to approach a flower, encouraging it to land inside a corolla, and initiating feeding cues (reviewed in Raguso 2008b). Floral scent can effectively signal the availability of nectar or pollen (Theis & Raguso 2005) in flowers at peak reproductive maturity (Theis et al 2007). It can promote outcrossing by attracting specialized pollinators and promoting floral constancy (Raguso 2008b), or by decreasing pollinators' visitation time and encouraging them to visit different individuals (Kessler & Baldwin 2007). Consequently, attracting pollinators via floral fragrances can increase seed production and have a direct positive effect on plant fitness (Majetic et al. 2009).

Flowers emit an incredible diversity of volatile organic compounds (VOC) that make up floral fragrances, with over 1700 compounds identified so far in angiosperms (Knudsen et al. 2006). These compounds belong to a few chemical classes because the major biosynthetic pathways that produce these compounds are ubiquitous (Dudareva et al 2020). They share biochemical

origins, and the same pathway can produce different compounds with little metabolic investment (Pichersky & Raguso 2018). The relative ease of producing volatile compounds means that individual species can emit up to 60 distinct compounds (Knudsen & Gershenzon 2020). This large number of compounds can attract different pollinators or change suites of pollinators in unpredictable environments (Pichersky & Raguso 2018). Across species within a genus, most taxa emit unique floral volatile profiles (Knudsen & Gershenzon 2020), and as such interspecific differences in scent can help to establish or maintain reproductive isolation (Waelti et al. 2007, Bischoff et al. 2014, Byers et al. 2014). Scent across closely related species or between populations can be strikingly different (Friberg et al. 2019), sometimes differentiating species in sympatry (Knudsen 1999, Peakall & Whitehead 2014, Okamoto et al. 2015). While variation of scent is phylogenetically and taxonomically informative in some clades (Raguso et al. 2006, Toth et al. 2016, Couto et al 2024), differences in scent profiles are most often correlated with differences in pollinators or in pollination mode (Clifford 2017, Wang et al 2018). Overall, the fragrance bouquet of a plant species is a labile and complex trait that is closely linked to pollinator signaling and can facilitate reproduction when other mechanisms of attracting pollinators are limited.

Plants with patchy distributions rely on floral scent to attract pollinators (Schatz 1990; Knudsen et al. 1999). Desert plants are often distributed unevenly across the landscape, with populations highly isolated from each other and occurring in low density, especially in dry years (Salguero-Gomez et al 2012). In this type of environment, floral scent can attract insects that detect scent

plumes from a distance and land on the source of the attractive fragrance (Cunningham et al. 2004). In addition, desert annuals depend on irregular rainfall to flower and complete their life cycle. Desert and tropical forest species are similar in their unpredictable flowering episodes and discontinuous distribution in space (Polis 1991), and scent has been shown to provide a cue for pollinators to visit tropical plant flowers (Grison-Pigé et al. 2002). Desert annuals also have a short growing season in which they must reproduce efficiently and quickly before conditions change, and scent can increase the chances of reproductive success in this critical period. Taken together, these features of desert living make effective pollinator signaling critical for desert plant species reproduction. Desert plants likely rely on scent for pollinator attraction, and reproductive isolation from congeners may be mediated by scent.

Linanthus is a flowering plant genus of 25 species that occur across the deserts and coastal areas of California and in the Great Basin. Many *Linanthus* species apparently have unique and recognizable scent profiles, yet this pattern of variation has not been well-characterized. Species are pollinated by a variety of insects including hawkmoths, bee-flies, bees, flies, and beetles with species attracting both generalist and specialist pollinators (Grant and Grant 1965; pers. obs). Because several species pairs of *Linanthus* co-flower temporally and spatially, it is likely that floral scent plays an important role in reproductive isolation and the maintenance of species boundaries. Scent has only been studied in one species of *Linanthus*. Fragrance composition between day and night blooming subspecies of *L. dichotomus* shows significant differences (Chess et al. 2008). The night blooming subspecies has higher concentrations of lilac

aldehydes, a monoterpene that attracts moths (Knudsen and Tollsten 1993), while the day blooming subspecies emits more phenylacetaldehyde known to attract a variety of insect visitors (Huber et al. 2005, Blight et al. 1997, Theis 2006). This study illustrated that scent variation is heritable in *Linanthus* and that certain compounds may be correlated with a switch from specialized moth pollination in the night blooming subspecies, to generalized pollination in the day blooming subspecies. These subspecies' scent variation pattern may indicate that floral scent in *Linanthus* is an evolutionary labile trait that can quickly respond to selective pressures, leading to local adaptation and reproductive isolation. Aside from this work on *L. dichotomus*, there has been no formal study quantifying scent variation in the genus.

This study is the first comprehensive comparative study describing variation in floral scent between and within species of *Linanthus*. We describe the variation in floral scent across species of *Linanthus* to test the hypothesis that floral profiles are associated with speciation. Our sampling includes samples within and between species to quantify scent composition across populations and species. Because floral scent likely plays a central role in plant-pollinator interactions, plant reproductive isolation, and the maintenance of species boundaries, we make the following predictions. First, we predict that species will emit many compounds, as indicated by the diversity of pollinators observed visiting *Linanthus* species. Second, we expect higher inter than intraspecific scent variation. Third, species scent profiles will differ significantly from each other.

Methods

Floral volatile collection

Sampling included 2 to 23 individuals from 15 of the 25 species of *Linanthus*. In total, we included 147 samples from 61 populations, including 108 floral samples and 39 ambient samples as controls (Supplemental Table 1A). We collected floral scent using a dynamic headspace method (Raguso and Pellmyr 1998; Tholl et al 2006) from plants grown in the greenhouse or in the field. To fit the size of the inflorescence, we cut non-reactive resin oven bags (Reynolds, Richmond, VA) into a 12 cm x 20 cm rectangle, resealed them on three edges using an impulse heat sealer FS-200 (Metronic, Pomona, CA), and made a slit into one of the sealed corners. We placed these headspace bags around an inflorescence and secured one end of the bag around the plant stem using a twist tie. We inserted the trap into the headspace through the slit and secured it in place with another twist tie. We connected scent traps to a micro air sampler (product #PAS-500, Spectrex, Redwood City, CA, USA) via Tygon tubing (4mm inner diameter) and continuously extracted air from the bag through a small slit at one of the corners of the bag where the scent trap was introduced and secured into the bag with a twist tie. We constructed scent traps by cutting clear 2.5 mm inner diameter glass tubing to 3 cm length, then plugged it with glass wool, filled with 6.5 mg of Tenax GC (product #21009-U, SUPELCO, Bellefonte, PA, USA; mesh size 80-100), and plugged the other end with glass wool. The micro air sampler pulled air from the headspace through the scent trap at 200 mL/min for 1 hour. We eluted scent traps with 200 uL of GC-MS-quality hexane into conical glass inserts (Agilent Technologies, Santa Clara, CA, USA) held in 2 mL auto-sampler vials (Agilent

Technologies). Auto-sampler vials were capped with screw cap PTFE/silicone lids (Agilent Technologies) and stored in a cooler in the field (if applicable), then at -20C until analyzed. Control samples were collected simultaneously in the same manner from oven bags filled with ambient air.

Gas-chromatography/mass spectrometry (GC/MS) analysis

We analyzed the samples by GC/MS using an Agilent 7890B GC with an HP-5 Ultra Inert non-polar capillary column (30 m length x 0.25 mm inner diameter x 0.25 um film thickness) and an Agilent 5977A mass spectrometer in Dr. Santiago Ramirez's lab at University of California, Davis. The autosampler extracted 1uL of sample and injected it in splitless mode. The GC oven temperature was held at 40 C for 2 mins, then ramped up to 210 C at a rate of 5 C/min, then ramped up to 250 C at a rate of 20 C/min, and held for 5 min (Karban et al 2014). Helium was used as a carrier gas and set to flow at 2 mL/min. The volatile peaks were automatically integrated using MassHunter GC/MS Acquisition software vB.07.00 (Agilent) and MSD ChemStation Enhanced Data Analysis Software vF.01.00 (Agilent).

We preliminarily identified compounds using the NIST05 mass spectral database and the NIST MS Search software v2.0. We verified compound identification by comparing their retention times (RT) in minutes to the first decimal to those of authentic standards where possible, and by calculating the Kovats Retention Indexes (KRI) and comparing them to compounds with published KRI. For the latter, we ran a C8-C30 alkane ladder using the same GC/MS equipment and settings as the samples and used the retention times of the n-alkanes to calculate the KRI.

We compared the calculated RI with those in the searchable databases NIST webbook (https://webbook.nist.gov/) and Pherobase (https://www.pherobase.com; El-Sayed 2024). Compounds that we were unable to identify confidently are listed by the first ten mass spectrum ions in descending order of abundance (Eisen et al 2022b).

Statistical analysis

We analyzed the scent data using the retention time rounded to the nearest decimal as compound ID. We used the area under the peak for each retention time as the emission quantity. We did not use an internal standard and cannot calculate the concentration of each compound, but we calculated the area under the peak using the GC/MS, which is proportional to the concentration. We filtered the data using the R pipeline bouquet (Eisen et al. 2022). We included compounds present in at least 10% of individuals within a species, with a minimum peak area of 40,000 units, and retention time between 4 and 40 minutes, a range at which floral volatile compounds are most likely to be eluted in the column. To pass our filtering, compound emission had to be three times greater in floral versus ambient samples or the floral versus ambient samples had to be significantly different with a p-value less than 0.05. The final matrix was filtered to 13 species, 76 samples and 44 volatile compounds from the initial 15 species, 108 samples and 392 compounds in the raw dataset (Supplemental Table 1B).

For multivariate analysis, we calculated Bray-Curtis dissimilarities between every pair of individuals. We then used this dissimilarity matrix to run four analyses to examine differences in scent profiles across species. First, we ran a permutational multivariate analysis of variance

(PERMANOVA) using the 44 volatile compounds as the dependent variable and species as factors. We used 999 permutations to test the significance of the model. To run the PERMANOVA, we used the adonis function in the R package vegan (Oksanen et al. 2018). Second, we used an unconstrained ordination method, non-metric multidimensional scaling (nMDS), to examine overall patterns of variation across the entire dataset as well as any differences in within-group variability and spread. To run this analysis, we used the metaMDS function in the vegan package (Oksanen et al. 2018) reducing the dimensions to two and using a minimum of 200 and a maximum of 500 random starting points in search of a stable solution. Reduction of dimensionality to three or four dimensions instead produced qualitatively similar conclusions (not shown). Third, we used a constrained ordination method, canonical analysis of principal coordinates (CAP, Anderson & Willis 2003) using the capscale function in the R package vegan (Oksanen et al. 2018). In contrast to the unconstrained ordination, CAP uses the information about group assignment to find the dimensions that maximize the among- to within-group variation. To test for differences among species, we used the CAP axis determined automatically by the capscale function and tested the significance of the model using the anova.cca function in the vegan package with 999 permutations (Legendre et al. 2011). Lastly, using both individual emissions as well as average emissions per species, we ran a hierarchical cluster analysis with the default clustering method of complete merging.

We used two approaches to analyze the data. First, we analyzed scent variation across all species and individuals. Second, to account for phylogenetic relationships and interpret

variation among closely related species, we partitioned the matrix into four submatrices corresponding to well-supported groups of closely related species (i.e., clades) identified with genomic data (Anghel et al. 2024; Fig. S2.1). Clade C is the coastal species clade and includes *L. concinuus, L. dianthiflorus*, and *L. orcuttii*. Clade D is the desert species clade and includes *L. bernardinus, L. demissus, L. killippii* and *L. parryae*. Clade N is the night blooming clade and includes *L. dichotomus* and *L. bigelovii*. Clade O is the early diverging grade and includes *L. campanulatus, L. filiformis, L. inyoensis* and *L. maculatus*.

Results

Across 13 species of *Linanthus*, we detected 44 floral scent compounds. On average, we found 18 compounds emitted per species ranging from three compounds in *L. demissus*, to 38 compounds in *L. parryae*. Individuals had between one and 27 compounds (Supplemental Table 1B-C). The most broadly distributed compound was an unidentified compound (RT 32.6) that was present in all but one of the 13 species. The next most common compounds were the sesquiterpene aromadendrene (RT 23.6) and an unidentified compound (RT 32.1), both of which were present in 11 of the 13 species each.

Scent variation was significantly different across *Linanthus* (PERMANOVA species: R^2=0.33391, $F_{12,63}$=2.6318, p=0.001), with interspecific differences accounting for 33% of the overall variation. Interspecific scent variation within clades was significantly different in the desert annuals clade (PERMANOVA clade D: R^2=0.14328, $F_{3,34}$=1.5051, p=0.011), annual night-blooming clade (PERMANOVA clade N: R^2=0.15876, $F_{1,12}$=2.076, p=0.032), and early diverging grade

(PERMANOVA clade O: R^2=0.35147, $F_{3,12}$=2.1678, p=0.004), but not in the coastal and Baja California annuals clade (PERMANOVA clade C: R^2 =0.18836, $F_{2,11}$=1.2764, p=0.159).

The nMDS visualization showed extensive intra and interspecific variation, with notable overlap between species and clades (Fig. 2.1). The first three constrained axes resulting from the CAP analysis were statistically distinguishable from random variation (ANOVA $F_{1,74}$=7.9905, p=0.001; $F_{1,74}$ =4.2473 , p=0.006; $F_{1,74}$ =3.6969, p=0.016 respectively). The dimensions of scent compounds between species were significantly different (ANOVA $F_{12,63}$=2.5055, p=0.001). The visualization of the CAP dimensions along axes 1 and 2, 2 and 3, and 3 and 4 showed a lot of overlap in the reduced multidimensional space (Fig. 2.2A-C). Results of the within clades CAP analysis showed marked interspecific differentiation, although species still showed intraspecific variation (Fig. 2.3A-D).

The hierarchical clustering analysis using all samples did not recover clusters matching taxonomic species (Fig. 2.4A). The heatmap visualizing the relationships between clusters of samples and clusters of compound emission showed that some samples clustered with other conspecifics (e.g., *L. parryae* in Fig. 2.4A), and most species did not cluster together cohesively. Some samples emitted many compounds, while others just a few (Fig. 2.4A). The dendrogram visualizing the relationships between clusters of species average emissions and clusters of compounds did not reflect phylogenetic relationships, either by clades or species (Fig. 2.4B, S1). Closely related species did not cluster together in chemospace. The exception were the species

in the early diverging grade, which formed one cluster in the dendrogram, also including *L. dichotomus, L. orcuttii*, and *L. parryae*.

Discussion

Linanthus flower visitors belong to various clades of insects, and many of these species are attracted by specific floral scents. *Linanthus* species have been observed to emit distinct scents that may attract particular pollinators. Here, we showed that *Linanthus* species have distinct scent profiles, especially when compared to most closely related species, which may have served as a reproductive isolation mechanism as species diversified in close geographic space. We extracted scent volatiles from 108 floral samples in 13 species representing 52% of the species in this genus. As predicted, we found a wide variation in scent compound presence and amount, and many types of compounds emitted by species, which may reflect the multiple functions of scent phenotypes.

Floral scent is variable in *Linanthus.* The pattern of scent variation shows overlap in the nMDS multidimensional space, though this is a visual effect of compressing multidimensional variation into fewer dimensions (Fig. 2.1). We found significant differences between species scent emissions. When using a constrained ordination method to find the axes of variation that maximize variation among species, we detected visual differentiation between some species (Fig. 2.2A-C). These visual differences are much more pronounced when we limit the analysis to species in the same clade (Fig. 2.3A-D), suggesting that closely related species differ markedly in floral fragrance. A similar pattern has been documented in *Streptanthus*, another annual plant

that occurs in dry areas of California, in which scent composition is highly divergent in close relatives versus more distantly related species pairs (Weber et al. 2018). These results suggest that variation in scent might facilitate reproductive isolation. Emitted scent quantity can quickly evolve when plants are exposed to different pollinators (Gervasi and Schiestl 2017). This rapid divergence leading to pollinator specialization can be a precursor to reproductive isolation and speciation, a process of angiosperm diversification known as the pollinator-shift model (Grant 1949, Stebbins 1970). It is plausible that this mechanism has facilitated the diversification of species in *Linanthus*. These species are visited by a variety of pollinators (Grant & Grant 1965, pers. obs.), and more than a third of the species co-occur sympatrically (pers. obs.). Floral scent may serve as a mechanism of reproductive isolation, especially in sympatric species that have an extremely compressed temporal span to produce seeds. Although we sampled multiple individuals per species, a more thorough investigation among populations of the same species and between sympatric species pairs and their pollinators is needed to test for reproductive isolation by floral chemistry.

The pattern of high scent compound variation across species could also be explained by a contrasting hypothesis. Instead of repeated pollinator shifts, species of *Linanthus* may attract a variety of pollinators, without specializing in a type or species of insect. Scent can play a role in the specialization-generalization spectrum, with certain floral compounds known to attract specific pollinators, a variety of insects, or generalist pollinators (Raguso 2008b). A generalist pollination approach is common in angiosperms and serves as a bet-hedging strategy

when certain pollinators are rare (Waser et al. 1996, Aigner 2001, Ohashi et al 2021). Generalization has been linked to plant species that flower unpredictably (Grison-Pigé et al. 2002) or have a scattered distribution (Schatz 1990), just like many *Linanthus* species. In *L. dichotomus* the night-blooming subspecies seems to be ancestral to the day-blooming taxon, which indicates that a shift from moth-specialization to generalized pollinator attraction has occurred in the genus (Chess et al 2008). To investigate whether scent variation across *Linanthus* is explained by a generalist or a specialist pollination strategy, future work is needed to determine the effective pollinators and evaluate the attractiveness of the floral scent to pollinators.

Another reason why floral scent may be so variable between and within species of *Linanthus* is that scent compounds may be under contrasting selective pressures. Floral scent can operate in multiple ways, with different components experiencing distinct selective pressures and potentially different evolutionary histories (Schiestl 2015). For instance, certain components of floral scent may discourage facultative pollinator visits (Junker and Blüthgen 2010) and deter florivores (Galen 1983) and nectar robbers (Galen 1983, Kessler et al. 2008), while other compounds deter individual pollinators from depleting one flower's reward (Schiestl 2015). The latter may promote outcrossing, which is likely important in isolated desert plant populations, such as in many species of *Linanthus*.

Scent volatile emissions exhibit plasticity with environmental factors, especially with high temperatures like in desert areas that *Linanthus* inhabits. Temperature can increase scent emissions concurrent with peak flowering time (Farré-Armengol et al. 2015), but high temperatures can stunt emissions past an optimal threshold (Farré-Armengol et al. 2014). Temperature can also change the relative proportion of compounds in a plant, potentially changing the pollinator's perception of the floral blend (Katzenberger et al. 2013). Different species have different optimal temperatures for floral volatile emissions (Farré-Armengol et al. 2014), adding another complexity to comparing species scent profiles. Plasticity due to temperature may have contributed to the high variation we saw between and within species in *Linanthus*. Though scent emissions may be plastic, those differences may still be ecologically functional (Delle-Vedove et al. 2017). Capturing the variation in scent across a variety of environmental factors and community contexts is capturing the real variation in scent blends of species. As climate warms and becomes less predictable, more variation in scent will likely occur because of these plastic responses. Higher scent emissions may increase the importance of fragrance as a signal in pollinator attraction (Katzenberger et al. 2013). Understanding scent emission patterns in desert plants like *Linanthus* can help us predict future scenarios in other species, and help us understand how plant reproduction will be affected by a warming world.

Scent may be a by-product of other metabolic processes. For example, pigment and scent share biosynthetic pathways, and a pleiotropy with floral color may maintain scent emissions in colored flowers without selective pressure on scent directly (Majetic et al. 2007, 2008). Indeed,

Linanthus has many color polymorphic species and pigment pleiotropy could contribute to the wide variation seen within *Linanthus* species. The repeated loss of anthocyanins across species in *Linanthus* (Anghel et al. in press), may have been a consequence of different mutations affecting various levels of the anthocyanin pathway, leading to different scent by-products. To date, few studies have explored the covariation between anthocyanin pigments, or loss of pigment, and scent produced through the same biosynthetic pathway. *Linanthus* would be an ideal system for further study of the pigment-scent correlation because 40% of species display anthocyanin-based color polymorphisms (Anghel et al. in press).

Figures

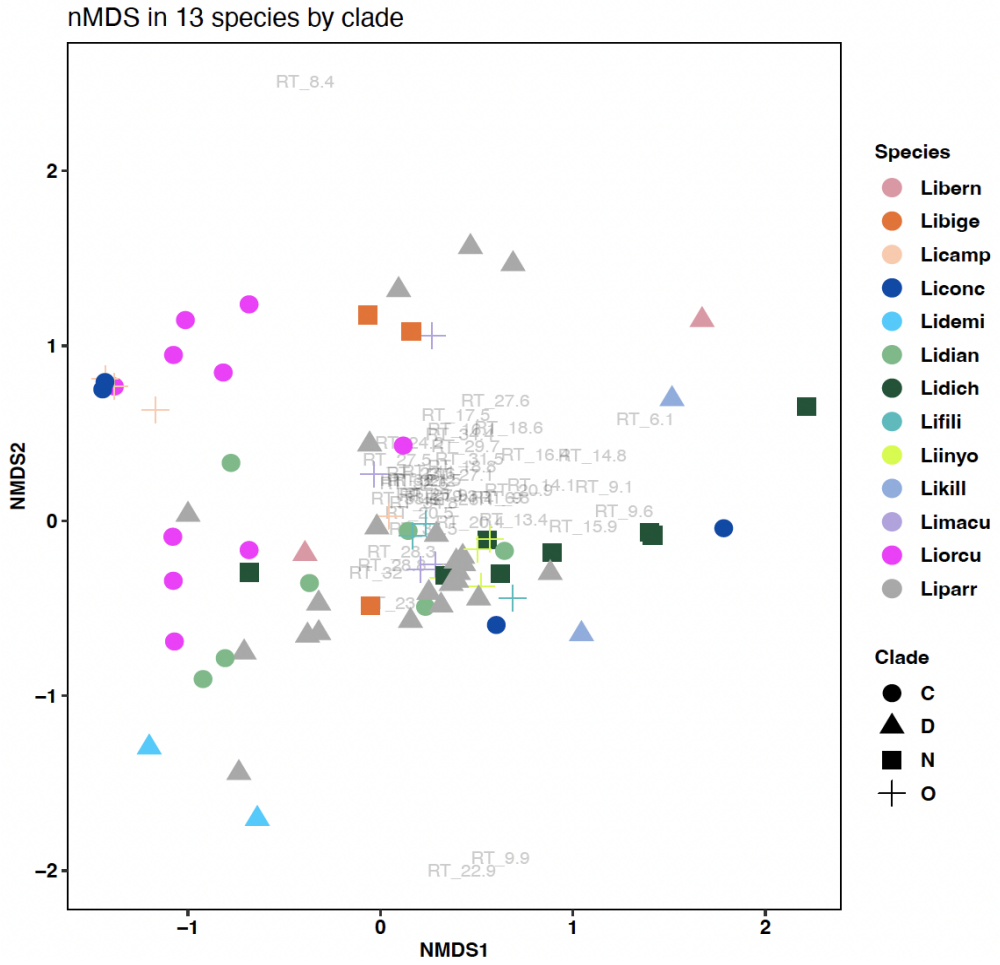


Figure 2.1. Individual scent profiles for 13 species of *Linanthus* with multiple individuals per species, in a non-metric multidimensional scaling (NMDS) of scent variation in two-dimensional space using Bray Curtis distances. We reduced the dimensions to five and obtained a stress score of 0.085. Each color is a different species, and each point shape represents a clade within the genus: C = coastal and Baja California annuals clade, D = desert annuals clade, N = annual night-blooming clade, O = early diverging grade (Fig. 2.S1).

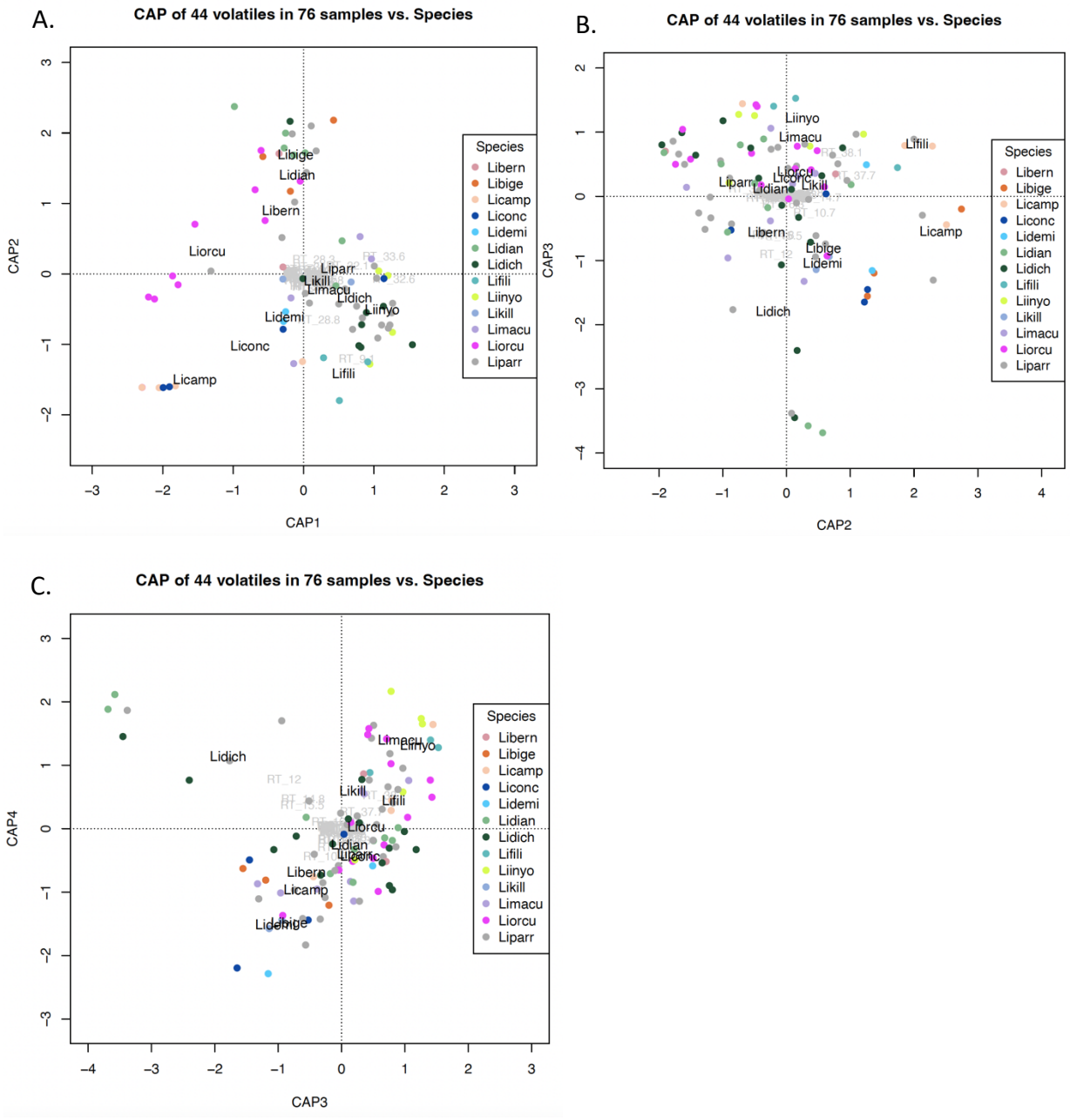


Figure 2.2. Canonical analysis of principal coordinates (CAP) constrained by species, with (A) all species along axes 1 and 2, (B) all species along axes 2 and 3, (C) all species along axes 3 and 4.

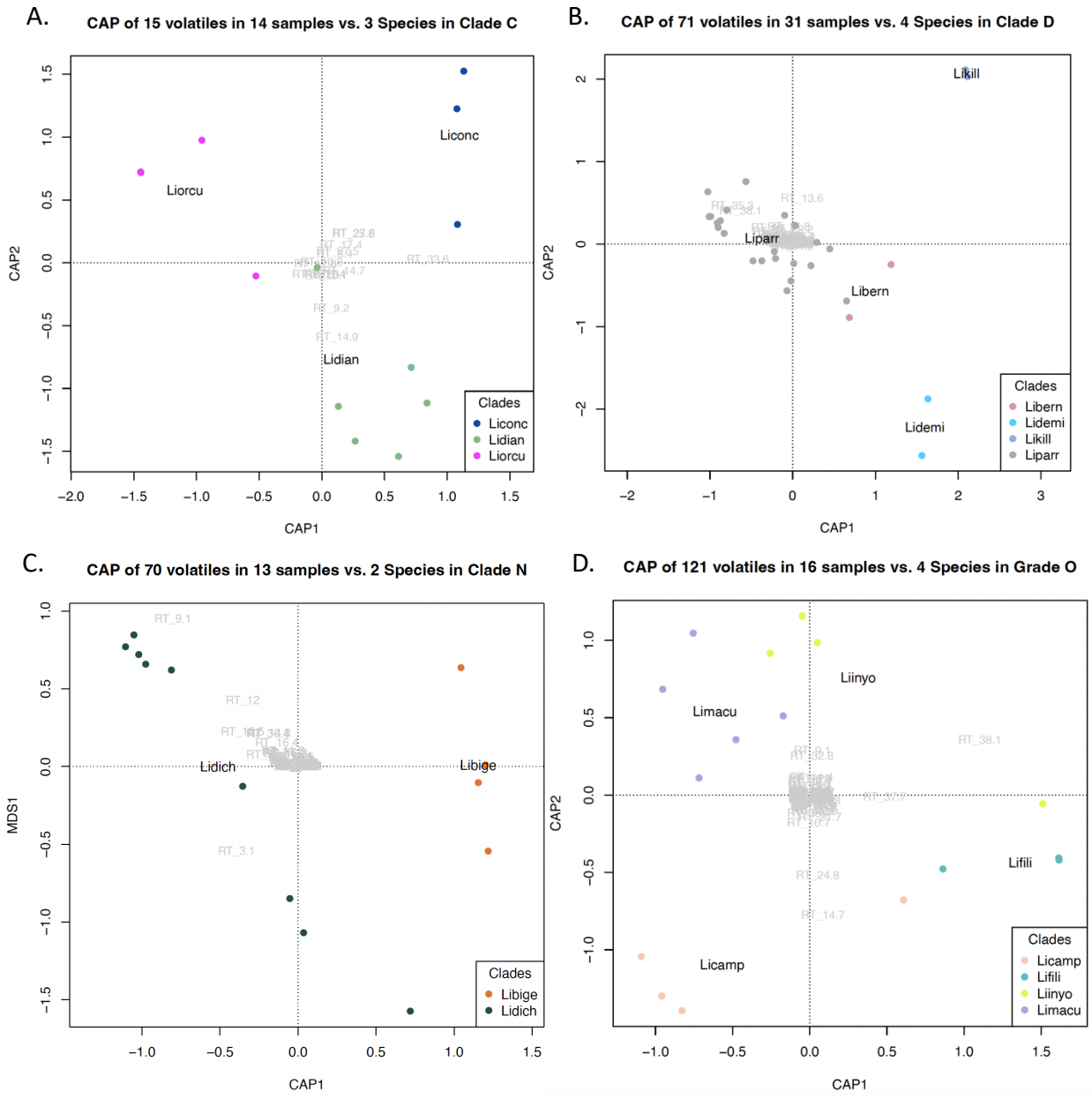
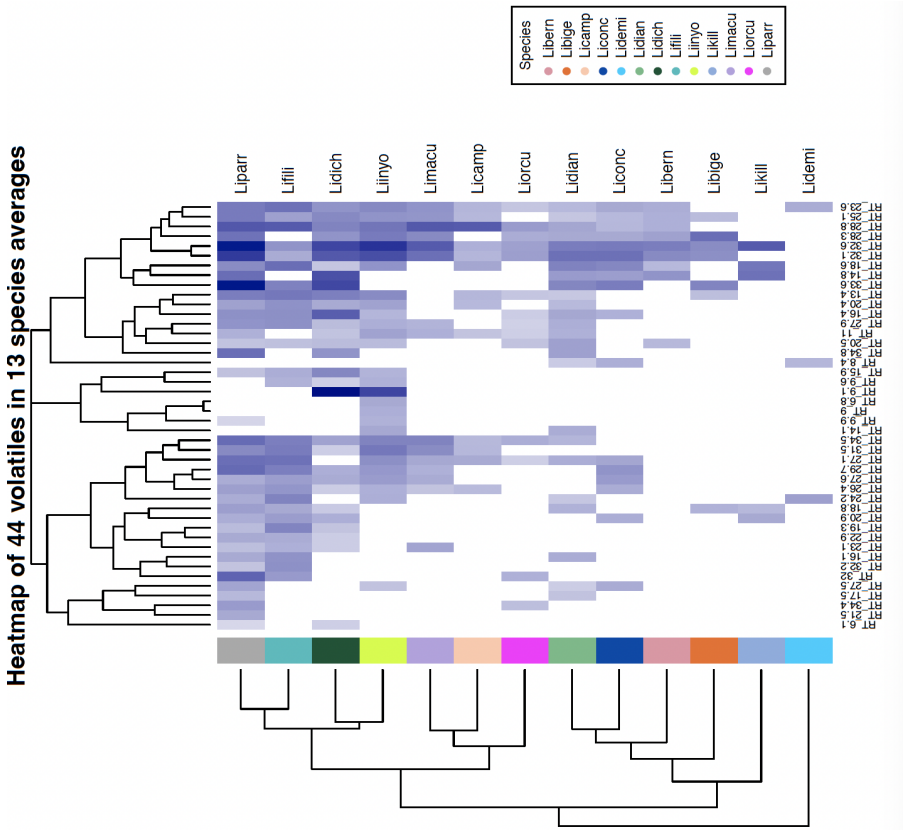


Figure 2.3. Canonical analysis of principal coordinates (CAP) constrained by species, with (A) species in clade C, (B) species in clade D, (C) species in clade N, and (D) species in grade O. Results of the within clades CAP analysis visually showed interspecific differentiation, although species still showed intraspecific variation.


Heatmap of 44 volatiles in 13 species averages
Species
Libern
Libige
Licamp
Liconc
Lidemi
Lidian
Lidich
Lifili
Linyo
Likill
Limacu
Liorcu
Liparr
Liparr
Lifili
Lidich
Linyo
Limacu
Licamp
Liorcu
Lidian
Liconc
Libern
Libige
Likill
Lidemi

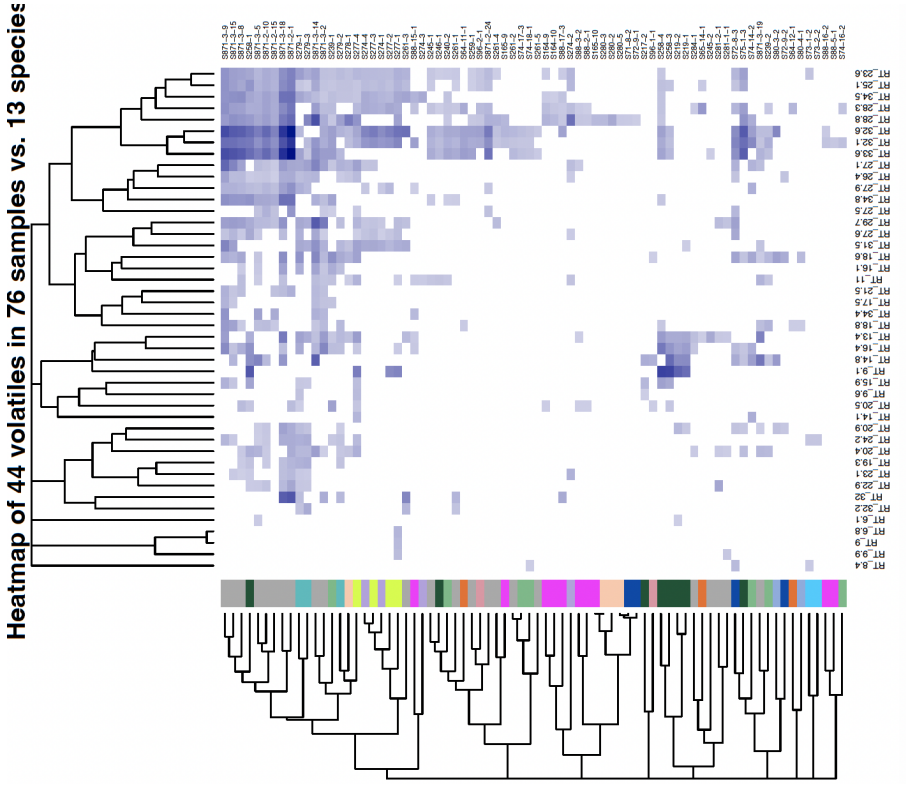

Heatmap of 44 volatiles in 76 samples vs. 13 species

Figure 2.4. Hierarchical clustering analysis of compound emissions is visualized with a dendrogram of retention times (i.e., compound) in the columns, and clusters of samples (A) or species averages (B) in the rows. The heatmap cells represent compound emission amount, with absence of compound in white and high compound emissions in dark blue. No compounds that passed filtering were shared across all samples or all species. Some samples or species emit many compounds, while others a few. (A) Samples forming chemical clusters do not match taxonomic species. While some samples cluster with conspecifics, most species do not cluster together. (B) Species forming chemical clusters do not match phylogenetic relationships (Fig. 2.S1)

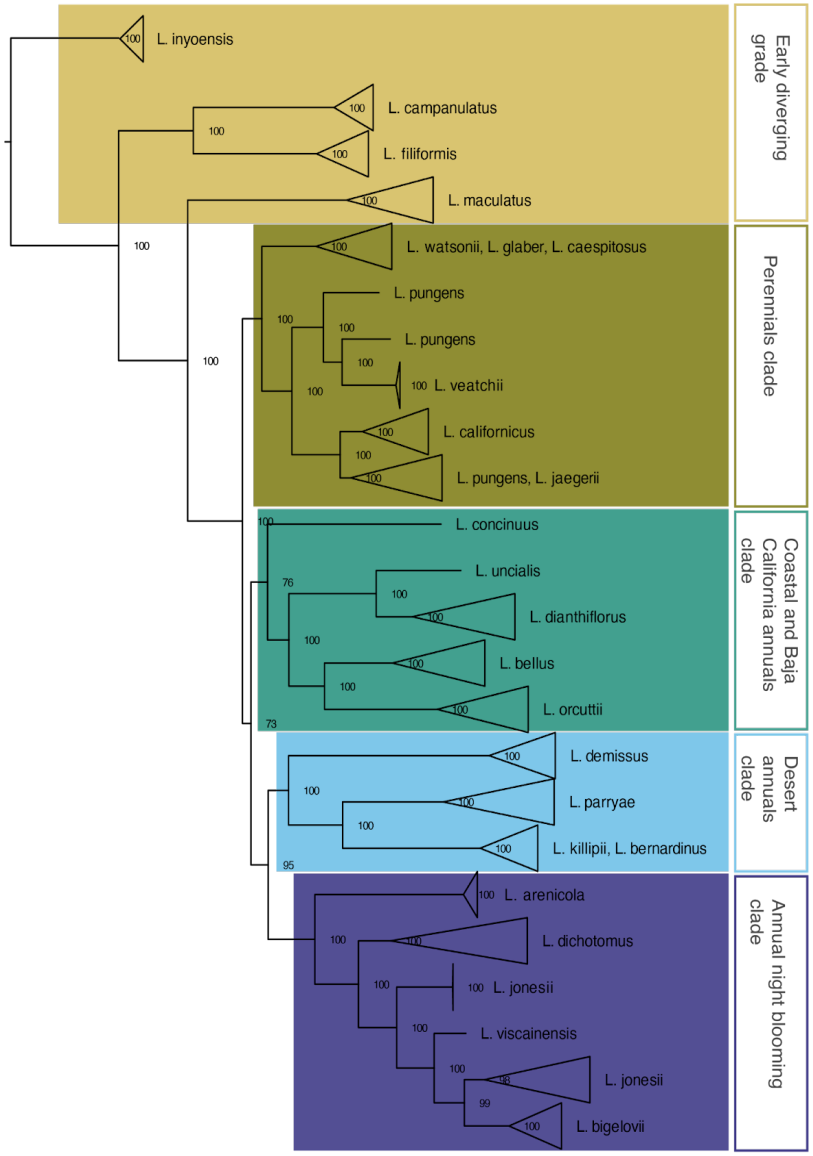


Figure S2.1. Phylogenetic relationships and major clades identified in Anghel et al. 2024. Clade C is the coastal species clade and includes *L. concinuus, L. dianthiflorus*, and *L.orcuttii*. Clade D is the desert species clade and includes *L. bernardinus, L. demissus, L. killippii* and *L. parryae*. Clade N is the nightblooming clade and includes *L. dichotomus* and *L. bigelovii.* Clade O is the early diverging grade and includes *L. campanulatus, L. filiformis, L. inyoensis* and *L. maculatus*.

References

Aigner, P. A. (2001). Optimality modeling and fitness trade-offs: When should plants become pollinator specialists? Oikos, 95(1), 177–184. https://doi.org/10.1034/j.1600-0706.2001.950121.x

Anderson, M. J., & Willis, T. J. (2003). Canonical Analysis of Principal Coordinates: A useful method of Constrained Ordination for Ecology. Ecology, 84(2), 511–525.

Anghel, I.G., Lichter-Marck, I., Smith, L. & Zapata, F. (2024) Phylogenomics of the North American Desert Radiation Linanthus (Polemoniaceae) Reveals Mixed Trait Lability and No Single Geographic Mode of Speciation (manuscript in review).

Baughman, O. W., Agneray, A. C., Forister, M. L., Kilkenny, F. F., Espeland, E. K., Fiegener, R., Horning, M. E., Johnson, R. C., Kaye, T. N., Ott, J., St. Clair, J. B., & Leger, E. A. (2019). Strong patterns of intraspecific variation and local adaptation in Great Basin plants revealed through a review of 75 years of experiments. Ecology and Evolution, 9(11), 6259–6275. https://doi.org/10.1002/ece3.5200

Bischoff, M., Jürgens, A., & Campbell, D. R. (2014). Floral scent in natural hybrids of Ipomopsis (Polemoniaceae) and their parental species. Annals of Botany, 113(3), 533–544. https://doi.org/10.1093/aob/mct279

Borcard, D., Legendre, P., & Gillet, F. (2011). Numerical Ecological with R (1st. ed). Springer.

Byers, K. J. R. P., Bradshaw, H. D., & Riffell, J. A. (2014). Three floral volatiles contribute to differential pollinator attraction in monkeyflowers (Mimulus). Journal of Experimental Biology, 217(4), 614–623. https://doi.org/10.1242/jeb.092213

Chess, S. K. R., Raguso, R. A., & LeBuhn, G. (2008). Geographic divergence in floral morphology and scent in Linanthus dichotomus (Polemoniaceae). American Journal of Botany, 95(12), 1652–1659. https://doi.org/10.3732/ajb.0800118

Clifford, M. R. (2017). Scents and Sense Ability: The evolution and role of chemical cues and sensing in the pollination and herbivory of Passiflora. PhD Dissertation. University of Washington, Seattle, Washington, USA.

Couto, M. A. M. S., Soares, G. L. G., & Turchetto, C. (2024). Exploring floral scent in wild tobacco: comparison of volatile compounds across pollinator functional groups and Nicotiana sections. Evolutionary Ecology. https://doi.org/10.1007/s10682-024-10301-8

Cunningham, J. P., Moore, C. J., Zalucki, M. P., & West, S. A. (2004). Learning, odour preference and flower foraging in moths. Journal of Experimental Biology, 207(1), 87–94. https://doi.org/10.1242/jeb.00733

Delle-Vedove, R., Schatz, B., & Dufay, M. (2017). Understanding intraspecific variation of floral scent in light of evolutionary ecology. Annals of Botany, 120(1), 1–20. https://doi.org/10.1093/aob/mcx055

Dudareva, N., Pichersky, E., Tholl, D., Knudsen, J. T., & Raguso, R. A. (2020). Biology of plant volatiles. In Biology of Plant Volatiles (2nd ed.). CRC Press.

Eisen, K. E., Geber, M. A., & Raguso, R. A. (2022). Emission Rates of Species-Specific Volatiles Vary across Communities of Clarkia Species: Evidence for Multimodal Character Displacement . The American Naturalist, 199(6), 000–000. https://doi.org/10.1086/715501

Eisen, K. E., Powers, J. M., Raguso, R. A., & Campbell, D. R. (2022). An analytical pipeline to support robust research on the ecology, evolution, and function of floral volatiles. Frontiers in Ecology and Evolution, 10(October), 1–20. https://doi.org/10.3389/fevo.2022.1006416

El-Sayed, A. M. (2024). The Pherobase: database of pheromones and semiochemicals. Website: https://www.pherobase.com

Farré-Armengol, G., Filella, I., Llusià, J., Niinemets, Ü., & Peñuelas, J. (2014). Changes in floral bouquets from compound-specific responses to increasing temperatures. Global Change Biology, 20(12), 3660–3669. https://doi.org/10.1111/gcb.12628

Farré-Armengol, G., Filella, I., Llusià, J., Niinemets, Ü., & Peñuelas, J. (2015). Optimum temperature for floral terpene emissions tracks the mean temperature of the flowering season. Functional Plant Biology, 42(9), 851-857

Filella, I., Primante, C., Llusià, J., Martín González, A. M., Seco, R., Farré-Armengol, G., Rodrigo, A., Bosch, J., & Peñuelas, J. (2013). Floral advertisement scent in a changing plant-pollinators market. Scientific Reports, 3, 1–6. https://doi.org/10.1038/srep03434

Friberg, M., Schwind, C., Guimarães, P. R., Raguso, R. A., & Thompson, J. N. (2019). Extreme diversification of floral volatiles within and among species of *Lithophragma* (Saxifragaceae). Proceedings of the National Academy of Sciences of the United States of America, 116(10), 4406–4415. https://doi.org/10.1073/pnas.1809007116

Galen, C. (1983). The Effects of Nectar Thieving Ants on Seedset in Floral Scent Morphs of Polemonium Viscosum. Oikos, 41(2), 245–249.

Gervasi, D. D. L., & Schiestl, F. P. (2017). Real-time divergent evolution in plants driven by pollinators. Nature Communications, 8, 1–8. https://doi.org/10.1038/ncomms14691

Grant, V. (1949). Pollination Systems as Isolating Mechanisms in Angiosperms. Evolution, 3(1), 82–97.

Grant, V., & Grant, K. A. (1965). Flower Pollination in the Phlox Family. In Taxon. Columbia University Press. https://doi.org/10.2307/1217069

Grison-Pigé, L., Bessière, J. M., & Hossaert-McKey, M. (2002). Specific attraction of fig-pollinating wasps: Role of volatile compounds released by tropical figs. Journal of Chemical Ecology, 28(2), 283–295. https://doi.org/10.1023/A:1017930023741

Junker, R. R., & Blüthgen, N. (2010). Floral scents repel facultative flower visitors, but attract obligate ones. Annals of Botany, 105(5), 777–782. https://doi.org/10.1093/aob/mcq045

Legendre, P., Oksanen, J., Braak, C. J. F., & Centre-ville, S. (2011). Testing the significance of canonical axes in redundancy analysis. 269–277. https://doi.org/10.1111/j.2041-210X.2010.00078.x

Katzenberger, T. D., Lunau, K., & Junker, R. R. (2013). Salience of multimodal flower cues manipulates initial responses and facilitates learning performance of bumblebees. Behavioral Ecology and Sociobiology, 67(10), 1587–1599. https://doi.org/10.1007/s00265-013-1570-1

Kessler, D., & Baldwin, I. T. (2007). Making sense of nectar scents: The effects of nectar secondary metabolites on floral visitors of Nicotiana attenuata. Plant Journal, 49(5), 840–854. https://doi.org/10.1111/j.1365-313X.2006.02995.x

Kessler, D., Gase, K., & Baldwin, I. T. (2008). Field experiments with transformed plants reveal the sense of floral scents. Science, 321(5893), 1200–1202. https://doi.org/10.1126/science.1160072

Knudsen, Jette T., and Jonathan Gershenzon. (2020). "The chemical diversity of floral scent." In Biology of Plant Volatiles (2nd ed.). CRC Press. 57-78.

Knudsen, J. T., & Tollsten, L. (1993). Trends in floral scent chemistry in pollination syndromes: floral scent composition in moth- pollinated taxa. 263–284.

Majetic, C. J., Raguso, R. A., Tonsor, S. J., & Ashman, T. L. (2007). Flower color-flower scent associations in polymorphic Hesperis matronalis (Brassicaceae). Phytochemistry, 68(6), 865–874. https://doi.org/10.1016/j.phytochem.2006.12.009

Majetic, C. J., Raguso, R. A., & Ashman, T.-L. (2008). The Impact of Biochemistry vs. Population Membership on Floral Scent Profiles in Colour Polymorphic Hesperis matronalis. Annals of Botany, 102(6), 911–922. https://doi.org/10.1093/aob/mcn181

Majetic, C. J., Raguso, R. A., & Ashman, T. L. (2009). The sweet smell of success: Floral scent affects pollinator attraction and seed fitness in Hesperis matronalis. Functional Ecology, 23(3), 480–487. https://doi.org/10.1111/j.1365-2435.2008.01517.x

Ohashi, K., Jürgens, A., & Thomson, J. D. (2021). Trade-off mitigation: a conceptual framework for understanding floral adaptation in multispecies interactions. Biological Reviews, 96(5), 2258–2280. https://doi.org/10.1111/brv.12754

Okamoto, T., Y. Okuyama, R. Goto, M. Tokoro, and M. Kato. 2015. Parallel chemical switches underlying pollinator isolation in Asian Mitella. J. Evol. Biol. 28:590–600.

Peakall, R., & Whitehead, M. R. (2014). Floral odour chemistry defines species boundaries and underpins strong reproductive isolation in sexually deceptive orchids. Annals of Botany, 113(2), 341–355. https://doi.org/10.1093/aob/mct199

Pichersky, E., & Raguso, R. A. (2018). Why do plants produce so many terpenoid compounds? New Phytologist, 220(3), 692–702. https://doi.org/10.1111/nph.14178

Polis, Gary A., ed. The ecology of desert communities. University of Arizona Press, 1991.

Raguso, R. A., and O. Pellmyr (1998) Dynamic Headspace Analysis of Floral Volatiles: A Comparison of Methods. Oikos 81:238–254.

Raguso, R. A., Schlumpberger, B. O., Kaczorowski, R. L., & Holtsford, T. P. (2006). Phylogenetic fragrance patterns in *Nicotiana* sections Alatae and Suaveolentes. Phytochemistry, 67(17), 1931–1942. https://doi.org/10.1016/j.phytochem.2006.05.038

Raguso, R. A. (2008a). Start making scents: The challenge of integrating chemistry into pollination ecology. Entomologia Experimentalis et Applicata, 128(1), 196–207. https://doi.org/10.1111/j.1570-7458.2008.00683.x

Raguso, R. A. (2008b). Wake Up and Smell the Roses: The Ecology and Evolution of Floral Scent. Annual Review of Ecology, Evolution, and Systematics, 39(1), 549–569. https://doi.org/10.1146/annurev.ecolsys.38.091206.095601

Salguero-Gómez, R., Siewert, W., Casper, B. B., & Tielbörger, K. (2012). A demographic approach to study effects of climate change in desert plants. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1606): 3100–3114. https://doi.org/10.1098/rstb.2012.0074

Schatz, G. E. (1990). Some aspects of pollination biology in Central American forests. In Reproductive Ecology of Tropical Forest Plants, ed. KS Bawa, MHadley, pp. 69–84. Paris: UNESCO/Parthenon

Schiestl, F. P. (2015). Ecology and evolution of floral volatile-mediated information transfer in plants. New Phytologist, 206(2), 571–577. https://doi.org/10.1111/nph.13243

Stebbins, G. L. (1970). Adaptive Radiation of Reproductive Characteristics in Angiosperms, I: Pollination Mechanisms. Annual Review of Ecology and Systematics, 1(1), 307–326.

Theis, N., & Raguso, R. A. (2005). The effect of pollination on floral fragrance in thistles. Journal of Chemical Ecology, 31(11), 2581–2600. https://doi.org/10.1007/s10886-005-7615-9

Theis, N., Lerdau, M., & Raguso, R. A. (2007). The Challenge of Attracting Pollinators While Evading Floral Herbivores : Patterns of Fragrance Emission in Cirsium arvense and Cirsium repandum (Asteraceae) International Journal of Plant Sciences, 168(5), 587–601.

Tholl, D., W. Boland, A. Hansel, F. Loreto, U. S. R. Röse, and J. P. Schnitzler. 2006. Practical approaches to plant volatile analysis. Plant J. 45:540–560.

Toth, P., Undas, A. K., Verstappen, F., & Bouwmeester, H. (2016). Floral volatiles in parasitic plants of the Orobanchaceae. Ecological and taxonomic implications. Frontiers in Plant Science, 1–15. https://doi.org/10.3389/fpls.2016.00312

Wang, T. N., Clifford, M. R., Martínez-Gómez, J., Johnson, J. C., Riffell, J. A., & Di Stilio, V. S. (2019). Scent matters: Differential contribution of scent to insect response in flowers with insect vs. wind pollination traits. Annals of Botany, 123(2), 289–301. https://doi.org/10.1093/aob/mcy131

Waser, N. M., Chittka, L., Price, M. V, & Williams, N. M. (1996). Emphasizing new ideas to stimulate research in ecology. Generalization in pollination systems, and why it matters. Ecology, 77(4), 1043–1060.

Weber, M. G., Cacho, N. I., Phan, M. J. Q., Disbrow, C., Ramírez, S. R., & Strauss, S. Y. (2018). The evolution of floral signals in relation to range overlap in a clade of California Jewelflowers (Streptanthus s.l.). Evolution, 72(4), 798–807. https://doi.org/10.1111/evo.13456

Chapter 3

Reference genome of the color polymorphic desert annual plant sandblossoms, *Linanthus parryae*

Reprinted from:

Anghel, I. G., Jacobs, S. J., Escalona, M., Marimuthu, M. P. A., Fairbairn, C. W., Beraut, E., Nguyen, O., Toffelmier, E., Shaffer, H. B., & Zapata, F. (2022). Reference genome of the color polymorphic desert annual plant sandblossoms, *Linanthus parryae*. *Journal of Heredity*, *113*, 712–721.

Journal of Heredity, 2022, **113**, 712–721
https://doi.org/10.1093/jhered/esac052
Advance access publication 15 September 2022
Genome Resources

American Genetic Association

OXFORD

Genome Resources

Reference genome of the color polymorphic desert annual plant sandblossoms, *Linanthus parryae*

Ioana G. Anghel[1], Sarah J. Jacobs[2], Merly Escalona[3], Mohan P.A. Marimuthu[4], Colin W. Fairbairn[5], Eric Beraut[5], Oanh Nguyen[4], Erin Toffelmier[1,6], H. Bradley Shaffer[1,6], Felipe Zapata[1]

[1]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA, United States,
[2]Department of Botany, California Academy of Sciences, San Francisco, CA, United States,
[3]Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA, United States,
[4]DNA Technologies and Expression Analysis Core Laboratory, Genome Center, University of California-Davis, Davis, CA, United States,
[5]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA, United States,
[6]La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles, CA, United States

Address correspondence to I. Anghel at the address above, or e-mail: ianghel@ucla.edu.
Address correspondence to F. Zapata at the address above, or e-mail: fzapata@ucla.edu.

Corresponding Editor: Kenneth Olsen

Abstract

Sandblossoms, *Linanthus parryae* is a widespread annual plant species found in washes and sandy open habitats across the Mojave Desert and Eastern Sierra Nevada of California. Studies in this species have played a central role in evolutionary biology, serving as the first test cases of the shifting balance theory of evolution, models of isolation by distance, and metrics to describe the genetic structure of natural populations. Despite the importance of *L. parryae* in the development of landscape genetics and phylogeography, there are no genomic resources available for the species. Through the California Conservation Genomics Project, we assembled the first genome in the genus *Linanthus*. Using PacBio HiFi long reads and Hi-C chromatin conformation capture, we assembled 123 scaffolds spanning 1.51 Gb of the 1.96 Gb estimated genome, with a contig N50 of 18.7 Mb and a scaffold N50 of 124.8 Mb. This assembly, with a BUSCO completeness score of 88.7%, will allow us to re-visit foundational ideas central to our understanding of how evolutionary forces operate in a geographic landscape. In addition, it will be a new resource to uncover adaptations to arid environments in the fragile desert habitat threatened by urban and solar farm development, climate change, and off-road vehicles.

Keywords: California Conservation Genomics Project, CCGP, desert plant, Polemoniaceae, polymorphism

Introduction

The annual desert wildflower *Linanthus parryae* is an iconic species of the Mojave Desert. In years with higher rainfall, the species germinates prolifically and covers the desert floor with its well-known display of white and blue flowers. The species is endemic to California and occurs from 360 to 2100 m in elevation. Its geographic range extends across the western Mojave Desert, north through the Great Basin Desert of the Owens Valley, and in scattered populations across the Tehachapi Mountains and the southern inner CoastRange north to Shell Creek in San Luis Obispo County (see Fig. 1; Jepson Flora Project 2022). The species can be abundant in years of high rainfall, while in dry years the seeds remain dormant in the soil seedbank. *L. parryae* germinates in the winter and completes its life cycle by May.

One of the best-known attributes of this small desert plant is the color polymorphism that it displays both within populations and across its range. Populations are white flowered, blue-purple flowered, or polymorphic with both blue and white flowers often co-occurring in the same local habitat patch. Based on an unpublished herbarium survey, we found that approximately 40% of populations were polymorphic, and that these populations were distributed throughout the range of the species.

This extreme level of flower color polymorphism in *L. parryae* has attracted the attention of evolutionary biologists for generations and served as a model in debates about the degree to which natural selection or genetic drift contribute to the evolution of adaptively important characters in nature (Schemske and Bierzychudek 2007; Ishida 2017). In a highly influential paper, Epling and Dobzhansky mapped the relative frequencies of blue and white flowers across populations in the western Mojave Desert and concluded that populations with different color morph frequencies must represent subtly

Received August 2, 2022; Accepted September 12, 2022

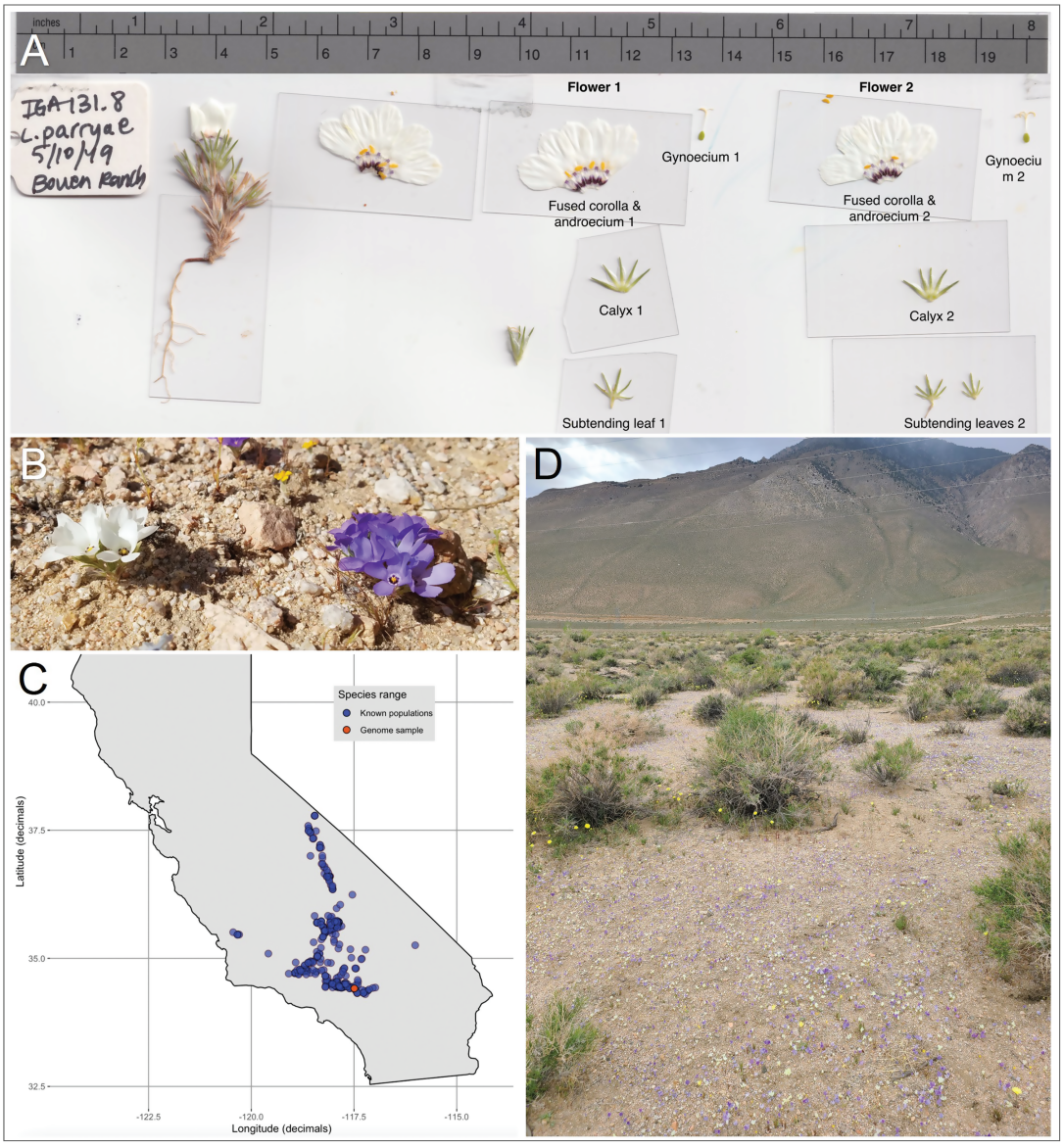


Fig. 1. *Linanthus parryae* is a minute annual wildflower, widespread in the Mojave Desert. (A) Dissected *L. parryae* flowers, with the gynoecium, fused corolla and androecium, calyx, and subtending leaves separated and labeled. (B) Blue and white corolla polymorphic populations of *L. parryae* can be found throughout the species' range. (C) A map of the species range including locations of herbarium specimens and observation reports aggregated by Calflora.org. The location of the specimen used for the genome assembly is marked in red. (D) A sandy wash west of Owens Lake, Inyo County, California, representative of *L. parryae* habitat.

differentiated "microgeographic races" with relatively restricted gene flow between them (Epling and Dobzhansky 1942). These data were then used to test Wright's shifting balance theory of evolution and assess the sometimes antagonistic roles of genetic drift and natural selection in maintaining intraspecific polymorphisms (Wright 1931, 1937). The *L. parryae* color polymorphism data were also used in the first empirical tests of models of isolation by distance (IBD) and were fundamental to the development of Wright's *F* statistics that describe the genetic structure of natural populations to infer migration frequencies, inbreeding coefficients, and other aspects of demographic history (Wright 1943). Later studies proposed that natural selection linked to variation in rain regimes is the main evolutionary force maintaining *L. parryae* flower color polymorphism (Schemske and Bierzychudek 2001, 2007). Given the importance of this species in these foundational studies, modern genomic approaches seem to be the inevitable next step in examining the genetic basis and

molecular mechanisms underlying this flower color polymorphism in this classic evolutionary genetic model system. These studies could also be expanded to include other flower color polymorphic species in the genus *Linanthus* including *L. dianthiflorus*, *L. killipii*, *L. bernardinus*, and *L. orcuttii*. Exploring the genetic basis for polymorphisms across closely related species and the roles of selection, drift, and gene flow operating on their respective landscapes will shed light on how polymorphisms arise and are maintained in nature.

Despite the central role of *L. parryae* in the development of phylogeography, landscape genetics, and population genetics, its habitat is in peril. Large swaths of the Mojave Desert are being developed for housing and photovoltaic power grids and utilized by Off-Highway Vehicles (OHVs). These flat sandy areas often coincide with the habitat preferred by desert annuals including *L. parryae*. Additionally, increases in temperature associated with global climate change can lead to range shifts and local extinctions for desert species that already occupy the upper extremes of plant thermal tolerance (Osmond *et al.* 1987; Lenoir and Svenning 2015). For such desert species to survive a changing climate, they must shift their range much faster than species in coastal or mountainous regions, which may be difficult to achieve in the shallow climatic gradients that characterize much of the desert (Loarie *et al.* 2009). At the same time, wetter microhabitats seem to be correlated with the ability of desert species to tolerate higher temperatures (Curtis *et al.* 2016). The large range of *L. parryae* throughout arid areas of California coupled with its adaptation to fluctuating patterns of rainfall make it an ideal species to study the ways that small desert annuals adapt to climate change, and to inform decision-makers on the extent to which protected areas will effectively support the survival of desert species. Understanding how desert annual species respond to global climatic changes and habitat destruction is therefore crucial to the conservation and protection of the desert biome that occupies 38% of California (Mooney and Zavaleta 2016).

This study reports the first chromosome-level genome assembly of *L. parryae*. To our knowledge, this is the second chromosome-level genome in the family Polemoniaceae (Jarvis *et al.* 2022), a group of significant historical importance in plant biosystematics and evolutionary biology (Grant and Grant 1965). This project was completed as part of the California Conservation Genomics Project (CCGP), an initiative with the goal of assembling genomic resources of endemic species in the state to inform conservation and management efforts (Fiedler *et al.* 2022; Shaffer *et al.* 2022). This reference genome will provide crucial resources to study the genetic mechanisms underlying and maintaining flower color polymorphism in the species. It will also serve as the foundation for studies identifying hotspots of genetic diversity and connectivity, assessing the genomic health of *L. parryae*, and investigating the genomic basis of adaptation to extreme environments.

Methods

Biological materials

We collected *L. parryae* individuals from a population with only white-flowered morphs. The 2 individuals sequenced in this project were collected on 17 May 2020, in the town of Phelan in the Mojave Desert, San Bernardino County, California. The location of the population was roadside, north of Muscatel Street, west of the intersection with Windemere Road, near the powerline, approximately 1 mile south of Phelan Road (34.412122° N, −117.493026° W). We collected entire plants, including leaves, flowers, stems, and roots. The plants were in the budding and flowering developmental stage. Immediately after collection, we placed each individual in separate 15 mL Nalgene bottles and kept the bottles in a liquid nitrogen dewar while in the field. The 2 individual plants were then stored in a −80 °C freezer. Individual plant IGA184.5 was shipped to the University of California Davis for high molecular weight (HMW) DNA extractions and Pacific BioSciences HiFi library preparation and sequencing (Pacific BioSciences—PacBio, Menlo Park, CA). Individual plant IGA184.4 was sent to University of California Santa Cruz for Omni-C library preparation and sequencing. A voucher for this population was previously collected (accession number UCR-112367) and it is stored at the herbarium of University of California, Riverside (UCR).

Nucleic acid library preparation and sequencing

Omni-C library preparation and sequencing

The Omni-C library was prepared using the Dovetail Omni-C Kit (Dovetail Genomics, Scotts Valley, CA) according to the manufacturer's protocol with slight modifications. First, specimen tissue from the whole plant (UCR112367; IGA184.4) was thoroughly ground with a mortar and pestle while cooled with liquid nitrogen. Nuclear isolation was then performed using published methods (Inglis *et al.* 2018). Subsequently, chromatin was fixed in place in the nucleus and digested under various conditions of DNase I until a suitable fragment length distribution of DNA molecules was obtained. Chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter-containing ends. After proximity ligation, crosslinks were reversed, and the DNA purified from proteins. Purified DNA was treated to remove biotin that was not internal to ligated fragments. An NGS library was generated using an NEB Ultra II DNA Library Prep kit (New England Biolabs, Ipswich, MA) with an Illumina-compatible y-adaptor, biotin-containing fragments were captured using streptavidin beads, and the postcapture product was split into 2 replicates prior to PCR enrichment to preserve library complexity, with each replicate receiving unique dual indices. The library was sequenced at Vincent J. Coates Genomics Sequencing Lab (Berkeley, CA) on an Illumina NovaSeq (Illumina, CA) platform to generate approximately 100 million 2 × 150 bp read pairs per Gb genome size.

DNA extraction

We extracted HMW genomic DNA (gDNA) from whole plant tissue of a single individual (600 mg; IGA184.5) using the method described in Inglis *et al.* (2018), with the following modifications. We used sodium metabisulfite (1%, w/v) instead of 2-mercaptoethanol (1%, v/v) in the sorbitol wash buffer and the lysis buffer. Using mortar and pestle, we pulverized the frozen tissue in liquid nitrogen for 15 min, then gently resuspended it in 10 mL of sorbitol wash buffer. The suspension was centrifuged at 3000 × *g* for 5 min at room temperature, and the supernatant was discarded. Using a paintbrush, we gently resuspended the ground tissue pellet in 10 mL of sorbitol wash buffer and repeated the wash

step 5 times to remove potential contaminants that may coprecipitate with DNA. We performed the lysis step at 45 °C (instead of the standard 65 °C, to avoid potential heat-induced DNA damage) for 1 h with gentle inversion every 15 min. The DNA purity was estimated using absorbance ratios (260/280 = 1.80 and 260/230 = 2.24) on a NanoDrop ND-1000 spectrophotometer. The final DNA yield (84 ng/μL; 32 μg) was quantified using a Quantus Fluorometer (QuantiFluor ONE dsDNA Dye assay; Promega, Madison, WI). The size distribution of the HMW DNA was estimated using the Femto Pulse system (Agilent, Santa Clara, CA), where 52% of the DNA fragments were found to be 45 kb or longer.

HiFi library preparation and sequencing

The HiFi SMRTbell library was constructed using the SMRTbell Express Template Prep Kit v2.0 (PacBio, Cat. #100-938-900) according to the manufacturer's instructions. HMW gDNA was sheared to a target DNA size distribution between 15 and 20 kb. The sheared gDNA was concentrated using 0.45 of AMPure PB beads (PacBio, Cat. #100-265-900) for the removal of single-strand overhangs at 37 °C for 15 min, followed by further enzymatic steps of DNA damage repair at 37 °C for 30 min, end repair and A-tailing at 20 °C for 10 min and 65 °C for 30 min, ligation of overhang adapter v3 at 20 °C for 60 min and 65 °C for 10 min to inactivate the ligase, then nuclease treated at 37 °C for 1 h. The SMRTbell library was purified and concentrated with 0.45 Ampure PB beads (PacBio, Cat. #100-265-900) for size selection using the BluePippin/PippinHT system (Sage Science, Beverly, MA; Cat #BLF7510/HPE7510) to collect fragments greater than 7 to 9 kb. The 15 to 20 kb average HiFi SMRTbell library was sequenced at UC Davis DNA Technologies Core (Davis, CA) using 3 8M SMRT cells, Sequel II sequencing chemistry 2.0, and 30-h movies each on a PacBio Sequel II sequencer.

Nuclear genome assembly

We assembled the *L. parryae* genome following the CCGP assembly protocol Version 4.0, as outlined in Table 1 which lists the nondefault parameters used in the assembly. As with other CCGP assemblies, our goal was to produce a high-quality and highly contiguous assembly using PacBio HiFi reads and Omni-C data while minimizing manual curation. We removed remnant adapter sequences from the PacBio HiFi dataset using HiFiAdapterFilt (Sim *et al.* 2022) and obtained the dual or partially phased initial diploid assembly (http://lh3.github.io/2021/10/10/introducing-dual-assembly) using HiFiasm (Cheng *et al.* 2021). We tagged output haplotype 1 as the primary assembly, and output haplotype 2 as the alternate assembly. We aligned the Omni-C data to the assemblies by using the Arima Genomics Mapping Pipeline (https://github.com/ArimaGenomics/mapping_pipeline) and then scaffolded both assemblies with SALSA (Ghurye *et al.* 2017, 2018). Next, we identified sequences corresponding to haplotypic duplications, contig overlaps and repeats on the primary assembly with purge_dups (Guan *et al.* 2020) and transferred them to the alternate assembly.

We generated Omni-C contact maps for both assemblies by aligning the Omni-C data against the corresponding assembly with BWA-MEM (Li 2013), identified ligation junctions, and generated Omni-C pairs using pairtools (Goloborodko *et al.* 2018). We generated a multiresolution Omni-C matrix with cooler (Abdennur and Mirny 2020) and balanced it with hicExplorer (Ramírez *et al.* 2018). We used HiGlass (Kerpedjiev *et al.* 2018) and the PretextSuite (https://github.com/wtsi-hpag/PretextView; https://github.com/wtsi-hpag/PretextMap; https://github.com/wtsi-hpag/PretextSnapshot) to visualize the contact maps, then checked the contact maps for major misassemblies. In detail, if in the proximity of a join that was made by the scaffolder, we identified a strong signal off-diagonal and lack of signal in the consecutive genomic region, we marked this join. All the joins that were marked, were "dissolved", meaning that we broke the scaffolds at the coordinates of these joins. After this process, no further joins were made. Using the PacBio HiFi reads and YAGCloser (https://github.com/merlyescalona/yagcloser), we closed some of the remaining gaps generated during scaffolding. We then checked for contamination using the BlobToolKit Framework (Challis *et al.* 2020). Finally, we trimmed remnants of sequence adaptors and mitochondrial contamination identified during the contamination screening performed by NCBI.

Genome size estimation and quality assessment

We generated k-mer counts from the PacBio HiFi reads using meryl (https://github.com/marbl/meryl). The k-mer database was then used in GenomeScope 2.0 (Ranallo-Benavidez *et al.* 2020) to estimate genome features including genome size, heterozygosity, and repeat content. To obtain general contiguity metrics, we ran QUAST (Gurevich *et al.* 2013). To evaluate genome quality and completeness we used BUSCO (Manni *et al.* 2021) with the embryophyta ortholog database (embryophyta_odb10) which contains 1,614 genes. Assessment of base level accuracy (QV) and k-mer completeness was performed using the previously generated meryl database and merqury (Rhie *et al.* 2020). We further estimated genome assembly accuracy via BUSCO gene set frameshift analysis using the pipeline described in Korlach *et al.* (2017). We identified and annotated repeat sequences using RepeatModeler and RepeatMasker (Smit and Hubley 2008–2015; Smit *et al.* 2013–2015).

Measurements of the size of the phased blocks are based on the size of the contigs generated by HiFiasm in HiC mode. We followed the quality metric nomenclature established by Rhie *et al.* (2021), with the genome quality code $x{\cdot}y{\cdot}P{\cdot}Q{\cdot}C$, where x = log10[contig NG50]; y = log10[scaffold NG50]; P = log10[phased block NG50]; Q = Phred base accuracy QV (quality value); C = % genome represented by the first "n" scaffolds, following a known karyotype $2n$ = 18 (Patterson 1979). Quality metrics for the notation were calculated on the primary assembly.

Results

The Omni-C and PacBio HiFi sequencing libraries generated 295.4 million read pairs and 4.8 million reads, respectively. The latter yielded 39.39-fold coverage (N50 read length 16,891 bp; minimum read length 48 bp; mean read length 16,034 bp; maximum read length 58,036 bp). Calculation of coverage is based on a flow cytometry estimated genome size of 1.96 Gb. The GenomeScope 2.0 genome size estimation was 1.6 Gb. Based on PacBio HiFi reads, we estimated a 0.182% sequencing error rate and 4.65% nucleotide heterozygosity rate. The k-mer spectrum based on PacBio HiFi reads (Fig. 2A) shows a bimodal distribution with 2 major peaks at ~23- and ~45-fold coverage, where peaks correspond to

Table 1. Assembly pipeline and software used.

Assembly	Software	Version
Filtering PacBio HiFi adapters	HiFiAdapterFilt	Commit 64d1c7b
K-mer counting	Meryl (k = 21)	1
Estimation of genome size and heterozygosity	GenomeScope	2
De novo assembly (contigging)	HiFiasm (Hi-C mode, –primary, output p_ctg.hap1, p_ctg.hap2)	0.16.1-r375
Remove low-coverage, duplicated contigs	purge_dups	1.2.6
Scaffolding		
Omni-C data alignment for SALSA	Arima Genomics Mapping Pipeline (https://github.com/ArimaGenomics/mapping_pipeline)	Commit 2e74ea4
Omni-C scaffolding	SALSA (-DNASE, -i 20, -p yes)	2
Gap closing	YAGCloser (-mins 2 -f 20 -mcc 2 -prt 0.25 -eft 0.2 -pld 0.2)	Commit 20e2769
OmniC contact map generation		
Short-read alignment	BWA-MEM (-5SP)	0.7.17-r1188
SAM/BAM processing	Samtools	1.11
SAM/BAM filtering	pairtools	0.3.0
Pairs indexing	pairix	0.3.7
Matrix generation	Cooler	0.8.10
Matrix balancing	HiCExplorer (hicCorrectmatrix correct --filterThreshold -2 4)	3.6
Contact map visualization	HiGlass	2.1.11
	PretextMap	0.1.4
	PretextView	0.1.5
	PretextSnapshot	0.03
Benchmarking		
Basic assembly stats	QUAST (--est-ref-size)	5.0.2
Assembly completeness	BUSCO (-m geno, -l embryophyta)	5.0.0
	Merqury	1
Contamination screening		
General contamination screening	BlobToolKit	2.3.3
Local alignment tool	BLAST+	2.10
Repeat element identification		
Repeat identification	RepeatModeler	2.0.3
Repeat annotation	RepeatMasker	4-1-2

Software citations are listed in the text.

homozygous and heterozygous states of a diploid species, respectively. The distribution presented in this k-mer spectrum supports that of a high heterozygosity profile.

The final assembly (ddLinParr1) consists of 2 pseudo haplotypes, primary and alternate, both genome sizes similar to the estimated value from GenomeScope 2.0 (Fig. 2A). The primary assembly consists of 123 scaffolds spanning 1.5 Gb with a contig N50 of 18.7 Mb, scaffold N50 of 124.8 Mb, largest contig of 69.4 Mb and largest scaffold of 194.5 Mb. The alternate assembly consists of 581 scaffolds, spanning 1.8 Gb with contig N50 of 10.6 Mb, scaffold N50 of 44.0 Mb, largest contig 41.8 Mb and largest scaffold of 145.2 Mb. Detailed assembly statistics are reported in Table 2, and a graphical representation for the primary assembly in Fig. 2B (see Supplementary Fig. 1 for the alternate assembly). The Omni-C contact map suggests that the primary assembly is highly contiguous (Fig. 2C).

We identified a total of 12 misassemblies, 6 per assembly, and broke the corresponding joins made by SALSA2 on both assemblies. We were able to close a total of 22 gaps, 9 on the primary and 13 on the alternate assembly. Finally, we filtered 15 contigs from the primary assembly, 1 matching to Oomycota and the rest to Ascomycota contaminants. No further contigs were removed. The primary assembly has a BUSCO completeness score of 88.7% using the embryophyta gene set, a per base quality (QV) of 68.93, a k-mer completeness of 58.96 and a frameshift indel QV of 46.04. The alternate assembly has a BUSCO completeness score of 84.4% using the embryophyta gene set, a per base quality (QV) of 66.36, a k-mer completeness of 60.95 and a frameshift indel QV of 47.1. We have deposited scaffolds corresponding to both primary and alternate assemblies on NCBI (see Table 2 and Data availability for details).

Discussion

Prior to sequencing, we estimated the genome size of *L. parryae* using flow cytometry. Our 1C estimates were between 1.93 and 1.99 Gb. This difference in size compared with the GenomeScope estimation of 1.6 Gb may be due to repetitive elements in the genome. Based on flow cytometry, the species seems to have an average genome size compared with other

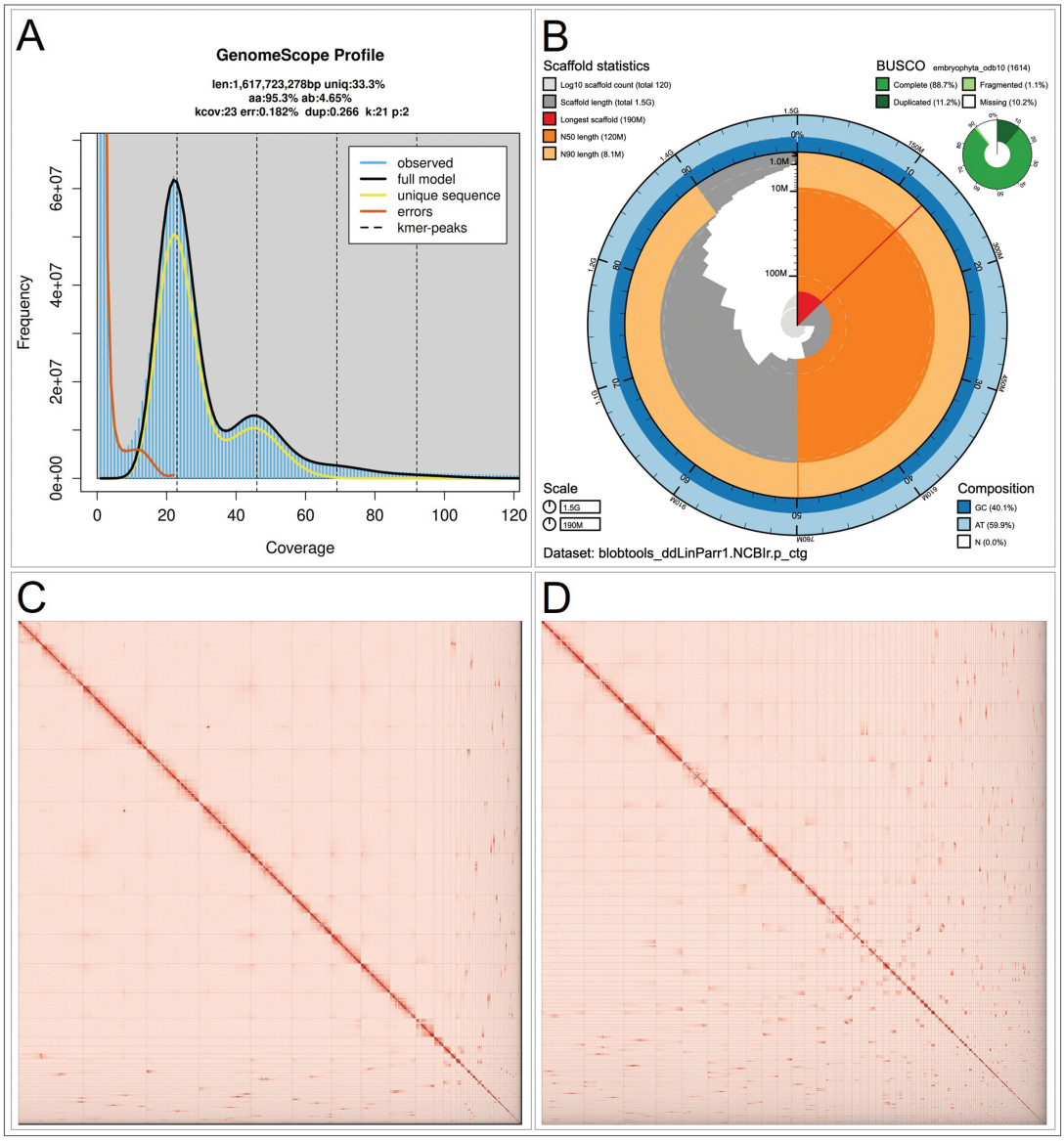


Fig. 2. Visual overview of the *Linanthus parryae* genome assembly metrics. (A) K-mer spectra output generated from PacBio HiFi data without adapters using GenomeScope 2.0. The bimodal pattern observed corresponds to a diploid genome and the k-mer profile matches that of high heterozygosity. K-mers at lower coverage and high frequency correspond to differences between haplotypes, whereas the higher coverage and low frequency k-mers correspond to the similarities between haplotypes. (B) BlobToolKit snail plot showing a graphical representation of the quality metrics presented in Table 2 for the *Linanthus parryae* primary assembly (ddLinParr1.0.p). The plot circle represents the full size of the assembly. From the inside-out, the central plot covers length-related metrics. The red line represents the size of the longest scaffold; all other scaffolds are arranged in size-order moving clockwise around the plot and drawn in gray starting from the outside of the central plot. Dark and light orange arcs show the scaffold N50 and scaffold N90 values. The central light gray spiral shows the cumulative scaffold count with a white line at each order of magnitude. White regions in this area reflect the proportion of Ns in the assembly; the dark versus light blue area around it shows mean, maximum and minimum GC versus AT content at 0.1% intervals (Challis ***et al.*** 2020). Hi-C contact maps for the primary (C) and alternate (D) genome assembly generated with PretextSnapshot. Hi-C contact maps translate proximity of genomic regions in 3D space to contiguous linear organization. Each cell in the contact map corresponds to sequencing data supporting the linkage (or join) between 2 of such regions. Scaffolds are separated by black lines and higher density of the lines may correspond to higher levels of fragmentation.

Table 2. Sequencing and assembly statistics, and accession numbers.

BioProjects and vouchers	CCGP NCBI BioProject		PRJNA720569			
	Genera NCBI BioProject		PRJNA765619			
	Species NCBI BioProject		PRJNA777190			
	NCBI BioSample		SAMN26264369			
	Specimen identification		IGA184			
	NCBI Genome accessions		Primary		Alternate	
	Assembly accession		JALGPY000000000		JALGPZ000000000	
	Genome sequences		GCA_023055425.1		GCA_023055565.1	
Genome sequence	PacBio HiFi reads	Run	1 PACBIO_SMRT (Sequel II) run: 4.8M spots, 77.2G bases, 52.5 Gb			
		Accession	SRX15304035			
	Omni-C Illumina reads	Run	1 ILLUMINA (Illumina NovaSeq 6000) run: 295.4M spots, 65.3G bases, 21.4G			
		Accession	SRX15304036, SRX15304037			
Genome Assembly Quality Metrics	Assembly identifier (quality code[a])		ddLinParr1 (7.8.P7.Q67.C78)			
	HiFi read coverage[b]		39.39×			
			Primary		Alternate	
	Number of contigs		208		730	
	Contig N50 (bp)		18,687,320		10,572,045	
	Contig NG50		17,395,258		11,393,586	
	Longest contigs		69,385,222		41,847,382	
	Number of scaffolds		123		581	
	Scaffold N50		124,808,130		43,960,851	
	Scaffold NG50		124,808,130		47,495,494	
	Largest scaffold		194,533,388		145,210,036	
	Size of final assembly (bp)		1,514,484,308		1,755,711,046	
	Phase block NG50		17,395,258		11,393,586	
	Gaps per Gbp (#Gaps)		56 (85)		85 (149)	
	Indel QV (frameshift)		46.04081985		47.09876692	
	Base pair QV		68.9289		66.3567	
					Full assembly = 67.3638	
	K-mer completeness		58.9657		60.9501	
					Full assembly = 98.0448	
	BUSCO completeness (embryophyta), *n* = 1,614		C	S	D	F	M
		P[c]	88.70%	77.50%	11.20%	1.10%	10.20%
		A[c]	84.40%	74.50%	9.90%	1.10%	14.50%

[a]Assembly quality code *x*·*y*·*P*·*Q*·C, where *x* = log10[contig NG50]; *y* = log10[scaffold NG50]; *P* = log10[phased block NG50]; *Q* = Phred base accuracy QV (quality value); C = % genome represented by the first "*n*" scaffolds, following a known karyotype 2*n* = 18 (Patterson 1979). BUSCO scores. (C)omplete and (S)ingle; (C)omplete and (D)uplicated; (F)ragmented and (M)issing BUSCO genes. *n*, number of BUSCO genes in the set/database. Bp: base pairs.
[b]Read coverage and NGx statistics have been calculated based on a genome size of 1.96 Gb.
[c]P(rimary) and (A)lternate assembly values.

species in the genus. Flow cytometry genome size estimations of 9 other species of *Linanthus* ranged from 1.31 to 3.51 Gb, with an average of 2.02 Gb.

This assembly is the second published genome in the family Polemoniaceae, and the first for *Linanthus*. The other chromosome-level genome assembly in the family is *Gilia yorkii*. *Gilia* and *Linanthus* last shared a common ancestor about 60 MYA (Landis *et al.* 2018). While these species have not shared an evolutionary history for a long period of time, we provide these metrics as a context for comparative genomics in Polemoniaceae. *G. yorkii* was sequenced using PacBio at 67× coverage compared with the *L. parryae* at 39.4× coverage. The genome size of *G. yorkii* was estimated at 2.80 Gb and the *L. parryae* at 1.96 Gb. The BUSCO completeness score for *G. yorkii* was 96.8% compared with the 88.7% in *L. parryae*. The *L. parryae* assembly has a longer contig N50 than the *G. yorkii* genome (18.7 vs. 2.5 Mb) and a shorter scaffold N50 (76.8 vs. 285.8 Mb). The *G. yorkii* reads were assembled in 3,947 contigs and 2,043 scaffolds, compared with 208 contigs and 123 scaffolds for *L. parryae* (Jarvis *et al.* 2022).

A total of 71.99% of the *L. parryae* genome was annotated as repetitive. This is similar to the values reported for *G. yorkii* (75.60%), but higher than the average of 45.49% across plant species (Luo *et al.* 2022). Long-terminal repeat (LTR) retroelements made up 38.26% of the genome, less

than for *G. yorkii* at 45.81% and more than the plant species average at 21.66%. Of the LTRs, Ty1/Copia made up 24.3% of the genome, which is similar to *G. yorkii* at 29.43%. Over a quarter of the genome (26.62%) was annotated as unclassified repeat elements, which is again similar to *G. yorkii* (21.92%); both species are about double the plant average (13.2%). This high percentage of unknown repeats may reflect the fact that plants in general, and Polemoniaceae in particular, are not well represented in the repeat database (Luo *et al.* 2022). The RepeatMasker annotation of repeat elements for *L. parryae* is summarized in Table 3.

This reference genome will contribute to conservation in California in several important ways. *L. parryae* occurs in a region of California considered a hotbed of plant diversity and neoendemism (Kraft *et al.* 2010). In concert with other species in the CCGP, this genomic resource will help determine whether these geographic areas also correspond to hotspots of genomic diversity and provide important information for setting realistic priorities for conservation. Although a desert species, *L. parryae* occurs across a relatively wide and heterogeneous environment gradient with an elevational range of 360 to 2100 m, a precipitation range of 50 to 500 mm/yr, and a geographic range that spans 470 km. The morphological and environmental variation along this species distribution may correspond with genetic differentiation between populations and potential local adaptation. Hence, this genome can serve as a foundation to investigate patterns of genomic diversity, connectivity, and health for other California plants, as well as shed light on the sorting of adaptive genomic variation across ecological gradients that are often associated with species divergence.

L. parryae occurs in fragile habitats that are particularly prone to habitat destruction, with at least 5 major ongoing threats. First, large swaths of the desert are now utilized by OHVs, including 200,000 acres of newly designated motorized recreation land established by the California Desert Protection Act of 2019 (Feinstein 2020). Second, photovoltaic power grid development is increasing with projects proposed or being built on an additional 30,000 acres of California desert (Wilson 2020). These developments can, and will, drastically impact unprotected desert lands (Hernandez *et al.* 2015). Third, urban sprawl spilling over from the Los Angeles Basin has increased habitat destruction in the southwestern edge of the Mojave Desert (Stewart 1997), an area with the highest density of *L. parryae*. Fourth, as water resources become more scarce and seasonal streams are diverted, ephemeral wash habitat for *Linanthus* species, as well as other desert dwellers, may become uninhabitable (Levick *et al.* 2008). Lastly, ecological pressure from ongoing invasive plants will be exacerbated with climate change (Smith *et al.* 2000). This reference genome will help determine the effects of these threats and enhance conservation plans in currently protected desert areas, helping resource managers determine which natural areas to prioritize for future protection.

Table 3. Classification of repeat elements generated from RepeatMasker.

	Number of elements[a]	Length occupied (bp)	Percentage of sequence (%)
Retroelements	649,089	1,119,332,631	42.05
SINEs	0	0	0.00
Penelope	0	0	0.00
LINEs	122,415	101,011,181	3.80
CRE/SLACS	0	0	0.00
L2/CR1/Rex	0	0	0.00
R1/LOA/Jockey	0	0	0.00
R2/R4/NeSL	0	0	0.00
RTE/Bov-B	7,307	1,975,180	0.07
L1/CIN4	115,108	99,036,001	3.72
LTR elements	526,674	1,018,321,450	38.26
BEL/Pao	0	0	0.00
Ty1/Copia	330,448	646,651,635	24.30
Gypsy/DIRS1	196,097	371,644,595	13.96
Retroviral	129	25,220	0.00
DNA transposons	62,196	51,067,856	1.92
hobo-Activator	15,958	5,783,269	0.22
Tc1-IS630-Pogo	3,961	1,515,451	0.06
En-Spm	0	0	0.00
MuDR-IS905	0	0	0.00
PiggyBac	0	0	0.00
Tourist/Harbinger	4,485	1,597,874	0.06
Other (Mirage, P-element, Transib)	0	0	0.00
Rolling-circles	4,483	5,689,821	0.21
Unclassified	1,755,132	708,389,341	26.62
Total interspersed repeats		1,878,789,828	70.59
Small RNA	22,615	11,470,955	0.43
Satellites	12,336	1,179,847	0.04
Simple repeats	357,637	16,506,411	0.62
Low complexity	48,902	2,518,203	0.09

[a]Most repeats fragmented by insertions or deletions have been counted as one element.

Funding

This work was supported by the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224].

Conflict of Interest

The authors declare that by publishing this manuscript they have no conflicts of interest.

Acknowledgments

PacBio Sequel II library prep and sequencing were carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by National Institutes of Health (NIH) Shared Instrumentation Grant 1S10OD010786-01. Deep sequencing of Omni-C libraries used the Novaseq S4 sequencing platforms at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

We thank the staff at the UC Davis DNA Technologies and Expression Analysis Cores and the UC Santa Cruz Paleogenomics Laboratory for their diligence and dedication to generating high-quality sequence data. We also thank the members of the Zapata lab for providing feedback on the manuscript.

Data availability

Data generated for this study are available under NCBI BioProject PRJNA777190. Raw sequencing data for sample IGA184 (NCBI BioSample SAMN26264369) are deposited in the NCBI Short Read Archive (SRA) under SRX15304035 for PacBio HiFi sequencing data, and SRX15304036, SRX15304037 for the Omni-C Illumina sequencing data. GenBank accessions for both primary and alternate assemblies are GCA_023055425.1 and GCA_023055565.1; and for genome sequences JALGPY000000000 and JALGPZ000000000. Assembly scripts and other data for the analyses presented can be found at the following GitHub repository: www.github.com/ccgproject/ccgp_assembly.

References

Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*. 2020;36(1):311–316.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit—interactive quality assessment of genome assemblies. *G3*. 2020;10(4):1361–1374.

Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmell NJ, Li H. Robust haplotype-resolved assembly of diploid individuals without parental data, arXiv [q-bio.GN], arXiv:2109.04785, 2021, preprint: not peer reviewed. doi:10.48550/arXiv.2109.04785

Curtis EM, Gollan J, Murray BR, Leigh A. Native microhabitats better predict tolerance to warming than latitudinal macro-climatic variables in arid-zone plants. *J Biogeogr*. 2016;43(6):1156–1165.

Epling C, Dobzhansky T. Genetics of natural populations. VI. Microgeographic races of *Linanthus parryae*. *Genetics*. 1942;27(3):317–332.

Feinstein D. United States Senator for California. 2020 [accessed 2020 Mar 9]. https://www.feinstein.senate.gov/public/index.cfm/2019/2/senate-passes-feinstein-bill-completing-25-year-effort-to-protect-california-desert

Fiedler PL, Erickson B, Esgro M, Gold M, Hull JM, Norris J, Shapiro B, Westphal MF, Toffelmier E, Shaffer HB. Seizing the moment: the opportunity and relevance of the California Conservation Genomics Project to state and federal conservation policy. *J Hered*. 2022; esac046. doi:10.1093/jhered/esac046

Ghurye J, Pop M, Koren S, Bickhart D, Chin C-S. Scaffolding of long read assemblies using long range contact information. *BMC Genomics*. 2017;18(1):1–11.

Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Koren S. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput Biol*. 2018;15(8):e1007273.

Goloborodko A, Abdennur N, Venev S, Hbbrandao, gfudenberg. *mirnylab/pairtools: v0.2.0*. 2018. https://zenodo.org/record/1490831

Grant V, Grant KA. *Flower pollination in the Phlox family*. New York (NY): Columbia University Press; 1965.

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*. 2020;36(9):2896–2898.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072–1075.

Hernandez RR, Hoffacker MK, Murphy-Mariscal ML, Wu GC, Allen MF. Solar energy development impacts on land cover change and protected areas. *Proc Natl Acad Sci USA*. 2015;112(44):13579–13584.

Inglis PW, Pappas MDCR, Resende LV, Grattapaglia D. Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One*. 2018;13(10):e0206085.

Ishida Y. Sewall Wright, shifting balance theory, and the hardening of the modern synthesis. *Stud Hist Philos Biol Biomed Sci*. 2017;61:1–10.

Jarvis DE, Maughan PJ, DeTemple J, Mosquera V, Li Z, Barker MS, Johnson LA, Whipple CJ. Chromosome-scale genome assembly of *Gilia yorkii* enables genetic mapping of floral traits in an interspecies cross. *Genome Biol Evol*. 2022;14(3):1–13, evac017.

Jepson Flora Project, editor. Jepson eFlora. 2022. Accessed on June 7, 2022. https://ucjeps.berkeley.edu/eflora/

Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobelt H, Luber JM, Ouellette SB, Azhir A, Kumar N, et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol*. 2018;19(1):1–12.

Korlach J, Gedman G, Kingan SB, Chin C-S, Howard JT, Audet J-N, Jarvis ED. De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience*. 2017;6(10):1–16.

Kraft NJ, Baldwin BG, Ackerly DD. Range size, taxon age and hotspots of neoendemism in the California flora. *Divers Distrib*. 2010;16(3):403–413.

Landis JB, Bell CD, Hernandez M, Zenil-Ferguson R, McCarthy EW, Soltis DE, Soltis PS. Evolution of floral traits and impact of reproductive mode on diversification in the phlox family (Polemoniaceae). *Mol Phylogenet Evol*. 2018;127:878–890.

Lenoir J, Svenning JC. Climate-related range shifts—a global multidimensional synthesis and new research directions. *Ecography*. 2015;38(1):15–28.

Levick LR, Goodrich DC, Hernandez M, Fonseca J, Semmens DJ, Stromberg JC, Tluczek M, Leidy RA, Scianni M, Guertin DP, et al. *The ecological and hydrological significance of ephemeral and intermittent streams in the arid and semi-arid American Southwest*. Washington, DC: US Environmental Protection Agency, Office of Research and Development; 2008.

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, arXiv, arXiv:1303.3997, 2013, preprint: not peer reviewed. doi:10.48550/arXiv.1303.3997

Loarie SR, Duffy PB, Hamilton H, Asner GP, Field CB, Ackerly DD. The velocity of climate change. *Nature*. 2009;462(7276):1052–1055.

Luo X, Chen S, Zhang Y. PlantRep: a database of plant repetitive elements. *Plant Cell Rep*. 2022;41(4):1163–1166.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38(10):4647–4654.

Mooney H, Zavaleta E. *Ecosystems of California*. Oakland, California: University of California Press; 2016.

Osmond CB, Austin MP, Berry JA, Billings WD, Boyer JS, Dacey JWH, Winner WE. Stress physiology and the distribution of plants. *BioScience*.1987;37(1):38–48.

Patterson R. Chromosome numbers in annual *Linanthus* species. *Madrono*. 1979:96–100.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11(1):1–10.

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke, T. 2018. High-resolution TADs reveal DNA sequences underlying genome organization in fies. *Nat Commun*. 9:1–5.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 2021;592:737–746.

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020;21(1):1–27.

Schemske DW, Bierzychudek P. Perspective: evolution of flower color in the desert annual *Linanthus parryae*: Wright revisited. *Evolution*. 2001;55(7):1269–1282.

Schemske DW, Bierzychudek P. Spatial differentiation for flower color in the desert annual *Linanthus parryae*: was Wright right? *Evolution*. 2007;61(11):2528–2543.

Shaffer HB, Toffelmier E, Corbett-Detig RB, Escalona M, Erickson B, Fiedler P, Gold M, Harrigan RJ, Hodges S, Luckau TK, et al. Landscape genomics to enable conservation actions: the California Conservation Genomics Project. *J Hered*. 2022;113:578–588.

Sim SB, Corpuz RL, Simmonds TJ, Geib SM. HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics*. 2022;23:1–7.

Smit AFA, Hubley R. RepeatModeler Open-1.0. 2008–2015. Accessed July 19, 2022. http://www.repeatmasker.org

Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015. Accessed July 19, 2022. http://www.repeatmasker.org

Smith SD, Huxman TE, Zitzer SF, Charlet TN, Housman DC, Coleman JS, Fenstermaker LK, Seemann JR, Nowak RS. Elevated CO_2 increases productivity and invasive species success in an arid ecosystem. *Nature*. 2000;408(6808):79–82.

Stewart WC. *Bioregional demographic trends and implications for biodiversity*. Sacramento: Fire and Resource Assessment Program; 1997.

Wilson J. Solar surges in the California desert. So why are environmentalists upset? *Palm Springs Desert Sun*. Jan 3, 2020. 2020 [accessed 2020 Mar 9]. https://www.desertsun.com/story/news/environment/2020/01/03/solar-surges-california-desert-environment-trump/2665799001/

Wright S. Evolution in Mendelian populations. *Genetics*. 1931;16(2):97–159.

Wright S. The distribution of gene frequencies in populations. *Proc Natl Acad Sci USA*. 1937;23(6):307–320.

Wright S. Isolation by distance. *Genetics*. 1943;28(2):114–138.

Chapter 4

Landscape genomics of *Linanthus parryae* reveals both isolation by distance and by environment and high genetic structure across its range

Abstract

Linanthus parryae played an important role in the development of population genetics theory when Sewall Wright used the distribution of blue and white-flowered individuals to illustrate the principle of isolation by distance, proposing genetic drift as the main microevolutionary force driving differentiation in natural populations. Subsequent biologists challenged these findings and proposed that natural selection played a more significant role in shaping patterns of population variation. These early studies relied on phenotype frequencies as a proxy for genetic divergence, but the availability of high-quality genomic data allows us to revisit this controversy. In this study, we used whole genome sequencing of individuals across the geographic range of *L. parryae* for two goals: (1) to analyze its population structure, and (2) to assess the genetic drift and natural selection hypotheses by evaluating the relative contribution of climatic variables and geographic distance to patterns of genetic variation. Our genetic structure analyses showed geographically distant individuals along the Eastern Sierra Nevada Mountains of California formed a consistent genetic cluster, while individuals from the southern Mojave Desert showed high genetic structure, with more genetic clusters emerging as the number of expected demes increased. Individuals from the western Transverse Range mountains were intermixed between the Eastern Sierra Nevada and southern Mojave Desert

populations. We found that desert wash dissection and isolation by distance are key drivers of genetic differentiation in *L. parryae*, supporting Wright's original findings of fine-scale genetic differentiation. However, we found that isolation by environment, particularly related to precipitation and temperature variability, also plays a role in genetic divergence, potentially reflecting the influence of fluctuating rainfall patterns on flower color morph variation. Overall, our study provides evidence that genetic divergence in *L. parryae* is driven by both genetic drift and selection, with both evolutionary forces contributing to the observed genetic structure across the species' geographic range.

Introduction

The flower color polymorphism in the minute desert annual *Linanthus parryae* inspired pioneering research in evolutionary biology. Populations across the geographic range of this species can be polymorphic, including both blue and white-flowered individuals, or monomorphic with one dominant flower color (Figure 4.1). Foundational evolutionary ideas explaining the maintenance of this polymorphism were developed based on field studies of this species (Epling and Dobzhansky 1942, Wright 1943a, b, Epling et al. 1960, Wright 1978). Epling and Dobzhansky (1942) collected a large dataset on the distribution of the two flower morphs and quantified changes in the frequency of white and blue morphs across 70 miles of the southern Mojave Desert. They concluded that genetic drift maintained this pattern of variation and that populations with differing color morph frequencies represented "microgeographic races" with limited gene flow between them. Wright (1943b) used these same data to illustrate his model of isolation by distance in an empirical system for the first time, demonstrating that

stochastic variation can become fixed in small populations with limited migration. These findings led to the development of the widely used *F*-statistic (F_{st}) to describe the genetic structure of natural populations and provide insights into migration rates, inbreeding coefficients, and other features of population demography (Wright 1943a).

After decades of study, Epling re-examined previous work and suggested that selective forces, and not only genetic drift, must contribute to the persistence of the two color morphs (Epling et al. 1960). Years later, Turelli et al. (2001) mathematically showed that the polymorphism in *L. parryae* can persist with yearly variation in fitness and a seed bank, and Schemske and Bierzychudek (2001, 2007) supported this finding with field experiments and observations. Rainfall variation across years provided fitness advantages to one morph or the other across opposite sides of a ravine, but such fitness differential was only found at one site. As a result, Schemske and Bierzychudek (2007) concluded that Wright was wrong and that natural selection instead of genetic drift was the microevolutionary force maintaining flower color variation in *L. parryae*.

These studies motivated debates about the degree to which genetic drift or natural selection contributes to the maintenance of genetic variation within and across natural populations (Ishida 2017). Schemske and Bierzychudek (2007) found that allozymes did not differ in frequencies across regions of their study site dominated by white versus blue flowers, showing that phenotypic differentiation was greater than putatively neutral genetic variation. Nevertheless, previous studies had used phenotypic data as a surrogate of genetic variation.

Prior investigations included only a portion of *L. parryae's* geographic range and could not make conclusions about the role of evolutionary forces across the species entire distribution. An explicit genetic approach at the landscape scale can provide valuable insight in assessing the relative contribution of genetic drift and natural selection to genetic variation in *L. parryae*.

Genetic drift and natural selection have very different impacts on the fitness of local populations. The relative roles of genetic drift and natural selection in shaping genetic variation can be distinguished by the spatial patterns of genetic variation they produce. When genetic drift is the primary force driving genetic divergence, the resulting pattern is typically explained by a model of isolation by distance, where genetic differences increase with geographic separation due to the random fixation of alleles in isolated populations (Wright 1943a, Slatkin 1993). By contrast, when natural selection drives genetic divergence, the pattern is better explained by a model of isolation by environment, where genetic differences are correlated with environmental variation rather than geographic distance (Wang and Bradburd 2014). In this scenario, populations in similar environments may converge genetically, regardless of their geographic proximity, as selection favors specific adaptations to local conditions.

The life history of *L. parryae* combined with the spatial and temporal patterns of variation in climate in the deserts of California provide an excellent opportunity to assess the roles of isolation by distance and by environment driving genetic variation across the geographic range of this iconic species. *Linanthus parryae* covers the Mojave Desert floor in years with high relative rainfall, likely connecting populations that remain isolated in drier years. Rain is

infrequent in the desert, and in most years, only scattered individuals are found across the landscape. Those populations that germinate in dry years likely undergo a bottleneck, but since many seeds can persist in the soil for a decade or more, populations may function demographically like longer-lived perennials (Epling et al. 1960). Epling thus predicted that these characteristics of *L. parryae* result in a larger effective population size than expected in each locality. As a result, it is plausible that a signature for isolation by distance is limited in *L. parryae.*

Conversely, other life history strategies of *L. parryae* predict high genetic differentiation across populations and a strong signal of isolation by distance. Pollen and seed dispersal mechanisms and barriers to migration could drive patterns of gene flow and genetic divergence in *L. parryae* (Levin 1981). In this species, pollen does not seem to be a vector for long distance dispersal. Wind pollination is unlikely, and the Melyrid beetles that visit *L. parryae* flowers travel only short distances between flowers, lingering inside and moving to another individual a few feet away (Epling et al. 1960, Schemske and Bierzychudek 2001). Likewise, *L. parryae* seeds drop into the soil close to the mother plant and do not travel far from the site where they germinate (Epling et al. 1960). Seeds are almost microscopic, measuring about one millimeter, making them an unlikely food source for animal dispersers. Wind gales are common in the desert, but the winds do not move the low growing plants and their seeds at the soil surface (Epling et al. 1960). Furthermore, *L. parryae* grows in ephemeral washes and alluvial fans, habitats that are periodically disturbed by rainfall events. During heavy precipitation events, water and sediment flow down these channels, often in unpredictable and shifting paths (Epling and Dobzhansky

1942). These ephemeral washes with unpredictable courses may also move seeds downhill, dissecting populations into smaller, isolated subpopulations. Therefore, while the large effective population size facilitated by the seedbank may buffer against genetic drift (Masel 2011), desert washes and limited dispersal may increase the effect of genetic drift and isolation by distance in *L. parryae*. Which of these patterns is prevalent in driving the genetic variation and population structure of this species is not known.

Natural selection may also drive patterns of genetic variation and divergence in *L. parryae*, consistent with a model of isolation by environment. If natural selection is strong enough to be detected at a local scale in polymorphic populations as suggested by Schemske and Bierzychudek (2001, 2007), we would expect this signal of selection to be apparent at the landscape scale across the geographic range of *L. parryae*. In desert environments, where precipitation patterns are highly variable, populations of *L. parryae* with different flower colors may persist due to the selective advantages conferred by these color variations in response to local environmental conditions. Specific flower colors may offer better adaptation to yearly rainfall variation, with fluctuating selection on drought tolerance having a pleiotropic effect on flower color (Schemske and Bierzychudek 2001, 2007). These adaptations would be driven by natural selection rather than geographic distance alone, as genetic variation would align more closely with environmental conditions than spatial separation. A genetic analysis is needed to determine whether isolation by environment, in particular precipitation, drives genetic divergence in *L. parryae.*

Another observation points to selection as a plausible key microevolutionary force behind genetic variation in *L. parryae*. Assuming that flower color is controlled by a single locus, Epling et al. (1960) predicted selection driving flower color distribution would result in homozygosity in monomorphic populations. Although flower color variation has not been well characterized across the entire geographic range of *L. parryae*, it does not appear to follow a geographic pattern. The presence of monomorphic and polymorphic populations throughout the species' range suggests that flower color is not structured by spatial distance; instead environmental variables may influence patterns of genetic and phenotypic diversity in *L. parryae*.

In this study, we sequenced whole genomes of individuals of *L. parryae* across its geographic distribution, conducting the first comprehensive range-wide study of genetic variation in this species. We used this data to address the hypotheses that genetic variation and differentiation in *L. parryae* at the landscape scale are primarily driven by (a) genetic drift or (b) natural selection. Specifically, we asked (1) Is there evidence of population structure in *L. parryae* across its geographic range? (2) Is geographic distance correlated with genetic distance, consistent with the isolation by distance model? (3) Are differences in environmental variables correlated with genetic distance, consistent with the isolation by environment model? These results can shed light on the phenotypic, theoretical and experimental work that built the historical cases for the forces of genetic drift or selection maintaining color polymorphism in *L. parryae*.

Methods

Sampling

Linanthus parryae samples were collected as part of the California Conservation Genomics Project, a state-wide landscape genomics effort (Shaffer et al. 2022). Between 2019 and 2022, we collected 27 individuals from different populations across the distribution of *L. parryae* (Figure 4.2). The samples were either flash frozen in liquid nitrogen and stored at -80C, or dried in silica gel. Locality and petal color information was recorded for each individual (Supplemental Table 1).

DNA Extraction and sequencing

DNA extractions were done by the mini-core facility of the California Conservation Genomic Project (CCGP) at UCLA with the NucleoMag Plant kit for DNA purification (Macherey-Nagel, Düren, Germany) with PVP and Proteinase K added at digest. Libraries were prepared using KAPA EvoPlus Kit (Roche, Basel, Switzerland). We used an initial digest time of 5 minutes, cleaned DNA with a solution of Solid Phase Reversible Immobilization (SPRI) beads at 0.8x ratio, and quarter reaction volumes. Libraries were size selected for fragment length of 400-800 bp on a PippinPrep with a 2% Agarose Gel (Marker L). Whole-genome sequencing was done on an Illumina NovaSeq X 25B using 150 bp, paired-end sequencing at the Center for Applied Technologies at UCSF.

Data filtering

We aligned reads to the *L. parryae* genome (Anghel et al. 2022) using BWA (Li and Durbin 2009) with the bwa mem command. We removed duplicate reads with sambamba v1.0 (Tarasov et al.

2015). We called variants using Genome Analysis Toolkit (GATK) and the HaplotypeCaller and GenotypeGVCFs functions (McKenna et al 2010), as implemented in the snpArcher Snakemake pipeline (Mirchandani et al. 2024). We applied GATK filters using VariantFiltration on SNPS, indels and mixed variants with quality by depth (QD) <2 and confidence that a site is a variant (QUAL) > 30. We excluded SNPs with a read position rank-sum test score (ReadPosRankSum) < -8.0, Phred-scaled probability that there is strand bias at the site (FS) > 60, strand odds ratio (SOR) > 3.0, root mean square mapping quality over all the reads at a site (MQ) < 40.0, and mapping quality rank-sum test score (MQRankSum) < -12.5. We excluded indels or mixed variants with ReadPosRankSum < -20.0, FS > 200.0, SOR > 10.0. We used bcftools (Danecek et al. 2021), to filter for minor allele frequency > 0.01, missingness across individuals < 0.75, and retain only biallelic sites.

Subsequent filtering and analyses of the resulting 100,098 SNPs were performed in the R package *algatr* v 1.0.0 (Chambers et al. 2023). For analyses that required no missing data, we imputed missing genotypes with sparse nonnegative matrix factorization (snmf) using the R package LEA (Frichot and Francois 2015) run through the 'str_impute' function in *algatr*. Cross entropy was calculated with *K* values from 1 to 4. We used *K* = 1 for imputation because it had the lowest cross entropy values. We calculated genetic distance using Euclidean distance (Shirk et al. 2017). The filtered VCF file was pruned for variants in linkage disequilibrium (LD) using the 'ld_pruned' function, resulting in 15031 SNPs with 1.43% missing data. To perform LD pruning, we used a minor allele frequency of 0.05, pruned out variants that had more than 0.9 correlation in each window, and used a maximum of 10 SNPs in each window.

Environmental data processing

We downloaded 19 climatic layers at 30 arc-seconds from the WorldClim database (Fick and Hijmans 2017) using the 'get_worldclim' function in the *algatr* R package and extracted the variables in the tiles corresponding to our sample localities with a 1% buffer area around each point. We corrected for collinearity between the climatic variables by reducing the dimensionality of the data with a principal component analysis (PCA) using the 'rasterPCA' function in the *RStoolbox* R package (Leutner et al. 2024).

Genetic structure

Using the linkage disequilibrium pruned matrix containing 15,031 SNPs, we transformed it into a genlight object using the R package *vcfR* and the function 'vcfR2genlight' (Knaus and Grunwald 2017). We used the 'ploidy' function to specify that *L. parryae* is a diploid species and performed a PCA of genomic variation using the function 'glPca' in the R package *adegenet* (Jombart 2011). To determine genetic clusters and interpolate those clusters in geography, we ran TESS3r (Caye et al. 2006) with the same SNP data which we converted to a dosage matrix using 'vcf_to_dosage' function in *algatr*. We selected the optimal partition of the data into genetic clusters (K = 1 to K = 12) using the 'tess_ktest' function and generated a matrix with the ancestry coefficients using the 'qmatrix' function. To project the geographic coordinates of each sampling locality onto spherical space, we transformed our coordinate data as simple features using the 'st_to_sf' function, then projected the data to the format NAD83 / California Albers (EPSG 3310) using the function 'st_transform' in the *sf* R package (Pebesma 2018). To interpolate the geographical extent of the genetic clusters, we used a process called kriging with

the function 'tess_krig' in *algatr*. We visualized the interpolated surfaces of the genetic clusters using the 'map.max' function.

Isolation by distance versus isolation by environment

We used generalized dissimilarity modeling (GDM) to test the relative contribution of isolation by distance (IBD) and isolation by environment (IBE) (Ferrier et al. 2007, Fitzpatrick et al. 2022). To run this analysis, we used the 'gdm_run' function in the *algatr* R package. We used the three environmental PCs and Euclidian geographic distance as predictors of compositional dissimilarity (i.e. genetic distance) between individuals. We ran a full model including all the predictor variables.

We visualized genetic variation by using the GDM model to investigate genetic turnover across gradients of the predictor variables. We used the 'gdm_map' *algatr* function to transform the original environmental layers using the GDM fitted functions by rescaling the rasters to a common biological scale, which facilitates directly comparing the effects of the environmental variables on genetic turnover (Fitzpatrick and Keller 2015, Fitzpatrick et al. 2022). The function uses PCA to reduce the environmental and geographic variables into three axes, and then assigns an RGB hue to each of the resulting axes. Points with similar hues are predicted to be genetically similar based on their environmental and geographic predictors. We projected the environmentally predicted genetic turnover onto a spatial map. We also generated a biplot of PC1 and PC2 of the GDM predictors, with the original environmental and geographic distance variables overlaid as vectors.

Results

Genetic structure

Visual inspection of the PCA of genomic variation readily revealed three main clusters, which largely reflected broad-scale geographic structure (Figure S4.1). PC1 explained 8% of the variation and separated individuals along latitude, suggesting three groups: one including southern Mojave Desert individuals, one including western Transverse Range individuals, and one including Eastern Sierra individuals. PC2 explained about 6% of the variation and separated individuals in the southern Mojave Desert area. The automatic selection of the number of genetic clusters (*K*) chose a *K* = 3 as the optimal partition (Figure 4.3). These three clusters differed from those suggested by the PCA (Figure S4.1). One cluster included the Eastern Sierra and western Transverse Range individuals, a second cluster included individuals from the southern Mojave Desert west of the Mojave River, and a third cluster included individuals from the southern Mojave Desert east of the Mojave River (Figure 4.3A-B). However, the cross-entropy criterion curve did not show a flattening at any point in the range of *K* between 2 and 12. We report genetic clustering plots for a range of *K* = 3-5 (Figure 4.3).

The southern Mojave Desert group of individuals were dissected by the Mojave River with *K* = 3 (Figure 4.3B) and fragmented into four clusters with *K* = 5 (Figure 4.3F). The division of the southern Mojave Desert cluster beginning with *K* = 3 only affected individuals in the alluvial fan of the Mojave River or to the east of it, while the individuals to the west continued to form a cohesive group irrespective of the values of *K*. The individuals from the east of the Sierra

Nevada Mountains (Owens Valley and northern Mojave Desert, hereafter referred to as Eastern Sierra) formed one cluster consistently in all the *K* iterations (Figure 4.3A-F). The individuals from the western Transverse Range were admixed between the populations in the Eastern Sierra and the western cluster of the southern Mojave Desert in all *K* iterations, except with *K* = 4, when they formed a unique cluster (Figure 4.3C-D). Two individuals furthest north in the alluvial fan of the Mojave River were consistently admixed between the west and the east of the Mojave River individuals (Figure 4.3A-F, samples 5 and 6).

Climatic variation across the landscape

The first three PCs of the WorldClim climatic layers explained 93.2% of the variation across all the climatic variables, with PC1 explaining 58.8%, PC2 30.7%, and PC3 3.8% of the variation, respectively (Figure 4.2; Table 4.2). PC1 had contributions from almost all the climatic variables, with the strongest positive contributions from precipitation seasonality (BIO15) and mean diurnal range (BIO2), and the strongest negative loadings from temperature seasonality (BIO4) and warmest month maximum temperature (BIO5). The highest loading was precipitation seasonality, which measures variability in the monthly total precipitation (O'Donnell and Ignizio 2012). PC2 had the strongest positive contributions from temperature annual range (BIO7) and wettest month precipitation (BIO13), and the strongest negative loadings from driest month and quarter precipitation (BIO14, BIO17). PC3 had the strongest positive contributions from precipitation in wettest and warmest quarters (BIO16, BIO18) and the strongest negative loadings from annual precipitation (BIO12) and wettest month precipitation (BIO13).

Isolation by distance versus isolation by environment

The GDM analysis showed a significant contribution from both IBE and IBD. The full GDM model incorporating all the predictor variables explained 32.94% of the deviance in genetic turnover (Table 4.3). The relative contribution of each predictor is indicated by the coefficient value and is visualized as the height of the GDM fitted functions (Figure 4.4). PC1, representing environmental variation (Table 4.2), had the tallest spline curve with a coefficient of 0.55, indicating that this predictor had the highest relative contribution to genetic turnover (Figure 4.4B). Genetic turnover was rapid at lower values of PC1 and slowed down to zero at higher values, as shown by the curve flattening out after the midpoint of the predictor variable axis (Figure 4.4B). The other two environmental PCs (Table 4.2) did not influence genetic turnover (Table 4.3), shown as flat lines (slope 0) in Figures 5C and 5D. Hereafter, we refer to environmental PC1 as the environmental axis. Geographic distance contributed to the model with a coefficient of 0.51. The steep slope at low values of the geographic distances indicates that the largest changes in genetic turnover were at shorter geographic distances. The longest distance between individuals sampled was 230 miles, so the steep slope between 0 and 0.5 in geographic distance corresponds to individuals separated by less than 33 miles (Figure 4.4A).

Predicting adaptive genetic variation

Figure 4.5A shows the spatial interpolation of genetic turnover based on the GDM PCA. Each GDM transformed axis is assigned red, green, or blue hues and is plotted on a map of Southern California with sample localities. The colors inside the points represent predicted genetic dissimilarity at those locations based on the environmental and geographic predictors. A visual

inspection of Figure 4.5A shows predicted rapid genetic turnover associated with environmental gradients represented by the sharp transitions of color. The points in areas in pink corresponded to localities at low elevations, predicted to be genetically dissimilar to points in areas in blue/green, corresponding to higher elevation locations.

Figure 4.5B displays the contribution of environmental and geographic variables to the expected genetic turnover in a two-dimensional space. The overlaid vectors show the direction and magnitude of each variable, and the colors of the points match the colors of the map in Figure 4.5A. The solid points represent the genetic dissimilarity associated with climatic variables of the samples included in this study, while the background colors represent predicted genetic dissimilarities in unsampled areas of the geographic space. The biplot shows that both the environmental and geographic axes were associated with genetic turnover. The biggest turnover associated with the environmental axis was predicted in the western Transverse Range localities (blue hue) and Eastern Sierra (purple hue), with little turnover in the clustered southern Mojave Desert localities (pink hue, Figure 4.5B). The western Transverse Range and Eastern Sierra individuals varied markedly both along environmental and geographic axes.

Discussion

This study of the genetic variation across the landscape of *Linanthus parryae* shows that both the genetic drift and natural selection scientists were right! The level of flower color polymorphism in *L. parryae* has captivated evolutionary biologists for generations, serving as a key model in debates over the relative roles of natural selection and genetic drift in maintaining

this polymorphism (Ishida 2017). Previous studies of *L. parryae's* floral color variation have come to contrasting conclusions (Wright 1943b, Schemske and Bierzychudek 2001, 2007). The goal of the current study was to revisit this debate using a landscape genomic approach and determine the contribution of isolation by distance (i.e., genetic drift) and isolation by environment (i.e., natural selection) across the geographic distribution of *L. parryae* in the deserts of California. Overall, we found a strong signal of large-scale geographic structure, with three main genetic clusters corresponding to (a) populations along the Eastern Sierra, (b) populations in the southern Mojave Desert, west of the Mojave River, and (c) populations in the southern Mojave Desert, east of the Mojave River. Populations in the western Transverse Range showed admixed composition between the Eastern Sierra and western Mojave genetic clusters (Figure 4.3). When we examined whether isolation by distance or by environment would explain this pattern of genetic structure, we found evidence for isolation by environment, largely driven by fluctuations in temperature and precipitation. The contributions from environmental variation were nonlinear, with values at the low end of the axis accounting for a majority of genetic turnover (Figure 4.4). We also found a nonlinear contribution for isolation by distance, with pairs of individuals separated by less than 33 miles having the highest genetic turnover (Figure 4.4). Taken together, our results provide evidence that genetic divergence in *L. parryae* is shaped by both genetic drift and selection, with both microevolutionary forces contributing to the genetic structure of this species across its geographical range.

Genetic structure

Linanthus parryae shows large scale genetic structure across regions within its species range. We found that the strongest signal of genetic structure partitioned the samples into 3 genetic clusters: two clusters in the southern Mojave Desert, and a third genetic cluster in the Eastern Sierra and western Transverse Range (Figure 4.3A-B). The barrier that separates the Eastern Sierra from the southern Mojave Desert populations corresponds to the Garlock fault, which runs from the western tip of the Mojave Desert at the San Andreas fault northeast to Death Valley. The area to the north of the Garlock fault includes parts of the northern Mojave Desert, the Sierra Nevada Mountains and Owens Valley, while the area to the south of the fault includes the rest of the Mojave Desert. Though few studies have looked at genetic structure in other Mojave Desert plants, populations of fleaworts (*Plantago ovata*) north of the Garlock fault are also genetically distinct from the rest of the Mojave Desert populations (Shryock et al. 2021), consistent with our results. In a study looking at genetic diversity in 17 Mojave dwelling animal taxa, most species' ranges did not expand north of the Garlock fault (Vandergast et al. 2013). This trend suggests that the Garlock fault may be a barrier that separates populations or limits species dispersal, by dividing the habitat through movement of the tectonic plates. The mountains to the north of the Garlock fault may also limit gene flow from the Mojave Desert northward to the Eastern Sierra populations.

The genetic cluster that consistently included the same samples across *K* values represents the populations in the southern Mojave Desert west of the Mojave River (Figure 4.3B, D, F, shown in green, blue, magenta, respectively). These populations show little admixture across *K* values

and do not divide with increasing *K* values. One possible explanation for this result is that the southern Mojave Desert area has been a refuge to *L. parryae* through periodic climatic changes during the Pliocene and Pleistocene. The "Mojave Assembly Model" proposes that 2-4 million years ago, when the Transverse Range and Sierra Nevada mountains were undergoing rapid uplift, the location of the current Mojave Desert changed geologically and climatically (Axelrod 1972, Bell et al. 2010, Graham et al. 2013). Desert taxa found refuge in arid areas like the Antelope and Phelan Peak basins (Bell et al. 2010), both of which correspond to the *L. parryae* populations to the west of the Mojave River. Then during the Pleistocene, as the Mojave River began draining into the Mojave Desert, the Phelan Peak basin divided into the Victorville and the Lucerne Valley basins (Bell et al. 2010), with the latter basin corresponding to the east of the Mojave River *L. parryae* populations. The genetic structure found in the southern Mojave Desert and the genetic admixture we found in populations in the floodplains of the current Mojave River is broadly consistent with this dynamic geologic history. Epling and Dobzhansky (1942) predicted that the Mojave River acts as a barrier restricting gene flow and that a continuous *L. parryae* population is dissected into "microgeographic races", while Epling (1960) described that outwashes after heavier rains could transport seeds downhill and be a mechanism for dispersal. Our results are consistent with the hypothesis that the occasional floods in desert streams may be a mechanism of reproductive isolation leading to fine-scale population genetic structure in *L. parryae*.

Isolation by environment

Isolation by environment (IBE) contributed to genetic structure across the range of *L. parryae* (Figures 4-6). With IBE, genetic connectivity can be high among individuals in similar environmental conditions regardless of geographic distance (Wang and Bradburd 2014). Selection against immigrants who are not locally adapted or successful dispersal to environments similar to those in the source population may result in a correlation between genetic and environmental distances (Sexton et al. 2014). In our study, isolation by environment was primarily associated with precipitation and temperature variability (Figure 4.5, Table 4.2). This pattern was especially important at the lower values of environmental variation with no turnover past the midpoint of this axis of variation (Figure 4.4B). Lower values of the environmental axis correspond to low precipitation seasonality, high temperature seasonality, low diurnal temperature ranges, and high maximum temperatures. This pattern indicates the combination of these climatic variables at low ranges of their gradient are associated with high genetic variation, while the high ranges have no association with genetic turnover in *L. parryae*. The highest positive loading of the environmental axis is precipitation seasonality, with low to medium rainfall fluctuation having high relative contribution to genetic turnover, which suggests that the species reaches a threshold of maximum precipitation variability that no longer contributes to genetic variation. The saturation of the effect of precipitation variability suggests that the degree of precipitation variability may be a key selective force driving genetic similarities across *L. parryae*. In support of this hypothesis, previous studies have found that rainfall variability patterns maintain the two color morphs in a *L. parryae* population, with the blue having a fitness advantage in the dry years and the white in the wet years (Schemske and

Bierzychudek 2001, 2007). Though flower color itself is likely not under selection, a linked or pleiotropic trait related to water uptake may have a differential fitness advantage with precipitation variation (Schemske and Bierzychudek 2007). This disparity in fitness between morphs associated with rainfall patterns may vary both temporally and spatially, maintaining genetic variation across the landscape of *L. parryae*. Further genomic sampling in combination with genomic association analyses to environmental variables is necessary to test this hypothesis rigorously.

Isolation by environment seemed to be most pronounced in the Transverse Range individuals, which were outliers in terms of predicted genetic dissimilarity based on climatic variables (Figure 4.5B). Despite the predicted genetic distinctiveness of these individuals, genetic structure analyses found that the western Transverse Range individuals were admixed between the Eastern Sierra and the western Mojave Desert populations (Figure 4.3, samples 25 and 27). While the Transverse Range individuals were predicted to be most genetically dissimilar based on environment, they do not form a distinct genetic cluster in most K values. A possible explanation for this finding is that the admixed origin of these individuals enabled them to occupy a new environmental space. Admixture may have provided the genetic variation or novel trait combinations that were adaptive in this new environment (Rius and Darling 2014). This potential mechanism of isolation by environment could serve as a natural experiment to understand how flower color segregates in these admixed populations, and whether pleiotropy or linkage are responsible for the association between precipitation and flower color in *L. parryae*.

Isolation by distance

We also found evidence for isolation by distance, with limited gene flow across the species range being one of the mechanisms contributing to genetic differentiation. What factors contribute to limited gene flow in this species? *Linanthus parryae* seeds are small, with no morphological features that would facilitate wind or animal dispersal, and likely land closely to the mother plant. Although long distance dispersal events can occur with wind gusts, they are unlikely to contribute to the genetic structure of *L. parryae*. Furthermore, dissection of the desert landscape by ephemeral streams may be an important mechanism that isolates populations on either side of the washes. Dry washes in the Mojave Desert have been shown to influence gene flow in *Acacia greggii*, another plant species (Kaddis et al. 2016). This shrub occurs only in desert washes which may promote genetic connectivity by dispersing its seeds downstream. By contrast, ephemeral streams may transport seeds of *L. parryae* into unsuitable habitats, leading to population dissection rather than connectivity. Indeed, we found genetic structure around the Mojave River, a mostly dry stream that floods occasionally during periods of high precipitation (Durbin and Hardt 1974), suggesting that these streams likely act as geographic barriers that dissect the seemingly continuous distribution of *L. parryae*. Dating the divergence of the population isolation events using environmental (e.g. fossils, packrat middens) and genetic (e.g. coalescent inferences) data could help us determine when these populations were separated and whether occasional flooding of dry washes may have led to micro-allopatry. Overall, our results are consistent with Wright's findings that *L. parryae* is genetically differentiated at a fine spatial scale (Wright 1943b).

Caveats

Including more individuals in our sample size may have helped identify more fine scale population structure across the *L. parryae* range. In addition, including both flower color morphs would have allowed us to more directly test the questions asked in the classic debates about the processes driving morph distribution in this species. The white flower morph is much more abundant and in years of high rainfall, it connects previously isolated populations by mass germinating across the landscape. The blue flower morph is rarer, and even in years with good germination, it has a patchy distribution. This may lead to higher gene flow between white morphs across the landscape and more isolation by distance in blues. Analyses including blue and white morphs will help us determine the relationship between color phenotype and genetic divergence in this species. Despite these limitations and opportunities for further investigation, the current sampling is sufficient to test the specific hypotheses we proposed in this study.

Conclusion

The species-wide patterns of genetic variation of *L. parryae* were striking, with the northern and southern populations separated by the Garlock Fault, and further genetic differentiation associated with the Mojave River. Isolation by environment associated with climatic variability had the highest contribution to genetic divergence across the species, which may reflect the impact of fluctuating rainfall patterns on flower color distribution found in previous studies in *L. parryae* (Epling et al. 1960, Schemske and Bierzychudek 2001, 2007). Isolation by distance also had a contribution to genetic differentiation in the species, supporting Wright's findings of fine-

scale genetic differentiation (Wright 1943b, 1978). Overall, both genetic drift and selection contributed to genetic divergence across the species' geographic range.

Linanthus parryae is an example of how desert plant may be more resilient than expected to increased rainfall stochasticity (Salguero-Gómez et al. 2012), and environmental variability may maintain genetic variation in populations (Hedrick 2006). In addition, local adaptation seems to be high in arid areas (Baughman et al. 2019). For these reasons, understanding genetic variation across the landscape in desert species can illuminate how dryland floras could survive an even more extreme future.

Acknowledgments

This work was supported by the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224]. We thank Sarah Jacobs, Isaac Lichter-Marck, and Susan Fawcett for help with field collections. We are grateful to Andrew Tully who extracted DNA and prepared libraries at the mini-core facility of the California Conservation Genomic Project at UCLA, and Erik Enbody, at University of California, Santa Cruz, who performed variant calling.

Figures

Figure 4.1. Linanthus parryae individuals have either blue or white petals. This color polymorphism varies across the landscape and across time. Some populations are polymorphic, while others are monomorphic with only one color. The frequency of colors can vary at a local scale and year to year with different rainfall conditions.

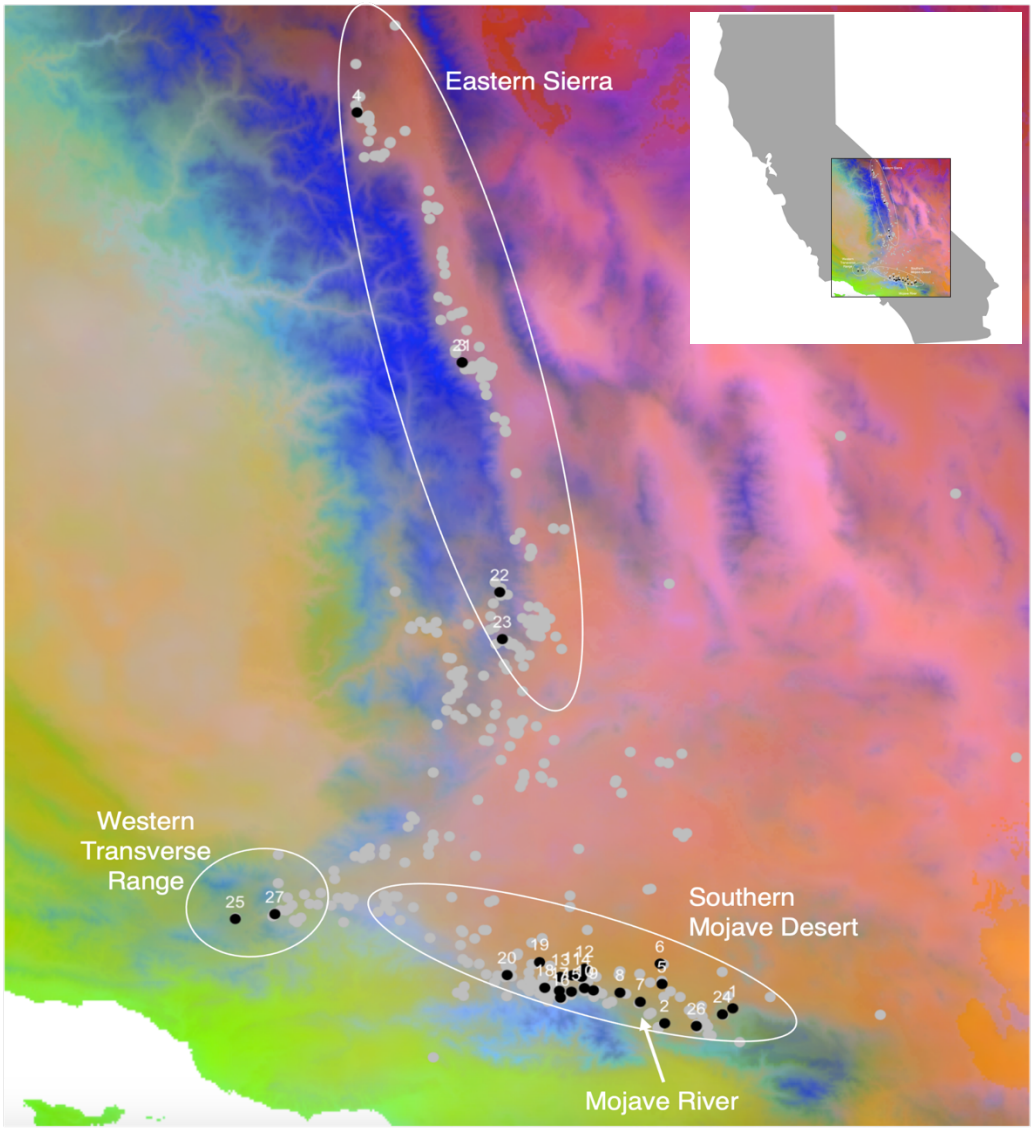


Figure 4.2. Map of sampling localities of the 27 *L. parryae* individuals included in this study (black) plotted against all occurrences of the species (gray). The background map represents the reduced dimensionality of the 19 environmental layers from the WorldClim database into three principal components (PCs) projected on a map of southern California (see top right inset). PC1 is represented by a red hue and it corresponds to the area occupied by the Mojave Desert and Great Basin lowlands. PC2 is represented by a green hue, and it corresponds to the coastal or foothill areas of California. PC3 is represented by a blue hue and it corresponds to the higher elevation areas of California. See Table 4.2 for loadings of climatic variables on each PC.

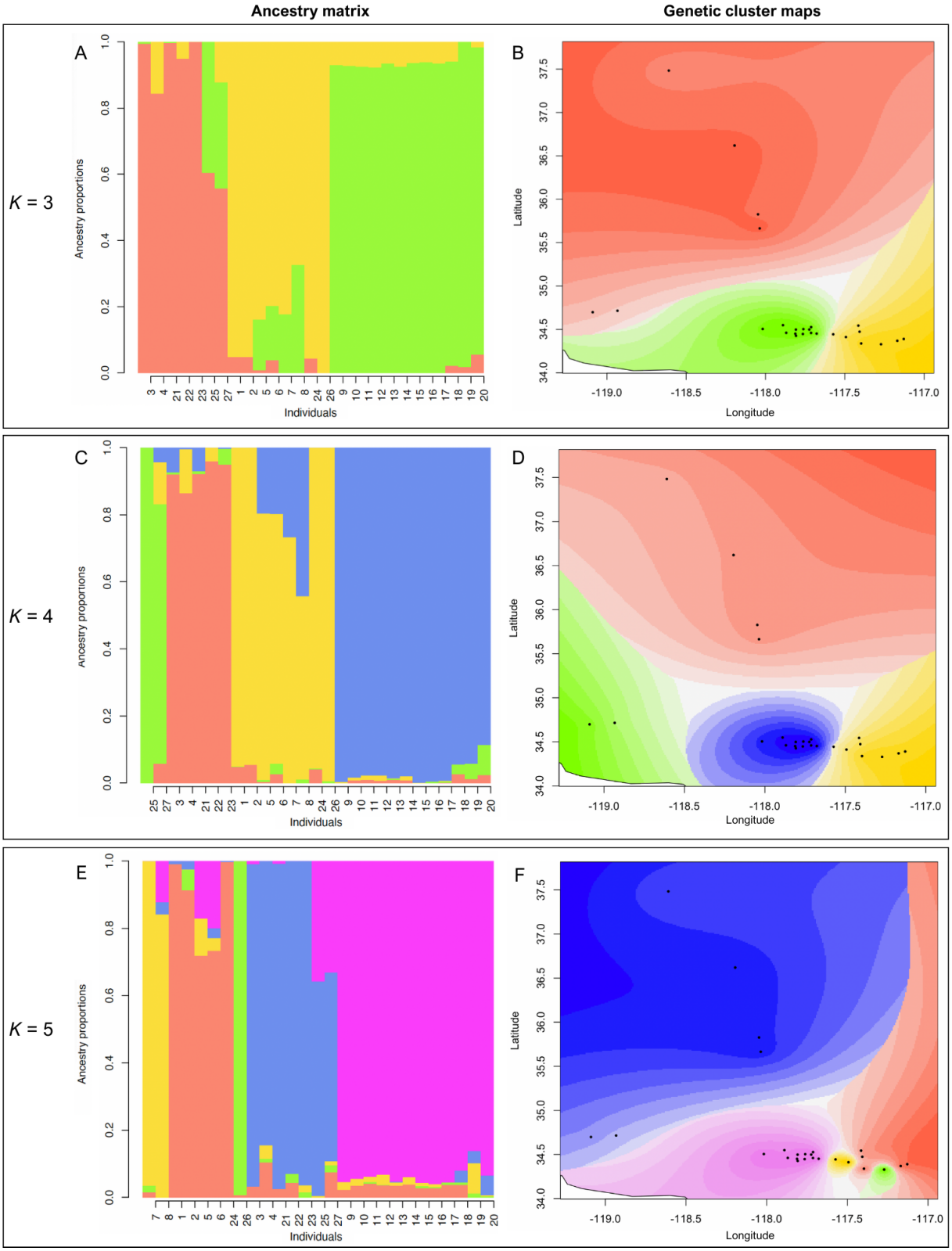

Ancestry matrix
Genetic cluster maps
K = 3
A
B
K = 4
C
D
K = 5
E
F
Ancestry proportions
Individuals
Latitude
Longitude

Figure 4.3. Spatial population structure of *L. parryae* with different values of genetic clusters, *K* =3 (A, B), K =4 (C, D), K=5 (E ,F). In the left column (A, C, E), the colors represent genetic clusters, and each bar represents the ancestry proportion of genetic clusters for each individual. The right column (B, D, F) shows spatial interpolation of ancestry coefficients across the range of the samples for each *K*, indicated with the same colors as in the bar plot on the left column.

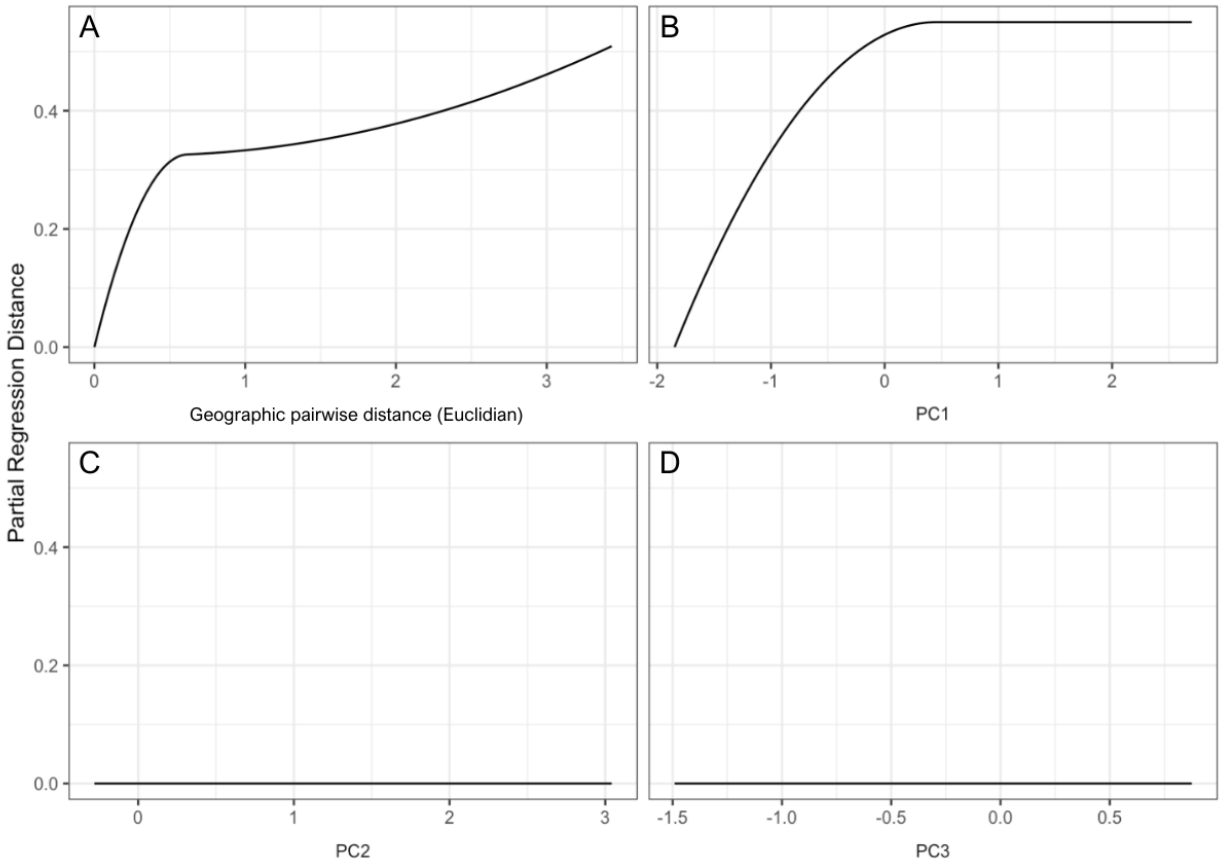


Figure 4.4. GDM fitted functions, modeling the relationships between predictor variables (geographic distance and three environmental PCs, shown individually) and the response variable (genetic distance) when all the other predictors are constant. The height of the curve was highest for PC1, indicating that this climatic variable had the highest contribution to genetic differences. Geographic distance also contributed to genetic differences. Genetic dissimilarities per unit were higher at shorter geographic distances than at longer geographic distances, illustrated by the steeper slope at the lower end of the x-axis.

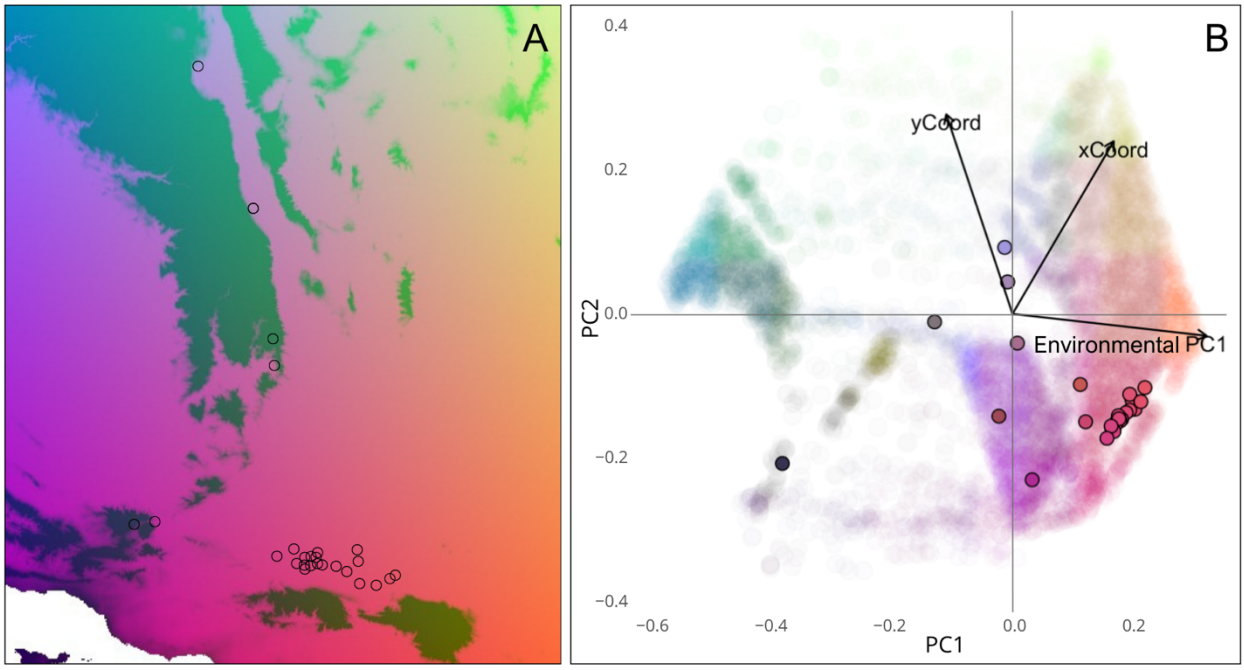


Figure 4.5. (A) Spatial interpolation of genetic dissimilarity associated with environmental gradients predict areas of rapid genetic turnover represented by sharp color changes across the landscape. The RGB hues represent the reduced dimensionality axes of the GDM predictors. Changes in color indicate regions where genetic composition is predicted to change rapidly in response to environmental gradients.
(B) Principal Component Analysis biplot of the GDM. The clustered points in the lower right quadrant in pink correspond to individuals from the southern Mojave Desert (i.e. southern individuals), while the dispersed points with a purple hue correspond to collections along the Eastern Sierra and western Transverse Range (i.e. northern individuals). The most extreme points in purple correspond to individuals in the western Transverse Range (lower left quadrant). The environmental axis contributed to the separation between the clustered southern individuals and northern individuals, while geographic distance (latitude and longitude) also had a high contribution to sample genetic turnover.

Tables

Table 4.1. Samples included in this study, including sample identifier, location and flower color.

Sample #	Sample ID	Longitude	Latitude	Location	Flower color
1	IGA131.11	-117.1283	34.3900	S Mojave Desert - East	white

2	IGA132.1	-117.3967	34.3385	S Mojave Desert - East	white
3	IGA138.1	-118.1952	36.6195	E Sierra Nevada	white
4	IGA141.10	-118.6096	37.4827	E Sierra Nevada	white
5	IGA181.2	-117.4072	34.4739	S Mojave Desert - Center	white
6	IGA183.2	-117.4148	34.5441	S Mojave Desert - Center	na, in bud
7	IGA184.1	-117.4930	34.4121	S Mojave Desert - Center	white
8	IGA185.2	-117.5731	34.4440	S Mojave Desert - Center	white
9	IGA186.2	-117.6776	34.4519	S Mojave Desert - West	white
10	IGA187.1	-117.7130	34.4597	S Mojave Desert - West	white
11	IGA188.1	-117.7613	34.5035	S Mojave Desert - West	white
12	IGA189.1	-117.7124	34.5271	S Mojave Desert - West	na, in bud
13	IGA190.1	-117.8084	34.4984	S Mojave Desert - West	white
14	IGA191.2	-117.7240	34.4987	S Mojave Desert - West	white
15	IGA192.1	-117.7644	34.4473	S Mojave Desert - West	white
16	IGA193.1	-117.8076	34.4269	S Mojave Desert - West	white
17	IGA194.2	-117.8120	34.4507	S Mojave Desert - West	white
18	IGA195.1	-117.8694	34.4609	S Mojave Desert - West	white
19	IGA196.2	-117.8900	34.5492	S Mojave Desert - West	white
20	IGA197.1	-118.0177	34.5049	S Mojave Desert - West	white
21	IGA198.1	-118.1951	36.6196	E Sierra Nevada	na, in bud
22	IGA199.3	-118.0472	35.8266	E Sierra Nevada	white

23	IGA200.2	-118.0367	35.6646	E Sierra Nevada	white
24	IGA220.1	-117.1691	34.3691	S Mojave Desert - East	white
25	IGA221.2	-119.0894	34.6985	W Transverse Range	white
26	IGA222.1	-117.2718	34.3289	S Mojave Desert - East	white
27	IGA224.1	-118.9334	34.7153	W Transverse Range	na, in bud

Table 4.2. Loadings of each WorldClim environmental layer for PC1-PC3. 59% of the variance is explained by PC1, 31% by PC2, and 4% by PC3. The colors correspond to the hues assigned to each PC in Figure 4.2. The red hue indicates the top four variables of PC1, largely driven by precipitation seasonality, mean diurnal range, warmest month maximum temperature and temperature seasonality variables (BIO15, BIO2, BIO5, BIO4). The green hue indicates the top four variables of PC2, driven by temperature annual range, wet month precipitation, and dry month and quarter precipitation (BIO7, BIO13, BIO14, BIO17). The blue hue indicates the top four variables of PC3, which was driven by precipitation variables (BIO12, BIO13, BIO16, BIO18).

Loadings:	PC1	PC2	PC3
BIO1 = Annual Mean Temperature	0.283	0.120	
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp)	**0.295**		
BIO3 = Isothermality (BIO2/BIO7) (×100)	0.244	0.232	
BIO4 = Temperature Seasonality (standard deviation ×100)	**-0.276**	0.136	
BIO5 = Max Temperature of Warmest Month	**-0.256**	0.202	
BIO6 = Min Temperature of Coldest Month	-0.246	-0.203	-0.121
BIO7 = Temperature Annual Range (BIO5-BIO6)		**0.377**	-0.106
BIO8 = Mean Temperature of Wettest Quarter	-0.252	0.212	
BIO9 = Mean Temperature of Driest Quarter	-0.239	-0.224	-0.142
BIO10 = Mean Temperature of Warmest Quarter	-0.212	-0.247	
BIO11 = Mean Temperature of Coldest Quarter	-0.256	0.200	
BIO12 = Annual Precipitation	0.202	-0.160	**-0.714**
BIO13 = Precipitation of Wettest Month		**0.343**	**-0.513**
BIO14 = Precipitation of Driest Month	0.155	**-0.339**	0.101
BIO15 = Precipitation Seasonality (Coefficient of Variation)	**0.296**		
BIO16 = Precipitation of Wettest Quarter	0.219	0.270	**0.122**
BIO17 = Precipitation of Driest Quarter	0.178	**-0.318**	-0.151
BIO18 = Precipitation of Warmest Quarter	0.237	0.125	**0.206**
BIO19 = Precipitation of Coldest Quarter	0.207	0.211	-0.277
Total Variance Explained	*59%*	*31%*	*4%*

Table 4.3. The Generalized Dissimilarity Model (GDM) incorporating all predictors explained 32.94% of the deviance in genetic turnover. The relative contribution of predictors to genetic dissimilarity across *L. parryae* is indicated by the relative coefficient values. PC1 has the highest coefficient, indicating that isolation by environment was the biggest contributor to genetic dissimilarity across the range of this species. Geographic distance also predicted genetic dissimilarity, while the other environmental PCs have no association with genetic dissimilarity.

Predictor	Coefficient
Geographic	0.51
Environmental PC1	0.55
Environmental PC2	0.00
Environmental PC3	0.00
% deviance explained by full model	32.94

Appendix

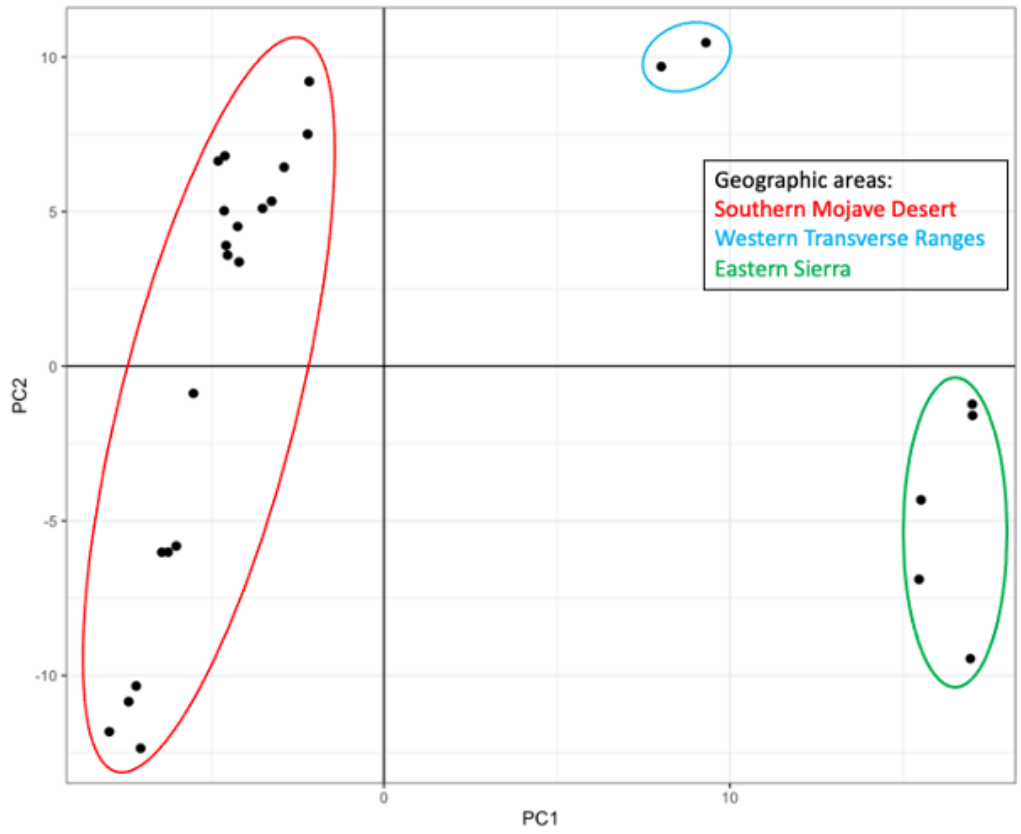


Figure S4.1. Principal component analysis (PCA) of the linkage disequilibrium pruned SNP data set of *L. parryae*. PC1 explained 8% of the genetic variation and separated individuals along latitude, with three genetically similar groups of individuals: southern Mojave Desert, western Transverse Range, and Eastern Sierra. PC2 explained 6% of the variation and separated individuals in the southern Mojave Desert.

References

Anghel, I. G., Jacobs, S. J., Escalona, M., Marimuthu, M. P. A., Fairbairn, C. W., Beraut, E., Nguyen, O., Toffelmier, E., Shaffer, H. B., & Zapata, F. (2022). Reference genome of the color polymorphic desert annual plant sandblossoms, *Linanthus parryae*. Journal of Heredity, 113, 712–721.

Axelrod, D. I. (1972). Edaphic Aridity as a Factor in Angiosperm Evolution. The American Naturalist, 106(949), 311–320.

Bell, K. C., Hafner, D. J., Leitner, P., & Matocq, M. D. (2010). Phylogeography of the ground squirrel subgenus *Xerospermophilus* and assembly of the Mojave Desert biota. Journal of Biogeography, 37(2), 363–378. https://doi.org/10.1111/j.1365-2699.2009.02202.x

Baughman, O. W., Agneray, A. C., Forister, M. L., Kilkenny, F. F., Espeland, E. K., Fiegener, R., Horning, M. E., Johnson, R. C., Kaye, T. N., Ott, J., St. Clair, J. B., & Leger, E. A. (2019). Strong patterns of intraspecific variation and local adaptation in Great Basin plants revealed through a review of 75 years of experiments. Ecology and Evolution 9(11): 6259–6275. https://doi.org/10.1002/ece3.5200

Caye K., Deist T.M., Martins H., Michel O., François O. (2016) TESS3: fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources* 16(2): 540-548. DOI: 10.1111/1755-0998.12471.

Chambers, E.A., Bishop, A.P., & Wang, I.J. (2023). Individual-based landscape genomics for conservation: An analysis pipeline. *Molecular Ecology Resources*.

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. (2021). Twelve years of SAMtools and BCFtools. Giga- Science 10(2): giab008.

D'Odorico, P., and A. Bhattachan. (2012). Hydrologic variability in dryland regions: Impacts on ecosystem dynamics and food security. Philosophical Transactions of the Royal Society B: Biological Sciences 367(1606): 3145–3157. https://doi.org/10.1098/rstb.2012.0016

Durbin, T. J., and W. F. Hardt. (1974). Hydrologic analysis of the Mojave River, California, using a mathematical model (No. 74-17). US Geological Survey

Epling, C. and T. Dobzhansky. (1942) Genetics of natural populations. VI. Microgeographic races of *Linanthus parryae.* Genetics 27(3): 317–332.

Epling, C., Lewis, H., & Ball, F. M. (1960). The Breeding Group and Seed Storage: A Study in Population Dynamics. Evolution 14(2), 238. https://doi.org/10.2307/2405830

Ferrier S., Manion G., Elith J., Richardson K. (2007) Using generalized dissimilarity modeling to analyze and predict patterns of beta diversity in regional biodiversity assessment.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology, 37(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modeling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, 18(1), 1–16.

Fitzpatrick M., Mokany K., Manion G., Nieto-Lugilde D., Ferrier S. (2022). gdm: Generalized Dissimilarity Modeling. R package version 1.5.0-9.1.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. Methods in Ecology and Evolution, 6(8), 925–929.

Gaddis, K. D., Thompson, P. G., & Sork, V. L. (2016). Dry-washes determine gene flow and genetic diversity in a common desert shrub. *Landscape Ecology*, *31*(10), 2215–2229. https://doi.org/10.1007/s10980-016-0393-7

Graham, M. R., Jaeger, J. R., Prendini, L., & Riddle, B. R. (2013). Phylogeography of the Arizona hairy scorpion (*Hadrurus arizonensis*) supports a model of biotic assembly in the Mojave Desert and adds a new Pleistocene refugium. Journal of Biogeography, 40(7), 1298–1312. https://doi.org/10.1111/jbi.12079

Hedrick, P. W. (2006). Genetic polymorphism in heterogeneous environments: The age of genomics. Annual Review of Ecology, Evolution, and Systematics, 37, 67–93.

Ishida, Y. (2017). Sewall Wright, shifting balance theory, and the hardening of the modern synthesis. Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences, 61, 1–10.

Jombart T.A.I. (2011). "adegenet 1.3-1: new tools for the analysis of genome-wide SNP data." Bioinformatics. doi:10.1093/bioinformatics/btr521.

Knaus, B. J., and N. J. Grunwald. (2017). VCFR: a package to manipulate and visualize variant call format data in R. Molecular Ecology Resources 17(1):44-53. http://dx.doi.org/10.1111/1755-0998.12549.

Li, H., and R. Durbin. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England), 25(14), 1754–1760.

Leutner, B., Horning, N., Schwald-Willmann, J., & Hijmans, R. J. (2019). RStoolbox: tools for remote sensing data analysis (Version 0.2.6). Retrieved from https://cran.rproject.org/web/packa ges/RStoolbox/index.html

Levin, D. A. (1981). Dispersal Versus Gene Flow in Plants. *Annals of the Missouri Botanical Garden*, 68(2), 233–253.

Masel, Joanna. (2011) Genetic drift. *Current Biology* 21(20): R837-R838.

McKenna, A., Hanna, M., Banks E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and M.A. DePristo. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res, 20:1297-303.

Mirchandani, C. D., Shultz, A. J., Thomas, G. W. C., Smith, S. J., Baylis, M., Arnold, B., Corbett-Detig, R., Enbody, E., and T. B. Sackton (2024). A Fast, Reproducible, High-Throughput Variant Calling Workflow for Population Genomics. Molecular Biology and Evolution, 41(1), 1–15.

O'Donnell, M. S., & Ignizio, D. A. (2012). Bioclimatic Predictors for Supporting Ecological Applications in the Conterminous United States. In *U.S Geological Survey Data Series 691*.

Pebesma, E. 2018. Simple features for R: Standardized support for spatial vector data. The R journal 10: 439.

Rius, M., & Darling, J. A. (2014). How important is intraspecific genetic admixture to the success of colonising populations? Trends in Ecology and Evolution, 29(4), 233–242.

Sala, O. E. and W. K. Lauenroth (1982). Small Rainfall Events: An Ecological Role in Semiarid Regions. Oecologia 53: 301–304.

Salguero-Gómez, R., Siewert, W., Casper, B. B., and K. Tielbörger (2012). A demographic approach to study effects of climate change in desert plants. Philosophical Transactions of the Royal Society B: Biological Sciences 367(1606): 3100–3114. https://doi.org/10.1098/rstb.2012.0074

Schemske, D.W., Bierzychudek, P. (2001) Perspective: Evolution of flower color in the desert annual Linanthus parryae: Wright revisited. Evolution 55(7):1269–1282.

Schemske, D.W., Bierzychudek, P. (2007) Spatial differentiation for flower color in the desert annual Linanthus parryae: Was Wright right? Evolution 61(11):2528–2543.

Sexton, J.P., Hangartner, S.B. and A. Hoffmann (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution, 68, 1–15.

Shaffer, H. B., Toffelmier, E., Corbett-Detig, R. B., Escalona, M., Erickson, B., Fiedler, P., Gold, M., Harrigan, R. J., Hodges, S., Luckau, T. K., Miller, C., Oliveira, D. R., Shaffer, K. E., Shapiro, B., Sork, V. L., & Wang, I. J. (2022). Landscape Genomics to Enable Conservation Actions: The California Conservation Genomics Project. Journal of Heredity, 113(6): 577–588.

Shirk, A. J., Landguth, E. L., & Cushman, S. A. (2017). A comparison of individual-based genetic distance metrics for landscape genetics. Molecular Ecology Resources, 17(6), 1308–1317.

Shryock, D. F., Washburn, L. K., DeFalco, L. A., & Esque, T. C. (2021). Harnessing landscape genomics to identify future climate resilient genotypes in a desert annual. Molecular Ecology, 30(3), 698–717. https://doi.org/10.1111/mec.15672

Slatkin M. (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, 47, 264–279.

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., & P. Prins. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics.

Turelli, M., Schemske, D. W., & Bierzychudek, P. (2001). Stable two-allele polymorphisms maintained by fluctuating fitnesses and seed banks: Protecting the blues in *Linanthus parryae*. Evolution 55(7): 1283–1298.

Vandergast, A. G., Inman, R. D., Barr, K. R., Nussear, K. E., Esque, T. C., Hathaway, S. A., Wood, D. A., Medica, P. A., Breinholt, J. W., Stephen, C. L., Gottscho, A. D., Marks, S. B., Bryan Jennings, W., & Fisher, R. N. (2013). Evolutionary hotspots in the Mojave Desert. Diversity, 5(2), 293–319.

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. Molecular Ecology, 23(23), 5649–5662.

Wang, I. J., Glor, R. E., & J. B Losos. (2013). Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, *16*(2), 175–182.

Wang, I. J., & K. Summers. (2010). Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology*, 19(3), 447–458.

Wright, S. (1943a). Isolation by distance. Genetics 28(2):114–138.

Wright, S. (1943b). An Analysis of Local Variability of Flower Color in *Linanthus parryae*. Genetics, 28(2): 139–156.

Wright, S. (1978). Evolution and the genetics of populations. VI. Variability within and among natural populations. Univ. of Chicago Press, Chicago.