UNCOVERING GENOMIC INSIGHTS TO THE DYNAMICS AND MECHANISMS OF SPECIATION USING *LIALIS BURTONIS* AND *HETERONOTIA BINOEI*

by

JAMES EVERETT TITUS-MCQUILLAN

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at

The University of Texas at Arlington May, 2019

Arlington, Texas

Supervising Committee:

Matthew K. Fujita, Supervising Professor

Todd A. Castoe

Jeff P. Demuth

Eric N. Smith

Matthew R. Walsh

Abstract

Uncovering Genomic Insights to the Dynamics and Mechanisms of Speciation using *Lialis burtonis* and *Heteronotia binoei*

James Everett Titus-McQuillan, Ph.D. in Quantitative Biology

The University of Texas at Arlington

Supervising Professor: Matthew K. Fujita

Species are fundamental entities in biology because their existence represents the cohesion that binds populations into a single unit. One facet of evolutionary biology is to identify the processes that cause the origin of species and maintain this cohesion apart from closely-related lineages. Two aims that speciation research attempts to answer are: (1) How do lineages diverge to form new species? and, (2) What prevents the merging of nascent species? One useful approach to address both questions is to study areas of contact where independent populations coexist and exchange genetic material. By investigating the size of contact zones, the extent of genetic exchange between lineages (gene flow), and the identity of introgressed genes; we can understand why and how species remain distinct despite ongoing gene flow between closely-related species. My dissertation focuses on uncovering the dynamics that gene flow has on potentially early and late stage speciation events. To address these goals, I employed next-generation sequencing to quantify species boundaries

and the extent of gene flow in two widespread Australian gecko lizard systems, *Lialis burtonis* and *Heteronotia*.

I found that *L. burtonis* has four distinct populations across Australia, with limited gene flow between groups, even when in close proximity. Each population appears to associate with an ecoregion in Australia. Currently there are four recognized lineages within the species: a population in the interior of Australia's arid zone, with a population in the Pilbara region (known for its high rates of endemism in native flora and fauna), the northern monsoonal tropic population, and a population occupying the eastern mesic zone bordering the arid zone. Migration is high within populations but not between populations. Findings show that migration between populations only happens where distributions overlap or are adjacent to one another. Migration within populations is likely because *L. burtonis* is an active and highly mobile predator.

Heteronotia binoei is hyperdiverse across its range, with over 60 mitochondrial populations discovered and more found every field season. *Heteronotia binoei* is constrained to outcrops, does not venture far from the established habitat of its home range. Broadly, this species colonizes and disperses across Australia. However, as lineages expand, isolation begins among distinct populations, allowing for lineages to diverge independently, erecting genetically distinct demes. Comparing 16 lineages across three contact zones at varying degrees of divergence, I found that more closely-related lineages, in overlapping range, exhibit greater gene flow compared to more divergent lineages. Secondary contact between divergent lineages occurs when expanding ranges between varying niche overlap. In this study, these lineages are not sister and have low or no gene flow when there is co-occurrence. These lineages do not encounter any gene flow between them and continue to diversify

independently. In this scenario, it is most probable that a speciation event happened, or late-stage speciation is currently occurring. *Heteronotia binoei* lineages that are sister taxa have continued gene flow or ancestral genotypes still within the population. Isolation by distance through parapatric speciation is the most probable cause for relatively recent lineage splits, where individuals close in geography have higher gene flow than those on the periphery of the range. To test whether there is habitat partitioning driving isolation, niche models were constructed using vegetation, radiation, temperature, precipitation, topography, and elevation variables. All models had no strong support for niche separation, confirming that *H. binoei* is a generalist gecko occupying most areas of Australia, where temperatures do not go below freezing.

Australia's climatic and geologic history, after splitting from Gondwana, has influenced the evolutionary history of its flora and fauna. Both *Lialis* and *Heteronotia* have their origins diversifying in isolation throughout the changing landscape and habitat of Australia. These two genera, however, are quite contrasting from each other. *Lialis burtonis* only has four distinct lineages, while *Heteronotia binoei* is hyperdiverse, with new lineages continually being discovered. Very little gene flow and high population structure and fixation indexes in *Lialis* indicate independent populations. Within *Lialis* population fixation indexes are low. By contrast, *Heteronotia* has many lineages at varying degrees of divergence. Many lineages in contact allow the migration of alleles between lineages, suggesting potential early speciation events. More deeply diverged lineages in secondary contact do not show allele migration or gene flow at overlapping ranges. These lineages experience the same biogeographic conditions but different evolutionary histories. Varying life histories between the species is presumably the root cause for the incredibly different demographies. *Lialis* is, comparatively, a large gecko, legless, and an active predator of secondary consumers (small

squamates). *Heteronotia* does not cover relatively large distances to forage; it is confined to small home ranges as an insectivorous secondary consumer.

Acknowledgements

I would have not completed this dissertation without the help and support of many. First, I wish to acknowledge my funding sources and thank the reviewers for having faith in my abilities to bring my ideas and proposals to fruition. The National Science Foundation provided an incredible experience through their EAPSI program, allowing me to conduct a majority of my *Heteronotia* work abroad, at the Australia National University. The American Philosophical Society for providing me with The Lewis and Clark Fund for Exploration and Field Research, which allowed me to gather needed samples in Australia and to finish my dissertation. I also thank the Phi Sigma Honor Society for providing the initial funds for data gathering, as well as travel funds for conference exposure and guest speaker presentations. The Teaching as Research (TAR) grant allowed me to conduct research while mentoring undergraduate students. And I thank The Graduate Student Council for providing the Graduate Studies Travel Grant, allowing me to conduct field work in Puerto Rico.

It would be remised to not also acknowledge my past and present peers and friends for their expertise and support. This list is as exhaustive as possible - Adnan Qureshi, Alex Hall, Alex Murray, Andrew Corbin, Anya Williford, Arm Panugong, Audra Andrew, Ayda Mirsalehi, Blair Perry, Bob Grinshpon, Brad Dimos, Catherine Rogers, Claudia Marquez, Corey Roelke, Dan Portik, Danielle Rivera, Daren Card, David Sanchez, Diwash Jangam, Drew Schield, Eli Wostl, Eric Watson, Gulia Pasquesi, Heath Blackmon, Jacobo Reyes-Velasco, Jesse Meik, Jill Castoe, Joe Maciag, Joe Mruzek, Jose Maldonado, Justin Jacobs, Kathleen Currie, Kris Row, Kyle O'Connell, Kyle Shaney, Lauren Fuess, Matt Mosely, Mehdi Eslamieh, Nick Long, Rachel Card, Rachel Wostl, Rich Adams, Shannon Beston, Stephen Zozaya, Suman Shrestha, Susana Domigues, Thor Larson, TJ Firneno, Trung Nguyen, Tuli Brunes, Utpal Smart, Walter Schargel, and Will Budnick. I want to

acknowledge my family and friends for their support during hard times, especially my best friend Greg Beilstein, mother Kate Titus, and Eric Van Cleave. Without them this journey would have been much tougher.

I would also like to thank my committee Dr. Todd Castoe, Dr. Jeff Demuth, Dr. Eric Smith, and Dr. Matthew Walsh for their help and insights into science. Finally, I want to give a special thank you to my adviser, Dr. Matthew Fujita. Who with much patience helped me with scientific writing, showed me the ins and outs on finishing projects, and gave me valuable resources and contacts.

Without these people, I would not be where I am today. So, thank you all from the bottom of my heart!

Preface

This dissertation is a long journey through my life where I poured my interests into herpetology. When I was around three years old, I went on an off-roading trip with my father, his best friend, and his daughter in Orange County California. I remember the jeep was like a roller coaster, traversing the rocky streams at the base of Saddle Back Mountain. We stopped for lunch; however, something incredible peaked my interests. In the stream, there was a long line of black-little fish. I was told that these round ovals with flailing tails would become frogs. My mind could not understand how something in the water would become a terrestrial creature that looked so very different. I would later find that these black-little fish would become the Western toad (*Anaxyrus boreas*). From that day forward I was hooked on amphibians, later reptiles, and evolution, albeit primitive at the time. As I grew up, I always had reptiles and amphibians in the house, usually rearing frogs from tadpoles. One of my favorite animals I kept were African clawed frogs (*Xenopus laevis*), which I had for 12 years. Even in first grade, I had a love for herpetology. The class made a town out of shoeboxes where each student crafted their shoebox to fit the occupation they wanted to become. I was the resident herpetologist. Surprisingly, I had no competition. From here my future endeavors throughout school were to become a herpetologist.

When deciding on where I would go to college, I looked for a program that had herpetologists as faculty along with a zoology major. After perusing through career resources of what it would take to be a research biologist, I knew I had to achieve a master's degree, at least, if not a Ph.D. So, I found myself at Cal Poly Pomona, not only because it matched all my criteria, but also because it had my Plan B major in computer science, and Plan C major

in architecture (in case working with animals did not work out). Plan A is where I continued my passions. I continued my studies, after my undergraduate degree, specializing with geckos under Dr. Aaron Bauer and Dr. Todd Jackman, and specializing in phylogenetics and systematics. This journey later brought me in contact with Dr. Matthew Fujita, at the First World Congress of Evolution in Ottawa, during the summer of 2012. I knew of Dr. Fujita from his paper elucidating the molecular systematics of North African gekkonids of the genus *Stenodactylus*, a congener of the taxon of my master's work. When I met Dr. Fujita, he had just accepted a job at the University of Texas at Arlington and would be a faculty member after the summer. I followed him to UTA.

Throughout my academic career, I have learned a substantial amount about both science and myself. My undergraduate study invigorated my passion for working with animals, specifically herps. It also allowed me to understand statistics and their value. My masters showed to me how real scientists work and how to conduct yourself within the field, along with giving me a taste of publishing. My Ph.D. enabled me to dig into computational biology, linking two passions of mine – biology and computers, and taught me the importance of writing. Oddly enough, I learned more about language and writing in the field of biology than in any English/Literature class I have ever taken.

This Ph.D. is the culmination of everything I have learned in 30 years. I will finally finish school in the 24th grade ready to become a full professional researcher.

Dedication

I dedicate this work to my mother. As a parent she allowed me to pursue my passions. Without her parenting style that allowed me the freedom to indulge in what I love; I would not be here today. I thank her for being a wonderful mother and mentor throughout life.

Table of Contents

List of Figures

Figure

List of Tables

Table

Chapter 1

Introduction: Paleoclimate Fluctuation Driving Australia's Diversity and Study Systems [*Lialis burtonis* & *Heteronotia binoei*]

A primary aim of evolutionary biology is to identify the processes that promote and maintain biodiversity (Coyne & Orr, 2004). Speciation research tends to focus on two significant questions: How do lineages diverge to form new species, and what prevents nascent species from merging back together? My dissertation aims to understand how genetics and environment can promote biodiversity. My dissertation's focus is on how the changing climate of Australia shaped the evolution of two gecko systems. These two distinct systems are *Lialis burtonis* and *Heteronotia binoei*.

Environmental change, especially climate fluctuation, significantly drives biotic diversification through expanding distributions during interglacial periods (Levins, 1968; Avise, 2000; Ackerly, 2003). The Southern Hemisphere experienced periods of arid expansions during interglacial periods from the mid-Miocene to the Pleistocene (Martin, 2006; Byrne *et al.*, 2008). Much of the biodiversity in Australia arose as a result of the aridification cycles in the Pliocene (Clapperton, 1990; Markgraf *et al.*, 1995). These aridification events also caused extinctions (Cogger & Heatwole, 1981; Crisp *et al.*, 2009), opening large new adaptive zones for lineages in refugia to expand and colonize previously occupied niches. The two most massive biomes that currently persist in Australia are a large arid zone within the interior of the continent and a tropical monsoon biome along Australia's north and northeastern coast. Paleoclimate fluctuations have resulted in both extensions and

retreats of these two biomes and likely influenced the demographic histories of populations, including population expansions and gene flow dynamics, which is observed in Australian reptiles (*e.g.*, *Heteronotia binoei*; Fujita, *et al.*, 2010) and nearly all flora and fauna across the continent (Byrne *et al.*, 2008, Brennan & Oliver, 2017; Brennan & Keoh, 2018). These periods of wet and dry phases also created land bridges between Papua New Guinea (PNG) and Australia, multiple times, when water was sequestered as glacial ice during the last glacial maxima (LGM) (Byrne *et al.*, 2008).

Chapter 2 aims to elucidate the contemporary and temporal processes influencing the biogeography of *Lialis burtonis* within Australia and between Australia and Papua New Guinea.

Burton's legless gecko, *Lialis burtonis*, is an excellent model system demonstrating regional differentiation with a continent-wide distribution. Covering all biomes throughout Australia and portions of Papua New Guinea (PNG), this species provides insights of biogeography. *Lialis burtonis* inhabits both tropical and arid biomes, regions that have led to distinct population dynamics (Fujita *et al.*, 2010; Pepper *et al.*, 2013). Land bridges allowed Gondwanan constituents to colonize PNG from the Australian continent. The aridification of the southern hemisphere, especially in Australia, is a driving force for a broad range of diversification in most flora and faunal groups. To investigate the processes and patterns of diversification, I choose to focus on the aridification of Australia.

With *L. burtonis*, I elucidate varied biogeography by incorporating all major biomes in Australia, while also incorporating dynamics of multiple land bridge connections and seaway barriers between Australia and PNG. To understand the processes influencing diversification I aim to observe how differentiating climate affects the evolutionary history of

populations, estimate the biogeographic processes affecting gene flow dynamics, and reveal cryptic haplotypes, and possibly unrecognized species. First, I hypothesize that geographic barriers have led to population separation dictated by paleoclimate, with limited gene flow between each population. Second, I hypothesize that cryptic diversity will be low within the species.

Chapter 3 aims to identify the genomic underpinnings of speciation in the widespread and hyperdiverse Bynoe's gecko (*Heteronotia binoei*) by quantifying the permeability of thousands of genes through three hybrid zones and between lineages of varying degrees of divergence.

The *Heteronotia binoei* (Bynoe's gecko) complex is an excellent model to study hybrid zones because there are several putative species whose distributions overlap at multiple locations across Australia. *Heteronotia* is one of the most abundant lizards in Australia and inhabits a wide variety of habitats, from the harsh arid interior to tropical monsoonal forests. Previous work by Matthew Fujita and Craig Moritz found extensive and cryptic diversity likely representing multiple speciation events influenced by the climatic cycles of the Pliocene and Pleistocene (Fujita *et al.*, 2010; observed in other organisms [Byrne *et al.*, 2008; Fujita *et al.*, 2010; Pepper *et al.*, 2006, 2013; Potter *et al.*, 2012, Eldridge *et al.*, 2014]). Further discoveries of hyperdiversity and local endemism in the monsoonal tropics of Australia continue to demonstrate the utility of *Heteronotia* as a model of recent biotic diversification (Moritz *et al.*, 2016; Rosauer *et al.*, 2016).

My project leverages this diversity of *Heteronotia* to investigate the processes of speciation operating at different timescales, by using three distinct hybrid zones between lineages of varying divergence. This project uses these hybrid zones to address two

hypotheses. First, I hypothesize that genomic admixture will be most significant between the least-differentiated lineages and lowest between the most differentiated lineages. Genes that do not cross the hybrid zone may be necessary for lineage divergence (candidate loci under selection). Second, I hypothesize that each hybrid zone will exhibit a distinct niche space that fails to show admixture. Because of local adaptation, I predict that each lineage will exhibit distinct niche space which mitigates gene flow and promotes diversification.

Chapter 2

Phylogeographic Patterns: *Lialis burtonis* During Australia's Aridification

Abstract

1. Population structure, and subsequent speciation events, across Australia, has been driven by climate through time. Glacial maxima and interglacial cycles promoted diverse species radiations in the southern hemisphere by creating aridification events during the LGM along with land connections to adjacent islands opening new niche space for expansion then constricting ranges during wetter times, subsequently restricting gene flow. These events have shaped the biodiversity of Australia. Identifying these processes leading the immigration and isolation are critical to understanding population dynamics leading to speciation and evolutionary history of a species.
2. I elucidated population structure, gene flow, and paleoclimate of Burton's legless lizard (*Lialis burtonis*) using genomic and geographical data. I estimated population structure, gene flow, and phylogeny using both nuclear (RADseq) and mitochondrial data using clustering, coalescence, and maximum likelihood methods. To understand the climatic dynamics from the past, influencing the evolutionary history of *L. burtonis*, I conducted niche analyses using paleoclimate data from the mid-Holocene, interglacial, and last glacial maximum.
3. I found four populations within *L. burtonis* across Australia and Papua New Guinea with limited gene flow between populations. The Eastern population spans across the northeastern portion of Australia and Papua New Guinea and likely crossed a land connection between the two landmasses during the LGM. Barriers to gene flow are

elevational across highland and lowland deserts. The Pilbara group is interestingly isolated to the Pilbara of Western Australia, an area of high endemism in many floral and faunal groups.

4. This project illustrates how changing climate and geographic barriers influence the processes of population structure and speciation within a species, and how genomic data in conjunction with integrative approaches are powerful enough to uncover intricate evolutionary patterns.

Introduction

Cycles of glaciation events have perpetuated recessions and expansions of diversity throughout time, where species retreat to refugia, undergo allopatry, and expand across environmental gradients as glaciers recede (Levins, 1968; Avise, 2000; Ackerly, 2003). This process occurred more so in the northern hemisphere during the Pliocene and Pleistocene, where glaciers acted as barriers between lineages confined to ancient refugia and expanded post-glaciation (Hewitt, 2004). The southern hemisphere mirrors, but also contrasts, the northern hemisphere during glacial periods. The southern hemisphere was not covered with significantly large ice sheets, and in turn, during glacial cycles, earth's water was confined to ice, creating periods of drying. This process of aridification can be tracked to the mid-Miocene and is observed throughout the southern hemisphere in South America (Ortiz-Jaureguizar & Cladera, 2006), South Africa (Richardson *et al.*, 2001; Cowling *et al.*, 2009), and Australia (Martin, 2006; Byrne *et al*., 2008). These aridification events are known to cause the recession of lineages into refugia from past wetter environments, but unlike sizeable glacial ice sheets, new large adaptive zones allowed for lineages in refugia to expand and colonize novel arid niche space (Cogger & Heatwole, 1981; Crisp *et al.*, 2009).

A substantial body of research has been given to the evolution of northern hemisphere taxa throughout the molecular age of evolutionary biology (Hewitt, 2000; Schafer *et al.*, 2010). However, new research over the last decade continues to uncover the underpinnings that aridification has had on evolutionary patterns across southern hemisphere taxa during interglacial cycles (Byrne *et al.*, 2008). Data gaps persist in the role aridification has on driving biodiversity in the southern hemisphere (Chambers *et al.*, 2017); however, Australia has a considerable amount of research addressing how aridification drives diversification through temporal and geological processes (Martin, 2006; Byrne *et al.*, 2008; McLaren & Wallace, 2010). This study is no exception in its role, adding to the study of aridification driving diversity.

Much of the contemporary biodiversity in Australia arose because of aridification cycles in the Pliocene (Clapperton, 1990; Markgraf *et al.*, 1995). The two most massive biomes that currently persist in Australia are a sizeable arid zone within the interior of the continent and a tropical monsoon biome along Australia's north and northeastern coast. The Pilbara region is an intermediate region between arid and tropical environments and one the oldest areas on earth, with rock formations over two billion years old (Australian Natural Resources Atlas, 2011). Paleoclimate fluctuations have resulted in both extensions and retreats of these two biomes and likely influenced the demographic histories of populations, including population expansions and gene flow dynamics, which is observed in Australian reptiles (Fujita *et* al., 2010; Pepper *et al.*, 2013; Moritz *et al.*, 2016; Moritz *et al.*, 2017) and a broad swath of floral and faunal lineages across the continent (Byrne *et al.*, 2008, Brennan & Oliver, 2017; Brennan & Keogh, 2018). These periods of wet and dry phases created land bridges between Papua New Guinea (PNG) and Australia three times—around 17k, 150k, and 250k years ago (Voris, 2000). Sea levels dropped due to the last glacial maxima (LGM)

(Voris, 2000; Byrne *et al.*, 2008). These land bridges allowed Gondwanan constituents to colonize PNG from the Australian continent. The aridification of the southern hemisphere, especially in Australia, is a driving force for a broad range of diversification in many flora and faunal groups. One such system that rapidly radiated and diversified across new niche spaces, driven by aridification, is the Australian herpetofauna.

This project aims to elucidate how biogeographic processes drive diversification. This study takes a novel perspective by using genomic scale data to explain the variation and biogeographic mechanisms promoting or restricting diversification. Through climate fluctuations that shrunk previous biomes into refugia and expanded novel biomes (arid), I aim to delineate how the varied climate and geography influence the contemporary fauna of Australia, focusing on *Lialis burtonis*. The continual expansion and contraction of environments into refugia by glacial cycles in the Pliocene and Pleistocene most likely played a part in the expansion and diversification of *Lialis burtonis*. As arid climates formed, expansion of *Lialis* into new niche space allowed for diversification of *Lialis burtonis*. During glacial cycles, the connection between Papua New Guinea and the Cape York Peninsula allowed access between the land masses, a unique happenstance within its family Pygopodidae. As a mobile, active predator, migration is a plausible mechanism for the continued gene flow between lineages of *L. burtonis*. This study incorporates RADseq data using next-generation sequencing to examine the dynamics of biogeography in *Lialis burtonis*. **My goals are to (1) elucidate the role aridification and land connection to Papua New Guinea had on *Lialis burtonis* diversification and (2) determine the mechanisms and processes that promote or limit genetic diversification.** *Lialis burtonis* is an excellent system to tackle these questions since the species is widespread across Australia and Papua New Guinea with dynamic boundaries between population groups influenced by

periods of glacial expansion and recession. Predominantly *L. burtonis* is an Australian species of pygopods; however, it does continue its range onto the island of New Guinea. I estimated population relationships to observe how geological and climatic events have influenced the evolution of this group.

Methods

Sampling

Samples of *Lialis burtonis* were obtained from the South Australia Museum (SAM), the Western Australia Museum (WAM), Louisiana State University Museum of Natural Science (LSUMNS), and Marquette University. A total of 81 individuals were obtained covering the distribution of *Lialis burtonis*, including Papua New Guinea (Table 1, Fig. 1).

Table 1. Point localities for all samples in this study in decimal degrees.

Catelogue #	Locality	State	Latitude	Longitude
ABTC3721	11kWNarrabri	NSW	-30.324835	149.782833
ABTC49515	Wipim	WP	-8.793	142.868
ABTC6485	SilverPlains	Qld	-13.983333	143.55
ABTC70388	1_2kSWWilpenaChalet	SA	-31.539	138.591
ABTC72801	BangBangJumpup On Burke Development Road Wandoolaturnoff	Qld	-18.525	140.658
ABTC72844	NormantonFloravilleRoadArmstrongCreekcrossing	Qld	-17.925	140.700
ABTC72863	BetweenKarumba&WalkersCreek	Qld		
ABTC77188	13kWMtSurpriseonGulfDevelopmentalRoad	Qld	-18.191	144.199
ABTC79176	Wegamu_Trans-Fly	PNG	8.43287	141.112
CCA3863	Brown River, North West of Port Moresby, 30-36 km from roundabout to airport	PNG	-9.1659667	147.1796167
CCA3901	Sogeri Road, ~ 5 km East of Port Moresby	PNG	-9.4124333	147.2343833
CCA5629	Amau Village, on the Amau River (near Kupiano)	PNG	-10.0368	148.5646333
ABTC06491	LiverpoolRiver	NT	-12.03	134.21
ABTC18024	BingBongHS	NT	-15.623	136.353
ABTC28071	MelvilleIs	NT	-11.582451	131.120003
ABTC28072	VictoriaRiver_GregoryNP	NT	-16.285251	130.689516
ABTC28237	McMinnsLagoon	NT	-12.540	131.060

ABTC28477	20kSWKatherine	NT	-14.604	132.143
ABTC28478	20kWKatherine	NT	-14.480	132.075
ABTC29298	GuluwuruIsland	NT	-11.521	136.429
ABTC29346	CapeCrawfordarea	NT	-16.680	135.720
ABTC29386	16kNWauchope	NT	-20.511	134.261
ABTC29662	LitchfieldNP	NT	-13.147	130.780
ABTC29916	EnglishCompanyIsles_PobassoIsland	NT	-11.902	136.454
ABTC29938	EnglishCompanyIsles_AstellIsland	NT	-11.877	136.422
ABTC30297	61kNThreeWays	NT	-18.893	134.088
ABTC30417	TawallahCreekLimmenGateNP	NT	-16.049	135.647
ABTC67994	BradshawStation	NT	-15.348	130.279
ABTC82421	PhosphateHill_CreekSite	Qld	-21.824776	139.964851
ABTC82436	TheMonument	Qld	-21.763	139.917
ABTC102707	15kSMountIsaonBouliaRoad	Qld	-20.876	139.454
ABTC102713	MicaCreek_powerstation_SMountIsa	Qld	-20.7792	139.4914
R139067	MANDORA	WA	-19.798	121.4478
R146044	KIMBOLTON	WA	-16.743	124.095
R151829	SIRGRAHAMMOOREISLAND	WA	-13.883	126.5667
ABTC68796	TriangleHill8kNEBimbowrieHS	SA	-31.979	140.217
R102424	BARLEERANGENATURERESER	WA	-23.096	116.0097
R117118	TUREECREEKHS	WA	-23.817	118.5667
R120834	PERON	WA	-25.875	113.5503
R123735	BOOLANPOOL	WA	-24.413	113.7631
R139138	MEENTHEENA	WA	-21.420	120.4258
R145508	PORTHEDLAND	WA	-22.500	119.09
R145685	ABYDOS	WA	-21.358	118.8831
R145904	SALUTATIONISLAND	WA	-26.535	113.7608
R157150	ROYHILL	WA	-22.770	120.5067
R157274	YANREY	WA	-22.154	114.5508
R161046	MARDAPOOL	WA	-21.063	116.234
R162792	TOMPRICE	WA	-22.835	117.382
R165296	CAPEBURNEY	WA	-28.852	114.64
ABTC06583	Wirrulla	SA	-32.4	134.53
ABTC12043	MtZeil	NT	-23.417	132.417
ABTC24061	LawrenceGorge	NT	-23.983	133.441
ABTC30915	CurtinSprings	NT	-25.3	131.75
ABTC34933	5_7kNEInvasionTank	SA	-26.051	140.881
ABTC35616	1_7kEStrangwaysSprings	SA	-29.161	136.566
ABTC35702	5_6kSWCallannaBore	SA	-29.641	137.889
ABTC36133	0_2kWAnvilHoleNativeWell	SA	-26.358	135.708
ABTC39401	8_5kNNEWillowSpringsHS	SA	-31.390	138.822
ABTC41583	10kNWMtKintore	SA	-26.497	130.412
ABTC41731	9_2kSSEAmpeinnaHills	SA	-27.150	131.143
ABTC42217	11_2kSWSentinelHill	SA	-26.141	132.359

ABTC42470	3_3kSWIndulkana	SA	-26.988	133.283
ABTC52028	OlympicDamRoxbyDowns	SA	-30.367	136.933
ABTC57020	MunyarooCP_8kSWMoonabieHS	SA	-33.294	137.208
ABTC57182	PooginookCP_6kNHighway64	SA	-34.083	140.133
ABTC57310	PeebingaCP	SA	-34.533081	138.747089
ABTC57448	12_2kNWMtCheesman	SA	-33.216667	137.116667
ABTC58064	StFrancisIsland	SA	-34.974	140.798
ABTC58538	Coonbah	NSW	-27.337	130.238
ABTC59720	95kSYuendumuRoad	NT	-32.983	141.617
ABTC59721	95kSYuendumuRoad	NT	-32.983	141.617
ABTC61782	95kSYuendumuRoad	NT	-23.641	132.355
ABTC61783	95kSYuendumuRoad	NT	-23.641	132.355
ABTC64149	5kSSWImmarnaSiding	SA	-30.547	132.138
ABTC64302	50kSWHalinorLake	SA	-29.525	130.150
ABTC68902	MonalenaRuins_OakdenHillsStation	SA	-31.715	137.178
ABTC70465	4_9kENETelowie	SA	-33.045	138.118
ABTC82562	MtGibsonStationsolanumtank	WA	-29.607522	117.409522
ABTC87339	44_9kWNWMaralinga	SA	-29.908	131.213
ABTC91518	3_2kNPungkulpirriWaterhole_WalterJamesRange	WA	-24.6286	128.7556
ABTC93835	1_3kSSWMtMisery	SA	-33.951	135.236
ABTC94340	184kSSWWatarru	SA	-28.50788	129.00473
ABTC94768	28_3kWSWAlyukurlpykurlpyTrig	SA	-29.594	135.374
ABTC95330	19_3kEVokesHillCorner	SA	-28.567	130.879
ABTC95758	14_3kENEKoongawa	SA	-33.123	136.0213889
ABTC101011	13_8kEFigTreeCorner	SA	-31.684	132.865
R122675	MEEDO	WA	-25.681	114.6217
R125696	QUEENVICTORIASPRING	WA	-30.428	123.5733
R125975	MARYMIA	WA	-25.067	119.85
R127542	BARNONGSTATION	WA	-28.633	116.2833
R132003	DRYANDRA	WA	-32.800	116.9167
R140548	DAWESVILLE	WA	-32.646	115.6336
R141725		WA	-27.869	120.1508
R151231	SALMONGUMS	WA	-32.788	121.42
R151472	LORNAGLENSTATION	WA	-26.075	121.4522
R154003	MUCHEA	WA	-31.642	115.9175
R154028	MUCHEA	WA	-31.638	115.9253
R157434	TANAMIDESERT	WA	-19.889	128.86
ABTC04195	CurtinSprings	NT	-25.183	131.683
ABTC14519	Manillatip	NSW	-30.749	150.710
ABTC18025	NathanRiver	NT	-15.578	135.429
ABTC18026	ElseyHS	NT	-14.960	133.322
ABTC21863	900mNGrindellHut	SA	-30.467	139.208
ABTC24002	70kEMoomba	SA	-28.117	140.867
ABTC24154	6kSSWClaravilleHS	NT	-23.431	134.726

ABTC39999	OakbankOutstation	SA	-33.128	140.606
ABTC51378	RiversleighStation	Qld	-19.031	138.739
ABTC51408	7kEMountIsa	Qld	-20.722	139.555
ABTC58591	Coonbah	NSW	-32.517	133.300
ABTC62019	WoodstockStation	WA	-21.302974	118.861156
ABTC76457	36kWChartersTowers	Qld	-20.183	145.967
ABTC81233	16kSBurra	SA	-33.818	138.926
ABTC102845	MountIsaarea	Qld	-20.720	139.480
ABTC102846	6kSMountIsaMinesturnoff	Qld	-20.777	139.481
ABTC102854	13_2kSMountIsaMinesturnoff	Qld	-20.8276	139.463
R163279	NEALEJUNCTION	WA	-28.303	126.2992
R164705	FITZROYCROSSING	WA	-18.681	125.8814
R171383	ADOLPHUSISLAND	WA	-15.111	128.1694

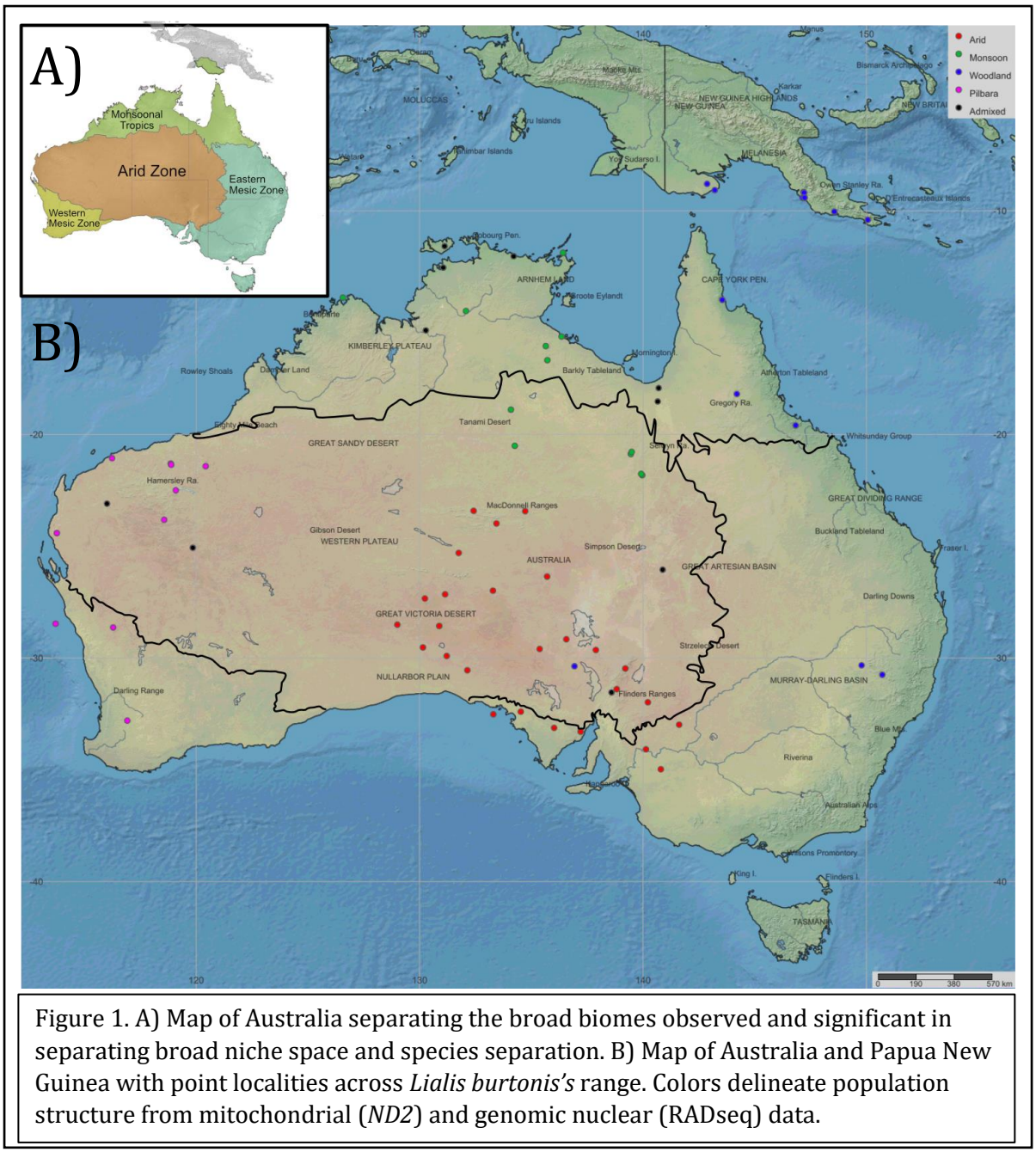


Figure 1. A) Map of Australia separating the broad biomes observed and significant in separating broad niche space and species separation. B) Map of Australia and Papua New Guinea with point localities across *Lialis burtonis's* range. Colors delineate population structure from mitochondrial (*ND2*) and genomic nuclear (RADseq) data.

Molecular

In order to investigate diversification in *L. burtonis*, thousands of single nucleotide polymorphisms (SNPs) were collected from individual nuclear genomes using restriction site associated DNA sequencing (RADseq) and Illumina technologies (Etter *et* al., 2011; Peterson *et* al., 2012), as well as mitochondrial gene NADH dehydrogenase subunit 2 (*ND2*).

Phenol-chloroform extractions were used to purify genomic and mitochondrial DNA from liver tissues (Walsh *et al.*, 1991). *ND2* was Sanger sequenced using primers mt30 and mt190 (Fujita *et al.*, 2010; Supplementary Table 1.). Genomic data were extracted using double digest RADseq protocols (Peterson *et al.*, 2012; Emerson *et al.*, 2010) to extract variable sites across individual genomes of *L. burtonis*. Unlike traditional Sanger sequencing methods that require species-specific primers, RADseq allows for the collection of thousands of homologous markers for hundreds of individuals with lower cost and effort (Peterson *et al.*, 2012; Emerson *et al.*, 2010). Restriction enzymes *Sbf*1 (rare cutter) and *Msp*1 (common cutter) were used to fragment the DNA. Single-end reads were then sequenced on an Illumina Hi-Seq 4000.

Bioinformatics

Raw Illumina reads were quality checked in FASTQC (Andrews, 2010) then imported into Stacks (Catchen *et al.*, 2011) for cleaning using 'Clone_filter,' 'process_short_reads,' and 'kmer_filter.' After the clean-up, the Stacks function 'process radtags' was used to demultiplex indexed libraries by barcode. iPyRAD (Eaton, 2018) was then used to filter reads by the quality score as a final check from Stacks. Assembly was performed as *De Novo* and clustered reads within sample for initial assembly. The joint estimation of heterozygosity and error rates per individual show that reads were clean and of high quality (Supplementary Table 2.). Consensus base calls were estimated through parameters set by sequence error rate estimation and heterozygosity from each cluster, where filtering is inferred from a binomial model and a maximum number of N bases. After the within individual clustering, I clustered among individuals. Finally, data went through a final filter set to drop cleaned filtered reads that did not have 90% coverage within genomic data (10% missing cut off) in iPyRAD. To

not drop SNPs by an excess amount of sequence data, I put in a parameter to drop individual taxa with long sequence gaps, also set to 90% complete data within the data set.

Population Structure

Discriminant Analysis Principal Component

I conducted a discriminant analysis of principal components (DAPC) to deduce population groupings from single nucleotide polymorphisms (SNPs) gathered from scrubbed RADseq data. A Principal component analysis (PCA) of genetic markers is instrumental in generating distinct population groupings since it makes no assumptions regarding Hardy Weinburg Equilibrium (HWE), linkage disequilibrium, or recombination. Since *a priori* information on mitochondrial races of *L. burtonis* is known, we ran a DAPC using the R package *adegenet* (Jombart, 2008; Jombart & Ahmed, 2011). Three discriminant analysis eigenvalues were kept along with 40 principal components (PCs) as to not induce high error in the analysis. Cross-validations show a high proportion of success for including 20-40 PCs.

Fixation Indexes

To quantify differentiation and, subsequently, independence among populations, I conducted an *Fst* test. Using the function 'populations' in stacks, I conducted *Fst*, *Φst*, and *Fis* to quantify fixation between populations and within populations (Table 2). Both *Fst* and *Φst* allow for estimating fixation indexes between populations. However, *Φst* is better equipped to handle heterozygosity. *Fis* was calculated to see fixation indexes within populations. *Fis* will enable observations of differentiation within populations instead of differentiation from one population concerning the next.

Table 2. Three matrices consisting of fixation statistics and one table with *Fis* statistics within each population.

Fst Summary				
	Arid	Monsoon	Woodland	Pilbara
Arid		0.118163	0.146961	0.0412291
Monsoon			0.122955	0.170864
Woodland				0.224994
Φst Means				
	Arid	Monsoon	Woodland	Pilbara
Arid		0.143035	0.176687	0.0290711
Monsoon			0.141583	0.175764
Woodland				0.251636
Fst Means				
	Arid	Monsoon	Woodland	Pilbara
Arid		0.104006	0.139494	0.0258153
Monsoon			0.101076	0.149073
Woodland				0.208413

Fis	
Pop ID	***Fis***
Arid	0.24585
Monsoon	0.15942
Woodland	0.1537
Pilbara	0.09738

Admixture

One of the primary aims of this project is to identify whether or not the distinct

populations of *Lialis* exchange genes or are genetically isolated. Hence, estimating introgression and gene flow between populations is essential to understand population dynamics, whether each group is evolving independently or admixing with adjacent populations. I used ADMIXTURE for our cluster analysis (Alexander *et al.*, 2009) using 100 bootstraps across *k*-values 2-10 to encompass all possible *k*'s across the sampled range. Our cross-validations generated by the analysis had the lowest error rate at $k = 4$, concordant with the mtDNA population groupings (Supplementary Fig. 1-6).

Haplotype Network and Phylogenetic Network

To estimate population relations across Australia and Papua New Guinea, I conducted a haplotype analysis on mitochondrial gene *ND*2 performed using SplitsTree5 v5.0.0_alpha (Huson, 1998, Huson & Bryant, 2006). The nexus input consisted of 80 taxa and the full protein coding portion of *ND*2 covering 1041 base pairs. In total 79 unique haplotypes were erected.

A super neighbor-network was conducted using RADseq SNP data. Phylogenetic networks are useful, unlike traditional phylogenetic methods that portray relationships as bifurcations, since they incorporate reticulation. Reticulated phylogenies help to estimate backcrosses between groups. I constructed a neighbor-network using SplitsTree5 v5.0.0_alpha (Huson, 1998; Huson & Bryant, 2006). The Hamming Distances method (Hamming, 1950) was used to obtain an 80x80 distance matrix. I used the median joining method (Bandelt *et al*., 1999), obtaining 85 nodes and 98 edges. The neighbor-net method (Bryant & Moulton, 2004) was used (parameters: CutOff = 1.0E-6, LeastSquares = ols, Regularization = nnls, LambdaFrac = 1.0) to obtain cyclic 374 splits. The splits network algorithm method (Dress & Huson, 2004) was used (parameters: Algorithm = EqualAngleConvexHull, UseWeights = true, BoxOpenIterations = 0, DaylightIterations = 0)

to obtain a splits network with 4719 nodes and 9062 edges.

Phylogenetic Analysis

First, a mitochondrial gene tree, using the protein-coding portion of *ND*2, partitioned by codon position, was constructed. I inferred the best-fit models for the partitioned codon positions of *ND*2 using the corrected Akaike information criterion (AICc) implemented in PartitionFinder2 v2.1.1 (Lanfear *et al.*, 2016; Akaike, 1998; Aho *et al.*, 2014). The Bayesian analysis was conducted using MrBayes v3.1.2 (Ronquist & Huelsenbeck, 2003). Each partition was given a mixed substitution model with invariant sites and a Γ parameter, with all parameters unlinked across all partitions. Analyses were initiated with random starting trees and run for 100,000,000 generations; Markov chains were sampled every 5,000 generations; the first 2000 trees, representing 10% of all trees, were discarded as burn-in.

A quartet tree was generated using TETRAD v.0.7.19 (Eaton, 2014) to construct an unrooted phylogeny using SNPs. TETRAD is open source and uses the same algorithm as SVDquartets (Clifman & Kubatko, 2014). A species tree was also run with the SNAPP (Bryant *et al.*, 2012) template of BEAST2.5.1 (Bouckaert *et al.*, 2014) using 39,967 SNPs across all *k* populations designations (4: Monsoon, Woodland, Arid, and Pilbara). Five individuals from each population were used that did not show any signs of introgression between other populations. The run was conducted with 10,000,000 MCMC generations logging trees every 1000 generations under a GTR Γ+ I substitution model.

Phylogeography

Effective Estimation of Migration Surfaces

Estimation of migration and diversity between populations of *L. burtonis* was conducted using effective estimation of migration surfaces (EEMS) (Petkova *et al.*, 2016). Effective Estimation of Migration Surfaces (EEMS) is useful for visualizing migration and

diversity along a Euclidean plane (here the geographic space of Australia and the island of New Guinea). Unlike traditional PCA and clustering approaches, EEMS explicitly represents genetic differentiation as a function of migration. EEMS uses a population genetic model involving migration on a unidirectional graph, $G = (V,E)$ where V are vertices (demes) connecting edges (E) defined by polygons on a graph (G) – the map defined here as Australia and New Guinea. Two parameters are used $m = \{m_e: e \in E\}$, migration, and $q = \{q_v: v \in V\}$, diversity. m defines a migration estimation on each edge and diversity q is the estimation of genetic dissimilarity within each deme. These estimations are accomplished through a Bayesian framework using a likelihood to measure how well m and q explain the observed data. A *prior* describes the expectation of m and q. In summary, EEMS estimates migration and diversity on a map (Euclidian plain) to estimate the observed from the expected (*prior*). The delineation is then put on a continuous scale with zero as the null, positive values are higher than expected, and negative values are below the expected model. Euclidean space is defined across *L. burtonis*'s range, Australia and Papua New Guinea. Total demes were set to 700 tessellations. MCMC was run 2,000,000 generations sampled every 1000 generations with a 50% burning, and an MCMC thinner of 9999 was implemented across three chains.

In conjunction with EEMS, I also estimated migration across branches of the phylogeny using TreeMix (Pickrell & Pritchard, 2012). To verify if k estimations hold from the cross-validation, a population test was conducted to verify proper population designation in the tree (Supplementary Table 3.) (Keinan *et al.*, 2007; Reich *et al.*, 2009). Drift parameters estimates were tested with all possible migration combinations – four populations have six events for migration and three migration schemes (two asymmetric and one unilateral). Following the product rule, this leads to a total of 18 migration combinations. Sample size correction was turned off with the *–noss* flag and run with 500 bootstrap

replicates.

Tree Dating and Species Distribution Models

Lialis burtonis bifurcations were dated by a time tree estimation using mitochondrial gene *ND*2. The phylogeny and divergence times were simultaneously estimated using BEAST2 v2.5.1 (Bouckaert *et al.*, 2014). I inferred the best-fit models for the partitioned codon positions of *ND2* using the corrected Akaike information criterion (AICc) as implemented in PartitionFinder2 v2.1.1 (Lanfear *et al.*, 2016; Akaike, 1998; Aho *et al.*, 2014). The partitioning scheme was set to have first and third codon positions using a GTR + I site model and the second codon position using the TVM site model. The tree model was set using an uncorrelated relaxed clock and Yule prior (Drummond *et al.*, 2006). I ran two replicate Markov Chain Monte Carlo (MCMC) analyses, each with 100 million generations retaining every 1000th sample. I used seven calibrations to constrain the minimum ages of nodes in the time tree analyses. Fossil calibrations consist of the most recent common ancestor (MRCA) of crown Gekkota, minimum age (Daza *et al.*, 2012; Daza *et al.*, 2014), MRCA of crown *Sphaerodactylus* (Kluge, 1995; Iturralde-Vinent & MacPhee, 1996; Daza & Bauer, 2012), MRCA of *Paradelma orientalis* + *Pygopus nigriceps* (Hutchinson, 1997; Jennings *et al.*, 2003; Lee *et al.*, 2009), MRCA of Helodermatidae + Anguidae (Nydam, 2000), and MRCA of Lepidosauria (Squamata + *Sphenodon*). I also incorporated a biogeographical calibration using MRCA of *Teratoscincus scincus* + *Teratoscincus roborowskii* (Tapponnier *et al.*, 1981; Abdrakhmatov *et al.*, 1996; Macey *et al.*, 1999) and a secondary calibration at the root Lepidosauria + Archosauria (Reisz & Müller, 2004) (Supplementary Table 4.). Output files were checked using Tracer v1.4 (Drummond & Rambaut, 2007); log files were combined with a 10% burn-in from each run, and a consensus was constructed with tree annotator at 20% burn-in from both runs.

I constructed a species distribution model to evaluate the effects that climate had on historical distributions of *L. burtonis*. During glacial periods, land bridges connected the two landmasses, perhaps allowing immigration of *Lialis* from Australia into New Guinea (Jennings *et al.*, 2003; Lee *et al.*, 2009). To see if land bridges played a role in the migration of *Lialis* from Australia to New Guinea, I ran a niche model analysis using MaxEnt v3.4.1 (Steven *et al.*, 2018) to track stability. Climatically suitable areas were initially tracked using the mid-Holocene (6 ka) climatic data. During the Last Glacial Maximum (22 ka, LGM), a land bridge between Australia and the island of New Guinea was present. This climate data was used, not only to estimate if the land bridge habitat was suitable for range expansion but to determine patterns of niche stability in the areas currently inhabited. Finally, climatic data from the Last Inter-glacial period (140-120 ka, LIG) was used to track stable niche spaces during interglaciation further. Following hypothetical expectations from Camargo *et al.* (2010), I expect variations in climate drive diversification along a geographic barrier, especially with the Torres Strait acting as an allopatric barrier bisecting *Lialis* between PNG and Australia.

I used 81 occurrence points and climatic layers for WorldClim v1.4 with a special resolution of 30 arc-seconds for LIG (Otto-Bliesner *et al.*, 2006) and the mid-Holocene and 2.5 arc-seconds for LGM (Hajimans *et al.*, 2005). Paleoclimatic models were provided by the Community Climate System Model (CCSM4) (Otto-Bliesner *et al.*, 2006). For range extensions and proper model definition, I used a buffer zone of 300 km from current distribution across continental Australia and New Guinea (Anderson & Raza, 2010). A pairwise Pearson's Correlation was used to select bioclimatic layers that have a correlation under 0.85 (Supplementary Table 5.). Of the 19 bioclimatic layers available, a summary is available in Table 3. Contributions of each variable of each species distribution model were

conducted with the Jackknife test. Each run was replicated using 50 cross-validations. Tuning to improve performance of MaxEnt models (Anderson & Gonzalez, 2011; Radosavljevic & Anderson, 2013) was performed on each population on current climate models; however, estimations of distribution presented here were conducted at the species level, since population estimates from the past may not match ranges of contemporary populations. I created a stability map of niche space estimations across paleoclimates in QGIS 3.2.1 (QGIS Development Team, 2019) to observe if evaluations of paleoclimate through time are suitable.

Table 3. Paleoclimate layers used in this study.

Geologic Category	Layer	Variable
LGM	Bio1	Annual Mean Temperature
Mid-Holocene, LGM, LIG	Bio2	Mean Diurnal Range
Mid-Holocene, LGM, LIG	Bio3	Isothermality
Mid-Holocene, LGM	Bio4	Temperature Seasonality
Mid-Holocene	Bio6	Min Temperature of Coldest Month
Mid-Holocene, LGM, LIG	Bio8	Mean Temperature of Wettest Quarter
Mid-Holocene, LGM, LIG	Bio9	Mean Temperature of Driest Quarter
LGM, LIG	Bio10	Mean Temperature of Warmest Quarter
LIG	Bio11	Mean Temperature of Coldest Quarter
Mid-Holocene, LGM, LIG	Bio15	Precipitation Seasonality
LGM	Bio16	Precipitation of Wettest Quarter
Mid-Holocene, LGM	Bio17	Precipitation of Driest Quarter
Mid-Holocene, LGM, LIG	Bio18	Precipitation of Warmest Quarter

LIG	Bio19	Precipitation of Coldest Quarter

Results

Phylogenetic Analysis

Phylogenetic analysis of mitochondrial data (Fig. 2A) has >0.95 posterior probability support for four significant clades across Australia and Papua New Guinea. These populations fall into four major regions of Australia: (1) a "Woodland" population in the northeast that includes savanna in Papua New Guinea; (2) a "monsoonal" population in the north within the monsoonal tropics and below 500 m elevation separated from the woodland by the Gap of Carpentaria (Joseph *et al.*, 2013; Pepper *et al.*, 2017); (3) a "Pilbara" population; and (4) an "Arid" population from the interior of Australia within the arid zone, mostly below the tropic of Capricorn, unless above 500 m, and west of the woodland biomes to the east where a transition to desert begins.

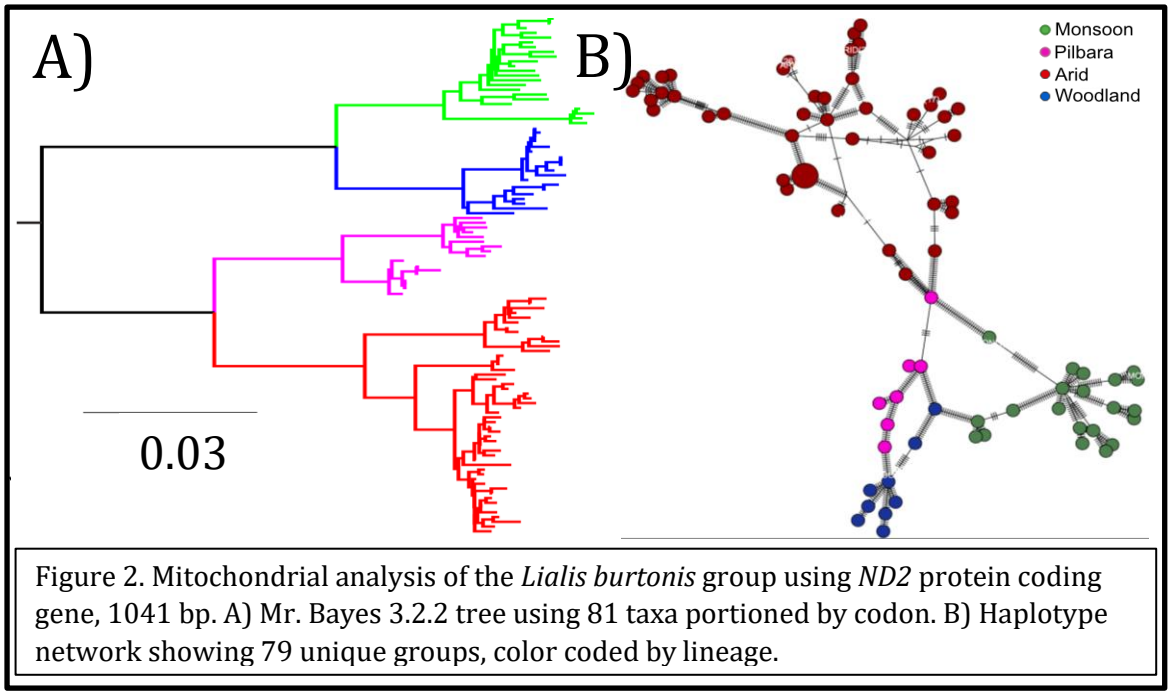


Figure 2. Mitochondrial analysis of the *Lialis burtonis* group using *ND2* protein coding gene, 1041 bp. A) Mr. Bayes 3.2.2 tree using 81 taxa portioned by codon. B) Haplotype network showing 79 unique groups, color coded by lineage.

Phylogenetic analysis of nuclear genomic data in SNAPP (Fig. 3) shows 1.0 posterior

probability support for each of the four defined groups. TETRAD also shows all four defined groups with lower bootstrap support than the SNAPP analysis. However, there is a deep split within Arid, more prominent on an unrooted tree, is observed in the TETRAD phylogeny (Fig. 4A). Unrooted neighbor network analysis shows clearly defined grouping, with the Arid population the most probable origin of every other population radiation (Fig. 4B).

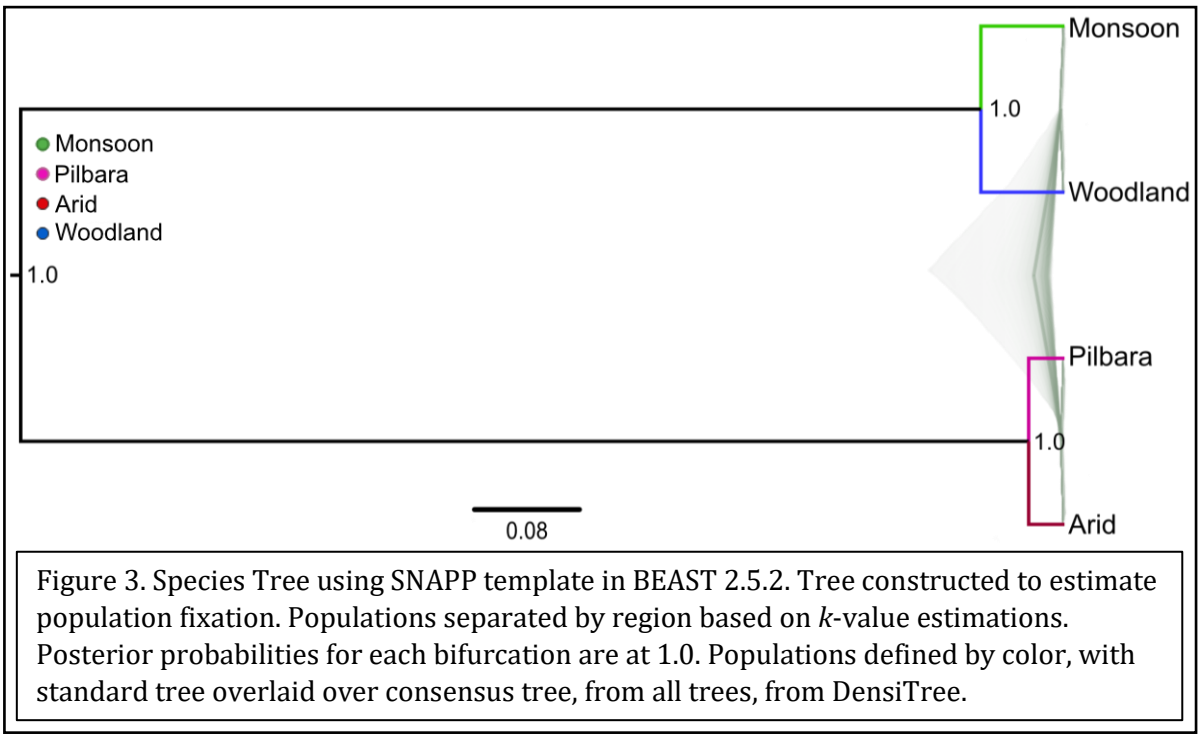


Figure 3. Species Tree using SNAPP template in BEAST 2.5.2. Tree constructed to estimate population fixation. Populations separated by region based on *k*-value estimations. Posterior probabilities for each bifurcation are at 1.0. Populations defined by color, with standard tree overlaid over consensus tree, from all trees, from DensiTree.

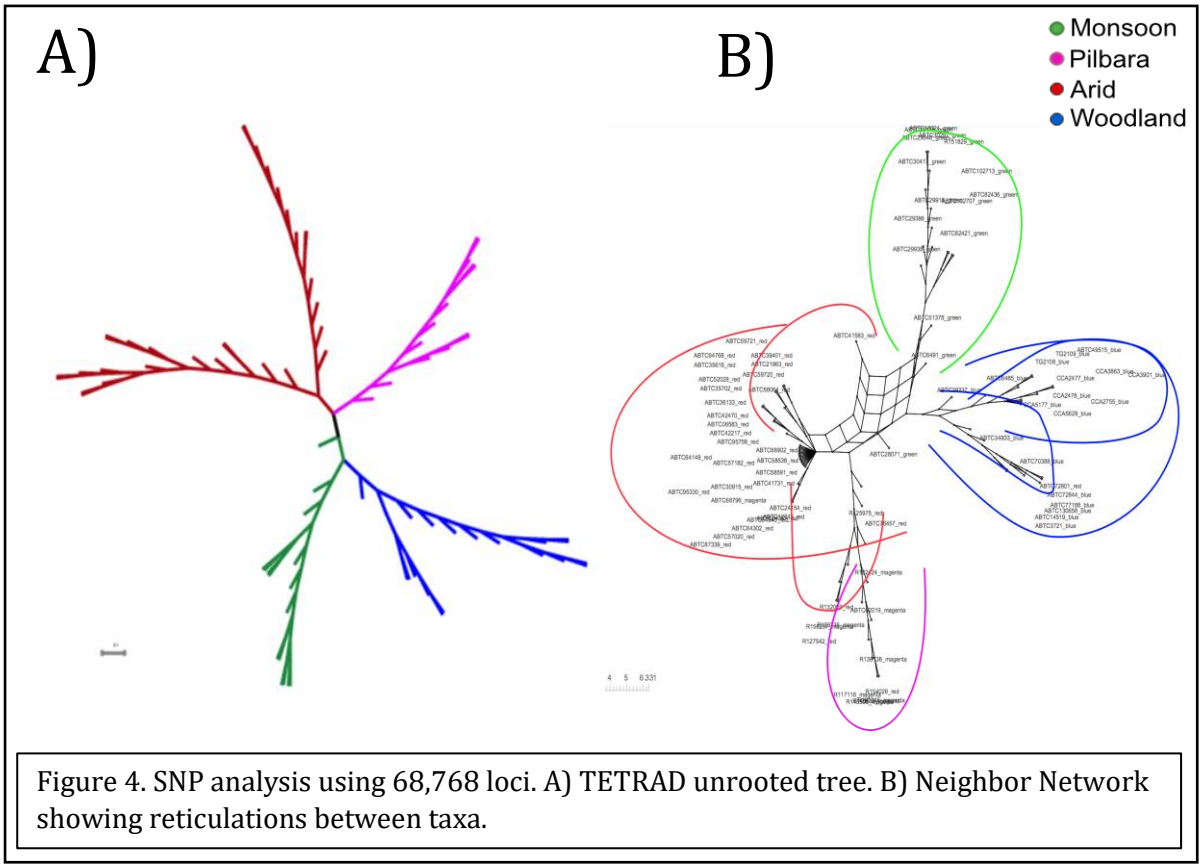


Figure 4. SNP analysis using 68,768 loci. A) TETRAD unrooted tree. B) Neighbor Network showing reticulations between taxa.

Gene Flow

DAPC shows a clear population structure between all populations of *L. burtonis* with potential gene flow among overlapping samples within centroids (Fig. 5). Among the defined *k* populations only 4 of the 81 individuals are classified outside their *a priori* designations. These samples were dropped from the final analysis, since missing data was high, attributing to erroneous population grouping.

There are reticulations (Fig. 4B) where populations are in close proximity or overlapping in range with the Arid population. The Arid population, being centrally located, is a source for gene flow in areas of overlapping range. Specifically, ABTC41583 (Arid) has admixture with the Monsoon group in the Kimberly, and ABTC28071 (Monsoon) admixes south of the Tanami Desert, where the Monsoonal group expands out to the arid biome.

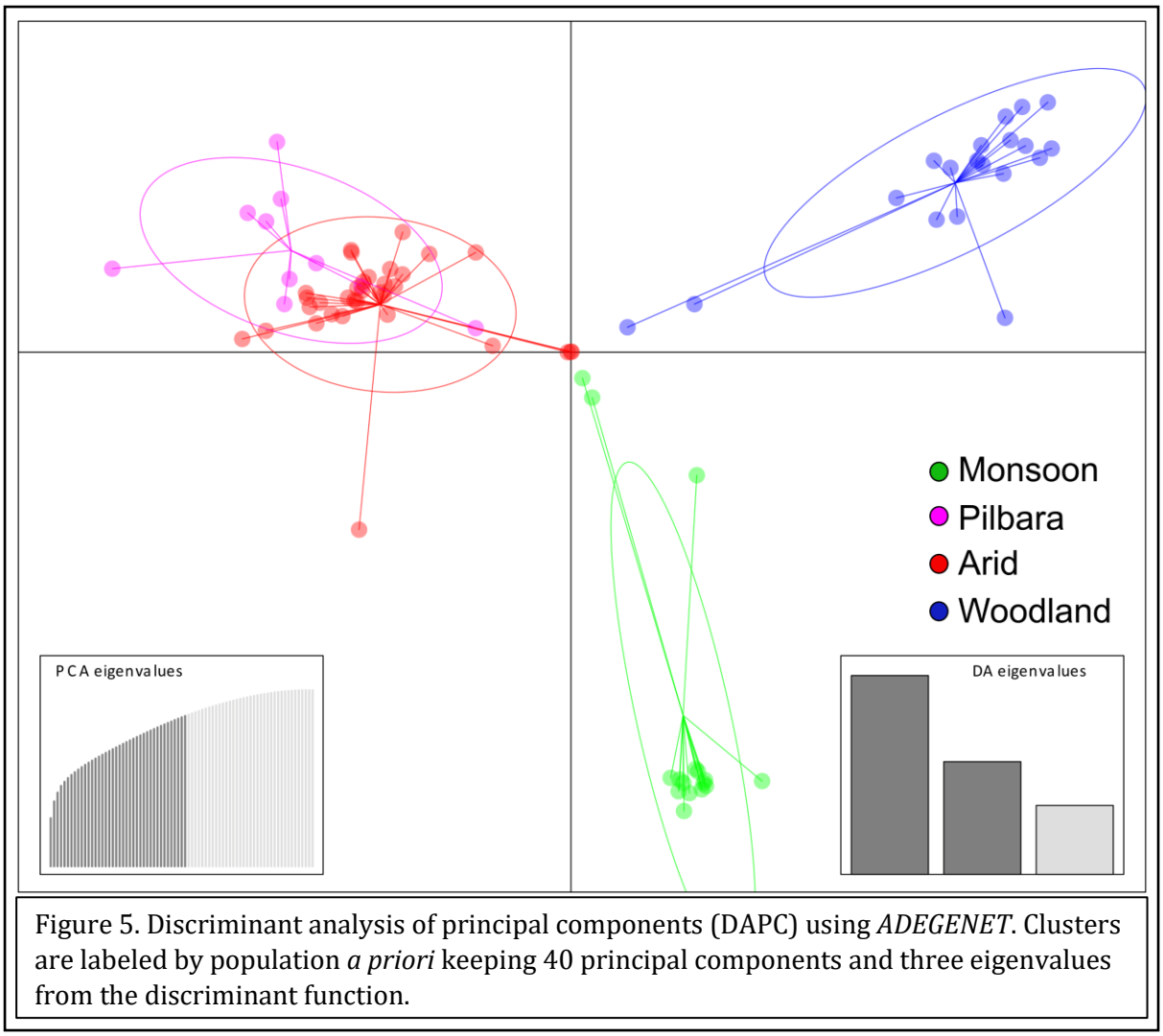


Figure 5. Discriminant analysis of principal components (DAPC) using *ADEGENET*. Clusters are labeled by population *a priori* keeping 40 principal components and three eigenvalues from the discriminant function.

ADMIXTURE results (Fig. 6) show that all groups have introgression along areas of their distribution bordering other populations. The population with the least amount of ADMIXTURE is the Woodland population in eastern Australia and PNG. This group has what looks to be ancestral alleles present or introgression from backcrosses with the Arid population (Fig. 7), without any gene flow into the Pilbara. The Pilbara population is closely related to the Arid population and has introgression between a few samples that expanded into the arid zone. Overall, however, within each population's distribution, there is minimal gene flow between populations. There is no introgression among *L. burtonis* in PNG with other continental populations. With that being the case, the Woodland population that persists

on the mainland does have introgression between the monsoon populations as well as the arid population near the border between each range. Among the *k*-values with the lowest error was a *k*-value of four, with *k*-value six only slightly higher in error from the cross-validation. Alternative *k*-value schemes show that the Arid population is split into two subpopulations – one to the west confined south of the MacDonell Range and into the Great Victoria basin. The other is broadly within the Simpson Desert. The other group split is the Woodland population, which is partitioned between the mainland and the island of New Guinea.

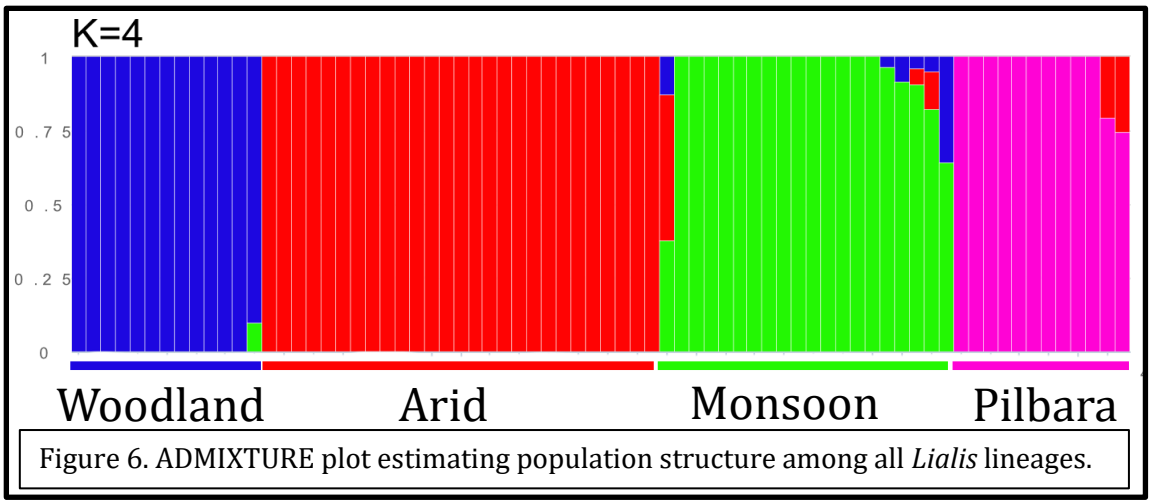


Figure 6. ADMIXTURE plot estimating population structure among all *Lialis* lineages.

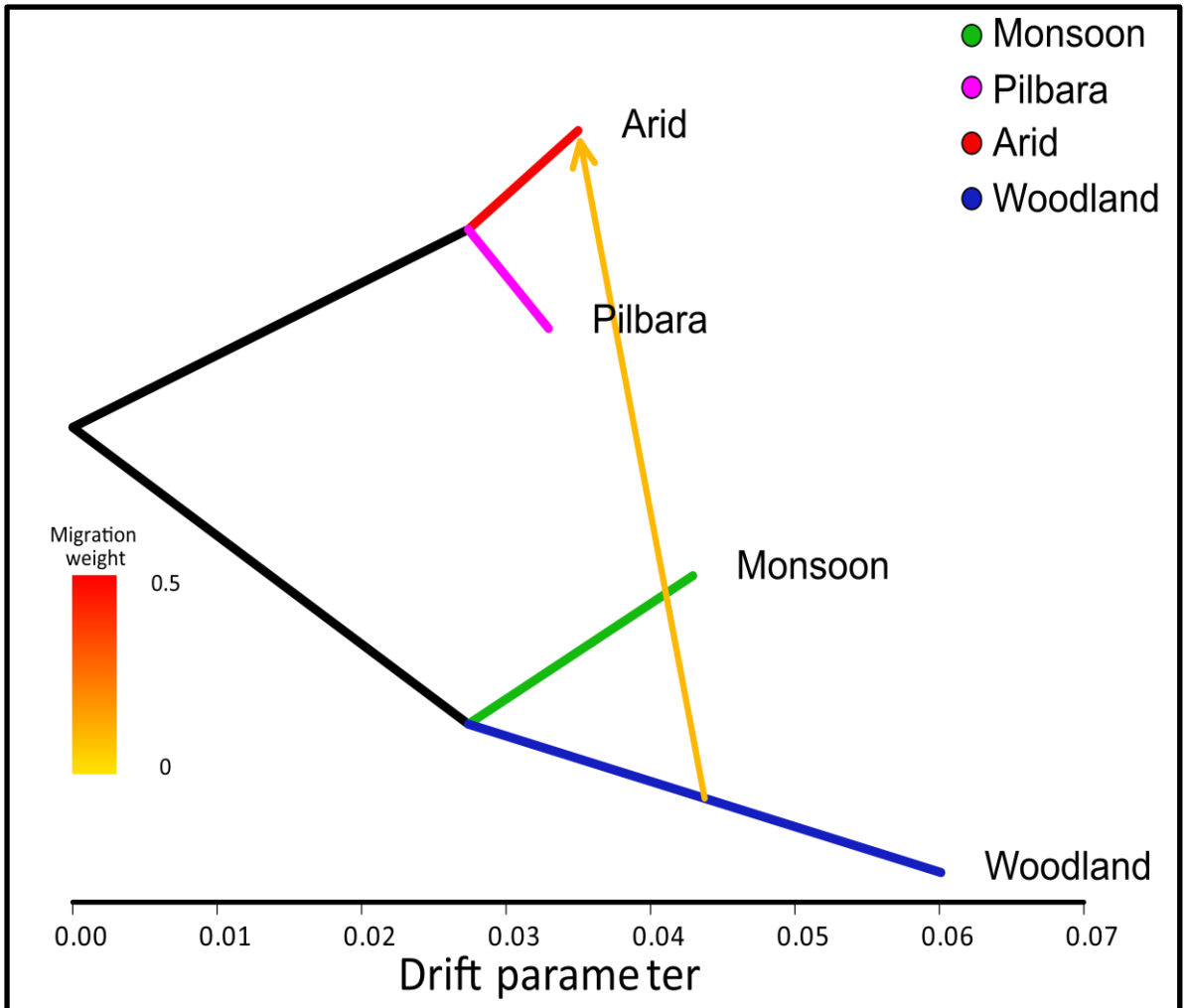


Figure 7. TreeMix ML estimation of directional migration between populations.

Phylogeography

Estimation of migration from TreeMix shows unidirectional migration from the Woodland population to the Arid population (Fig. 7). This admixture is found in the Great Artesian Basin. Estimating effective migration surfaces (EEMS) shows a clear barrier between lineages where elevation is over ~500 m, and where transitions from woodland habitat to the arid zone are present (Fig. 8). Diversity across the mainland is lower than the expected models to the east, where the Woodland population is genetically conserved within the group. There is higher than expected diversity to the west between the other three

populations, with a stark border between high and low diversity following the boundary of the Great Artesian Basin.

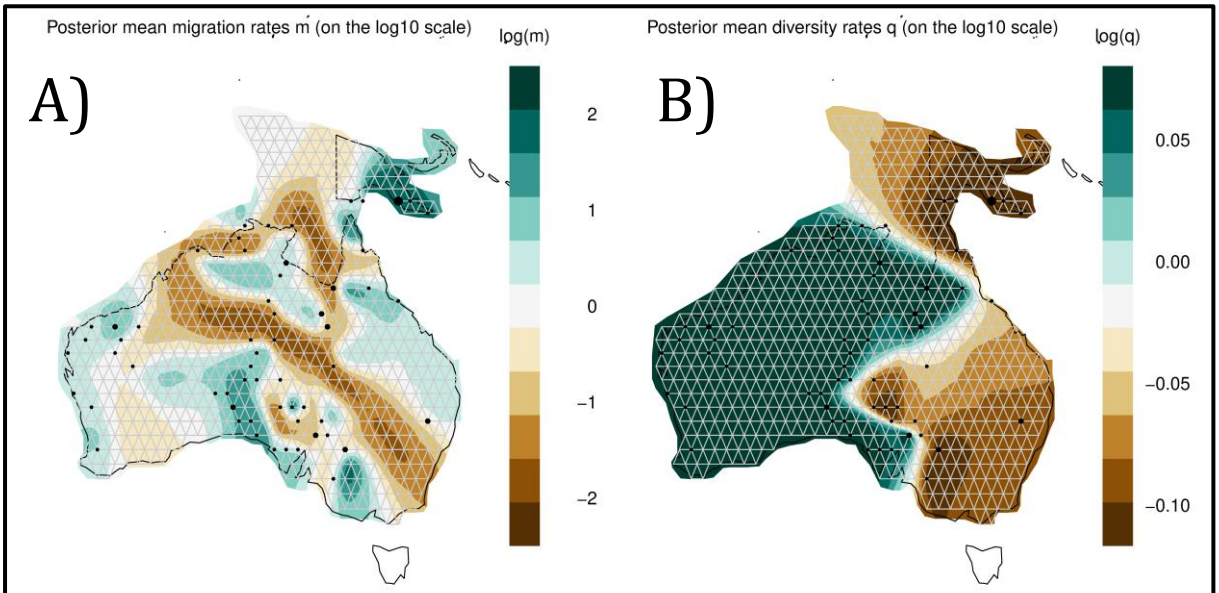


Figure 8. *EEMS* estimation of *Lialis* and its demes. A) Migration estimations where orange-brown colors are estimations of allele migrations between populations under the expected model (0), while indigo colors are estimation above the expected model of allele migration. B) Diversity estimations where orange-brown colors are estimations of allele diversity between populations under the expected model (0), while indigo colors are estimation above the expected model of allele diversity.

Time tree estimations (Fig. 9), from mitochondrial data, show *Lialis* populations diverged at the CT boundary. *L. burtonis* populations' initial split is approximately 66.1 mya, which is concordant with many studies of pre-Pleistocene origins (Jennings *et al.*, 2003). Two subsequent bifurcations happened around the same time in the arid aone and monsoonal zone in the Eocene around 49.7 mya and 45.5 mya, respectively. Respective populations become independent between 24 and 30 mya, with the most recent of those being the Woodland population. The first potential colonization of PNG was during the early Miocene. The Woodland population, residing in PNG, split from the mainland population around 4 mya, a pre-Pleistocene split.

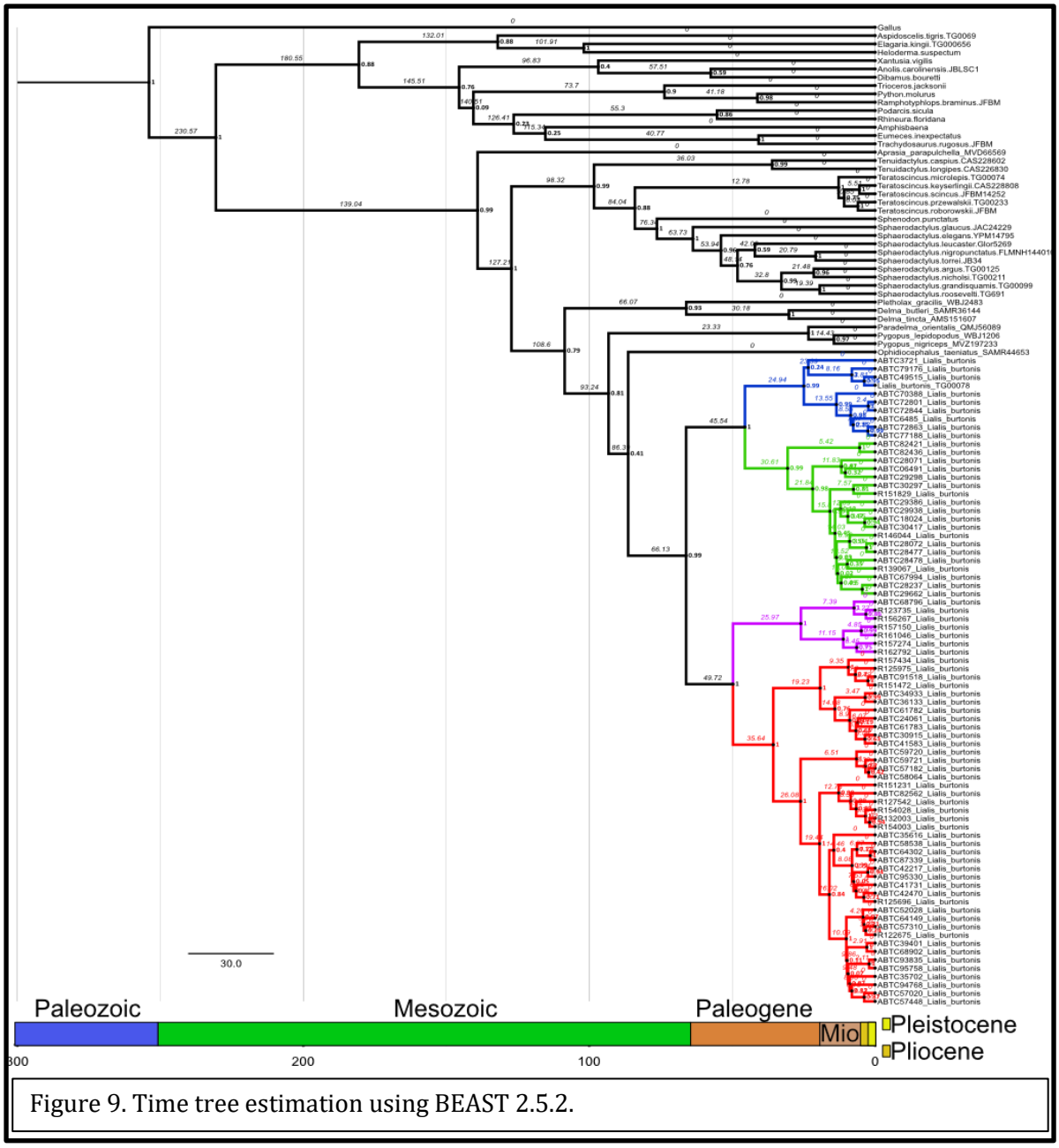


Figure 9. Time tree estimation using BEAST 2.5.2.

Species distribution models have AUCs of 0.674, 0.615, and .647 for LGM, LIG, and the mid-Holocene, respectively. Each model illustrates the stability of habitable niche space through time between New Guinea and mainland Australia. Including the Sahul Shelf at 150 m drop in sea level during glaciation maxima (Fig. 10).

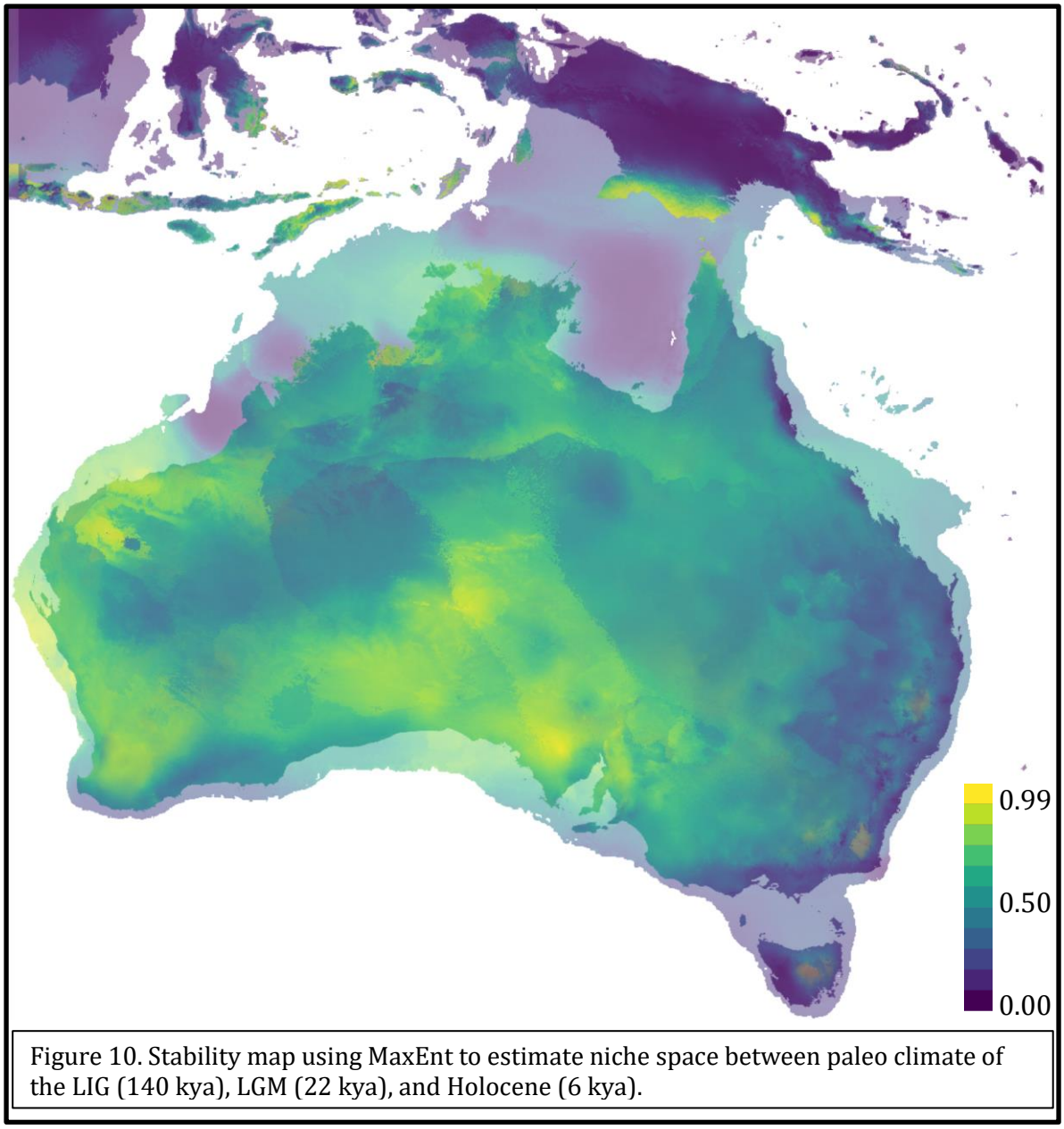


Figure 10. Stability map using MaxEnt to estimate niche space between paleo climate of the LIG (140 kya), LGM (22 kya), and Holocene (6 kya).

Discussion

With little gene flow and low migration between each population of *L. burtonis*, the populations are likely diversifying in isolation and in an early stage speciation event. Being that the oldest node in the time tree is 66.1 million years old, however, it is prudent to understand that fluctuations of wet and dry periods may link populations together again. *L. burtonis* is an anomaly to life histories of other pygopod species, being an active squamate

predator. Because of this, its home range is more extensive than most small geckos. Within population, *L. burtonis* has low fixation based on the size of its distribution *e.g.*, not diverse within populations. This low fixation is probably attributed to its life history of active predation and the presumably broader range it has being an active predator. The species is better adapted to migrate, potentially mitigating diversification within populations.

Over the LGM, LIG, and mid-Holocene stable niche space has occurred between Australia and PNG over the last 150k years. Along with the Sahul Shelf, *L. burtonis* has had many opportunities to emigrate from Australia into PNG (Byrne *et al.*, 2011). It is not fully understood if *L. burtonis* colonized PNG during the LGM since time tree estimates place the split between Australia and PNG as pre-Pleistocene. The most likely scenario is that the woodland population had emigrated multiple times to PNG when sea levels were at historic lows, exposing the Sahul Shelf and new stable niche space, allowing access to New Guinea.

Allopatry is likely the mechanism acting as a barrier to gene flow between populations. Each population resides in each major biome of Australia, known drivers of diversification (Byrne *et al.*, 2008). Only when individuals colonize outside their home biome does gene flow occur. The Monsoon population does not extend past the Carpentarian Gap (MacDonald, 1969), and only extends to the arid zone of Australia while not extending past the MacDonnell Ranges to the south or the Kimberley Plateau to the west. The monsoonal tropical zone has shifting climate compared to the arid zone; however, since individuals extend past respective borders of each biome but not past the topographical barrier, physical barriers are a mechanism driving isolation of populations.

The Pilbara population has a low *Fst* between the Arid population and does not have any truly defined physical limiters with the Arid population. A unique aspect to the Pilbara population is the region has a high rate of endemism (Pepper *et al.*, 2013) which may

correlate to this relatively recent independence, though the Pilbara group is not truly isolated to the Pilbara region, and not an actual physical barrier. The Kimberley Plateau in conjunction with the King Leopold Range separates suitable habitats between the Monsoon population (Catullo *et al.*, 2014).

The highest *Fis* values are found within the Arid population. This result is attributed to the range of this region being the largest. The expansive size or the arid zone allows for isolation by distance, which is a probable driver given the high fixation index within the population. The Great Artesian Basin is an elevational gradient that does not allow penetration into the Woodland population's range. The Arid population is also a sink to all other *L. burtonis* populations, as all other populations have introgression at the periphery of their range with the Arid population.

The Woodland population is interesting since it has colonized PNG from Australia. There have been many traceable land bridges between Australia and PNG across the Arafura Shelf and Sahul Shelf (Hall, 2009). The last land bridge was during the LGM. Niche stability maps estimate stable niche space where the population currently resides on PNG. However, if a *k*-value of six populations is used, the Woodland population is split in two. This split does not, however, sequester one subpopulation to PNG and the other to Australia. Instead, the PNG subpopulation resides within PNG and on the Cape York Peninsula. The split between this group is approximately 3.81 mya, well outside the timeframe of the LGM land bridge. Based on the results, the most parsimonious scenario is that this group diversified in Australia then secondarily colonized PNG during a land bridge event. The Torresian Barrier may be a zone promoting isolation between the PNG and Australian subpopulation of the Woodland population (Ford 1986; 1987). However, this barrier separates lowland the dry climate habitat from the wetter, higher elevation, tropical habitat, and there is a visible corridor, though niche

stability estimations put the probability of suitable niche space below 0.50.

Barriers to gene flow are concordant with many other taxa within Australia (Jennings *et al.*, 2003), The monsoon population also is sequestered to the monsoonal tropical zone in Northern Australia. Adaptation and selection within populations are undoubtedly due to climate, but even more so due to elevation gradients, where uplift of the land to the south is a barrier to entry within the arid zone. Elevation also seems to have a role in isolation between the woodland group and the arid group.

Chapter 3

How Gene Flow Regulates Speciation: Genomic Insights to the Mechanisms Leading to Isolation and Subsequent Diversification in *Heteronotia binoei*

Abstract

1. The dynamics between gene flow and diversity play a pivotal role to understand the processes driving speciation and maintenance of a species. In this chapter, the *Heteronotia binoei* system is used to uncover the processes and mechanism promoting or limiting gene flow between areas of contact between independent lineages.
2. *Heteronotia binoei* is a hyperdiverse and widespread species ranging across all major biomes in Australia. Three contact zones, at varying degrees of divergence with independent lineages overlapping in range, are used to elucidate the dynamics between gene flow and diversity, and its role in speciation. I used approximately 3,000 exons, nine introns, and one mitochondrial gene using traditional Sanger Sequencing techniques along with next-generation targeted sequence capture. Niche modeling was incorporated to estimate environmental and geological barriers as mechanisms influencing gene flow.
3. I found that contact zones with high degrees of divergence had less gene flow, while contact zones between sister taxa had continued gene flow and migration of alleles. Transition zones between wet and dry areas in the mesic habitats in Queensland were a driver mitigating gene flow, and subsequent isolation, leading to higher rates of diversity.
4. The eastern contacts zone between the CYA6 and EA6 groups was highly influenced by Australia's aridification events, where three transition zones adjacent to the Great

Dividing Range act as barriers to gene flow. The south-central contact zone in Queensland has two divergent lineages in secondary contact that have no observable gene flow or allopatric restrictions. Two lineages (EIU and GULF) are in sympatry and likely isolated by genetic distance, causing pre- or post-zygotic isolation as a reproductive barrier. Finally, the third contact zone to the western border of Queensland is enigmatic and does not have a clear signal of gene flow. This group is not highly diversified with respect to the lineages in contact. In this contact zone, a likely response to unclear gene flow is sampling size, but some physiological character is likely leading to a lack of gene flow caused by pre- or post-zygotic isolation as a reproductive barrier.

Introduction

Uncovering the mechanisms and processes of speciation gives new insight into life's continued persistence on Earth. As evolution continues across time, lineages separate into new species to adapt to the dynamic changes of life. A myriad of mechanisms promotes or limit gene flow through geography, genetic incompatibilities, or behavioral preferences. Gene flow has a significant influence on the processes driving speciation, especially in its initial stages. A high degree of gene flow limits speciation (Kearns *et al.*, 2018) since ongoing admixture between lineages mitigates divergence.

Conversely, as species become diversified independently, gene flow is hindered, driving isolation. Lineages can then evolve independently. The geographic features within a species' range have the potential to limit or promote gene flow, which in turn influences diversification. As dynamic topography changes a species' range, barriers to gene flow can form – promoting isolation. Climate, too, can cause species to expand or recede in their

range, leading to refugial niche space, further leading to isolation and lack of gene flow. Isolation, and therefore, the lack of gene flow, can promote speciation as lineages become more divergent, potentially driven by geography. To understand how geography mechanistically influences the rate of gene flow and the dynamic processes gene flow contributes to speciation, this study focuses on an Australian gecko, Bynoe's gecko (*Heteronotia binoei*), to elucidate the driving mechanisms of gene flow and how it plays a role in divergence.

To understand the mechanisms promoting or limiting gene flow, we look to Australia's well-studied geologic and climatic history. The Australian climate has experienced extreme shifts and differentiation in topographic features throughout geologic time. The contemporary environment is split into four, large climatic regions (from here on referred to as biomes) (Cogger & Heatwole, 1981; Crisp *et al.*, 2009). These four biomes are the dry arid biome, eastern mesic biome, south western mesic biome, and tropical monsoon biome. The largest of these biomes is the arid covering, 70% of Australia (Byrne *et al.*, 2008) and resides in the central portion of the country. The tropical monsoon biome in the north consists of both subtropical and tropical environments *e.g.*, grasslands, savannas, rainforests, and scrublands. It is influenced by an annual monsoonal season during the southern hemisphere's summer. On the east coast, the Great Dividing Range provides a gradient of low elevation to high elevation, creating a rain shadow. Ecoregions vary along the eastern mesic biome, with scrublands, grasslands, and temperate forests farther south; and a Mediterranean climate exists in the south western mesic biome of the continent encompassing the southwestern region of Australia.

The climatic history of Australia has influenced the diversification of *Heteronotia binoei*. The genus has an unusually cosmopolitan range throughout Australia, including

lineages in the desert, arid scrub, and mosaic rock outcrops in the arid biome, and north to dry tropics that include savanna, scrub, and southern rainforest in the monsoonal biomes (Fujita *et al*., 2010; Pepper *et al*, 2011). *Heteronotia* is said to be hyperdiverse across its range (Kearney & Shine, 2004; Strasburg *et al.*, 2005; Fujita *et al.*, 2010; Pepper *et al.*, 2011; Moritz *et al.*, 2016). Rock-specialized constituent species, *Heteronotia spelea* and *Heteronotia planiceps*, highlight the relevance of an understanding of the geological setting around which the genus has evolved. Regions where rock dwelling species reside, in the Kimberley and Pilbara regions, are highly diverse in their geology and topography. To the south lie uniform expanses of linear sand ridges in inland deserts, which act as barriers leading to isolation between groups of *Heteronotia binoei* lineages (Beard, 1979; Fujita *et al.*, 2010; Pepper *et al.*, 2013). Biogeographic patterns of *Heteronotia* were influenced by aridification, which significantly drove biotic diversification through expanding distributions during interglacial periods (Ackerly, 2003; Avise, 2000; Levins, 1968). The Southern Hemisphere experienced arid expansion during interglacial periods from the mid-Miocene to the Pleistocene (Byrne *et al.*, 2008; Martin, 2006). As with *Heteronotia*, much of the within-species biodiversity in Australia arose as a result of aridification cycles in the Pliocene (Clapperton, 1990; Markgraf *et al.*, 1995). These aridification events are known to cause extinctions (Cogger & Heatwole, 1981; Crisp *et al.*, 2009), thus opening large new adaptive zones that allowed invasions of new lineages from previously occupied mesic environments to new arid niches (Rabosky *et al.*, 2007). These paleoclimatic fluctuations resulted in both extensions and retreats between biomes that influenced the demographic histories of populations, including population expansions and gene flow dynamics in *Heteronotia* (Fujita *et al.*, 2010; Pepper *et al.*, 2011; Pepper *et al.*, 2013).

Divergent time analyses from Pepper *et al.* (2011) and Fujita *et al.* (2010) suggest the divergence between species (*H. binoei*, *H. planiceps* and *H. spelea*) is placed conservatively around 5 mya to 2.4 mya, which corresponds with the Pliocene through to the early Pleistocene. These analyses suggest near-simultaneous diversification (rapid radiation), corresponding to interglacial periods of aridification and refugial populations migrating to newly opened niches. The divergence time estimates coincided with a period of extreme environmental change associated with deepening aridity and increased seasonality (Byrne *et al.*, 2008; Fujita *et al.*, 2010; Pepper *et al.*, 2011). The *H. binoei* complex diverges within 2.27 and 0.69 mya for monsoonal tropical lineages and 1.14 and 0.3 mya for arid biome lineages, with strong signatures of recent population expansion in the north <1 mya (Fujita *et al.*, 2010).

The various mechanisms by which environmental gradients drive diversification on populations are well studied (Haldane, 1948; Fisher, 1950; Endler, 1977). Using allele frequency to plot genetic variation along a geographic plane shows that, as environments change, ecotones can have abrupt genetic discontinuities, hybrid zones, and distinct species boundaries (Endler, 1977; Slatkin, 1978; Kirkpatrick & Barton, 1997; Barton, 1999). Changing environments can cause a lack of gene flow (reproductive isolation) between populations, through both allopatric barriers and parapatry, by limiting migration at the periphery of a population's range, influencing phylogenetic relationships (Irwin, 2012a).

The interplay of gene flow and divergence has different influences depending on if the populations are in early (higher gene flow) or later (higher divergence) stages of speciation. *Heteronotia* has tremendous utility in understanding the mechanisms and processes of speciation through an examination of gene flow between lineages in overlapping ranges across heterogeneous geographic areas in Australia. In this study, we aim to

investigate the dynamics of gene flow between zones of secondary contact, at varying degrees of divergence, across heterogeneous environments, to understand the mechanisms and processes driving speciation events. Using the hyperdiverse *Heteronotia binoei* complex, we identified geographic areas of presumed gene flow to understand what mechanisms and processes facilitate speciation or hamper it. Individuals from nine distinct lineages across three explicit contact zones at varying degrees of divergence in Queensland, Australia were sampled. The *Heteronotia binoei* complex has utility in understanding the interplay that gene flow and divergence has on speciation.

Methods

Taxonomic Sampling

Sampling of *H. binoei* took place over four weeks in May 2015 across Queensland, Australia (Fig. 11, Table 4). Any specimens identified, by an expert, as a newly discovered lineage were sacrificed under the proper protocols given by Australian animal ethics. Whole bodies were formalin fixed and preserved in 70% ethanol. Liver samples were preserved in RNALater. Individuals that could be assigned to an existing lineage were released after collecting weight, SVL, photographs, and tail tissue for subsequent DNA analysis.

Table. 4. Point localities for exon captured *Heteronotia binoei* samples in this study.

Sample Name	Sample Origin	Latitude	Longitude	Lineage
JET01_CCM5138	Milgarra	-18.19008	140.69069	GULF-W
JET01_CCM5100	Maryvale	-17.93463	141.75894	GULF-E
JET01_CCM5137	Milgarra	-18.19008	140.69069	GULF-W
JET01_CCM5232	Mt Mulgrave	-16.36888	143.97714	CY6-N

JET01_CCM5237	Archer River	-13.47466	142.97212	CY6-N
JET01_CCM5249	Fairlight Station tip	-15.74851	144.04677	CY6-N
JET01_CCM5264	Rookwood tip	-17.04801	144.35088	CY6-N
JET01_CCM5321	Yarraden Station	-14.31951	143.27472	CY6-N
JET01_CCM5340	West Musgrave	-14.79313	143.40062	CY6-N
JET01_CCM5347	Mt McLean Station tip	-15.90897	144.83844	CY6-N
JET01_CCM5401	11km E of Kalpowar	-24.69246	151.39261	Biloela
JET01_CCM5155	Cobbold dump	-18.80151	143.43626	CY6-S
JET01_CCM5158	8.5km east of Emnford	-17.41587	145.09802	CY6-S
JET01_CCM5188	34km E of Georgetown	-18.27407	143.83688	CY6-S
JET01_CCM5194	N. Forsayth	-18.5667	143.56871	CY6-S
JET01_CCM5201	9km N of Mt Garnet	-17.63921	145.06319	CY6-S
JET01_CCM5230	48km NW of Chillagoe	-16.8878	144.21411	CY6-N
JET01_CCM5254	8.5km E of Emuford	-17.41586	144.09802	CY6-S
JET01_CCM5255	8.5km E of Emuford	-17.41586	144.09802	CY6-S
JET01_CCM5332	Bamboo Range	-14.62601	143.45819	CY6-N
JET01_CCM5374	Mt Stuart	-19.35871	146.80519	EA6
JET01_CCM5381	Lynchs beach	-19.45667	147.48341	EA6
JET01_CCM5130	Gilbert River	-18.19342	142.89709	GULF-E
JET01_CCM5165	Cobbold camp	-18.79654	143.42195	EIU
JET01_CCM5171	Cobbold dump	-18.80151	143.43626	EIU
JET01_CCM5172	Cobbold sandstone	-18.81071	143.40161	EIU
JET01_CCM5183	Gilbert River	-19.03035	143.33352	EIU
JET01_CCM5206	Cobbold sandstone	-18.81071	143.40166	EIU
JET01_CCM5099	Maryvale	-17.93463	141.75894	GULF-E
JET01_CCM5122	Mayvale Station tip	-17.93452	141.75897	GULF-E
JET01_CCM5140	Milgarra	-18.19008	140.69069	GULF-E
JET01_CCM5177	Old Robin Hood	-	143.6097	GULF-

		18.95333		W
JET01_CCM5107	Coolullah Station	-19.82248	140.16267	GULF-W
JET01_CCM5109	Bang Bang Jump Up	-18.5194	140.67236	GULF-W
JET01_CCM5353	Maitland Downs	-16.22185	144.7083	GULF-W
JET01_CCM5361	Glencoe Road junction	-18.09938	145.19406	GULF-E
JET01_CCM0147	Camp Bushy Park	-21.18974	139.75507	CQLD
JET01_CCM0119	Blac Tip 37	-19.53237	144.06488	CY6-S
JET01_CCM0117	Blac Tip 27	-19.58618	144.01645	EIU
JET01_CCM0121	Rifle Creek Station	-20.92199	139.55646	SM6-Ne
JET01_CCM1714	Willeroo Station Tip	-15.29606	131.58511	SM6-NW
JET01_CCM0370	Calvert Hills Campsite	-17.19790	137.43430	SM6-Nd
JET01_CCM0257	Thorntonia Station Tip	-19.50150	138.94908	NWQ
JET01_NTM37234	DNA exon capture			planO
JET01_WAM132771	DNA exon capture	-16.0361	128.7753	SM6NW
JET01_CCM0603	Vic River Research Station Tip	-16.12643	130.95403	SM6NW
JET01_CCM0228	Thorntonia Station Tip	-19.50150	138.94908	NWQ
JET01_CCM1837	Delamere Tip	-15.6236	121.6334	SM6-NW

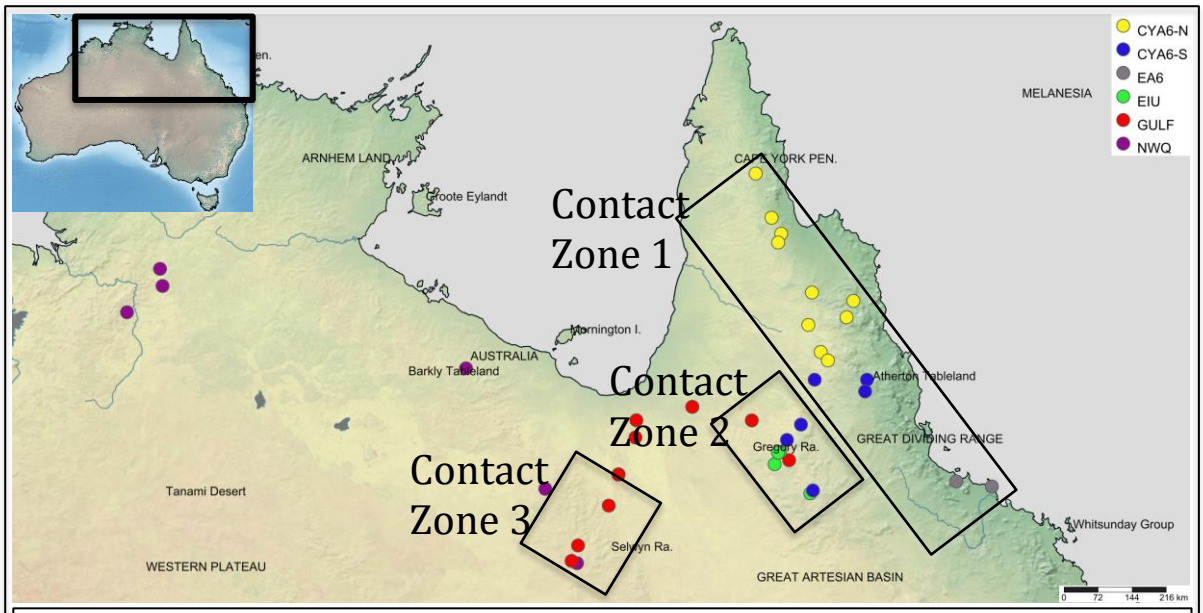


Figure 11. Locality of map of *Heteronotia* lineages across north east Queensland. Contact zones between lineages labeled one through three. Each lineage is color coded.

Exon Capture

DNA was extracted from either tail-tip or liver tissues. 15 uL of protease K and 330 µL of TNES (2.5 mL Tris [1M], 4.0 mL NaCl [5M], 2.0 mL EDTA [500 mM], 5.0 mL SDS [5%], and diluted to 50.0 mL DI water to create stock) were added to macerated tissue. Tissue solution was vortexed and incubated at 55° C for 1-3 hours. 170 mL of NaCl [5M] was added to the solution, vortexed, and spun down at 1,400 rpm for 10 minutes. The supernatant was transferred into a new tube, and an equal amount of 200 proof ethanol was added and briefly mixed by inverting the tube several times. The sample was spun down at 1,400 rpm for 10 minutes. The supernatant was removed by aspiration with care to not disturb or pipette the DNA pellet at the bottom. 70% ethanol was added and spun down once more at 1,400 rpm. The supernatant was aspirated off, and the sample was allowed to dry to remove all traces of ethanol briefly. The purified DNA was then resuspended in 20 uL of Milli-Q H_2O (Sunnucks & Hales, 1996).

For acquiring the genetic data, I used a probe set of approximately 4000 exons for a sequence capture system previously designed by the lab of Dr. Craig Moritz. It uses liver-derived transcriptomes from *Heteronotia binoei* to design sequence capture probes. ~4,000 orthologous exons were identified with an average base pair length greater than 200 bp following the strategy of Bi *et al.*, (2012) and Bragg *et al.*, (2016). Exon capture probes were generated *in silica* and synthesized by NimbleGen as a SeqCap EZ Developer Library kit. Illumina library preparation followed Meyer & Kircher (2010) with slight modifications, as described in Bi *et al.* (2013), and hybridizations were done using the manufacturer's protocols (SeqCap EZ Developer Library; NimbleGen). The first DNA fragment size selection was conducted on a Biorupter Sonicator (Diagenode) and checked by gel electrophoresis for target length. If fragments were not of the adequate size range, samples were sonicated again. Capture pools consisted of 56 samples individually indexed with standard Illumina P7 and P5 adapters per pool with custom barcodes from Meyer & Kircher, (2010). Probes were then combined with initial phase libraries for hybridization. These hybridizations incubated initially for five minutes at 95° C and then for 60 hours at 47° C. To eliminate custom barcodes from hybridizing a set of barcode specific blocking oligos (1,000 pmol) were implemented. After hybridization, I did a probe wash and recapture of sequence targets, following manufacturer protocols (SeqCap EZ Developer Library; NimbleGen). Two enrichment PCRs of postcapture libraries were run for 17 cycles. A Labchip DS (Caliper Life Sciences) was used for DNA quantifications. qPCR was performed on targeted loci to assess global enrichment efficiency (Bi *et al.*, 2012). Confirmation of target enrichment was done by amplifying targets from aliquots of captures and nontargets, as a control. Illumina NextSeq 550 at ANU's Research School of Biology was used for genetic sequencing.

Bioinformatics

The resulting sequence data was quality checked with fastqc (Andrew, 2010). Data had adapters trimmed, cleaned of duplicates, scanned for bacterial and adapter contamination, and dropped low complexity reads using modified Perl scripts from Bi *et al.* (2012) and Singhal (2013) pipelines. The Phyluce pipeline (Faircloth, 2016) was followed to generate exon contigs. I used a 20× sequence coverage threshold, with coverage for each exon across all taxa set to 85%, or 15% missing data. Resulting in 3253 exons per individual (Appendix III).

Phylogenetics and Diversification

Heteronotia binoei phylogeny was constructed with 3253 concatenated exons using RAxML v8.2.8 (Stamatakis, 2014) to search for the best scoring ML tree and perform a rapid bootstrap analysis with 1,000 rapid bootstrap replicates (-x flag). Partitions were applied to each exon and a substitution model as GTR Γ. SNAPP (Bryant *et al.*, 2012) was implemented to construct a phylogeny using a Bayesian framework from 3176 SNPs using BEAST 2.5.1 (Bouckaert *et al.*, 2014). Clade delineations were defined from mitochondrial lineages. The analysis ran for 10,000,000 MCMC generations logging trees every 1000 generations under a GTR Γ+ I substitution model. 1,000 trees (10%) were dropped as burn-in.

Gene Flow

Structure

To infer gene flow, I used *Structure* v2.3.4 (Pritchard *et al.*, 2000; Pickrell & Pritchard, 2012) to conduct a cluster analysis. Individuals were partitioned by contact zone. *Structure* was run for 2,000,000 generations with three replicates. To combine and match clusters CLUMPP v1.1.2 was used on each contact zone's cluster replicate (Jakobsson & Rosenberg, 2007). CLUMPP was automated using Structure Harvester v0.6.94 (Earl, 2012).

To test for real admixture or error due to incomplete lineage sorting, I conducted an ABBA-BABA test using Patterson's D-statistic from the CalcD function in EvoBiR (Durand *et al.*, 2011; Eaton *et al.*, 2013; Blackmon & Adams, 2015). Only contact zone 1 and contact zone 3 have any observable introgression. Because of this, ABBA-BABA testing was omitted from contact zone 2. For both tests, *Heteronotia planiceps* exon data was used as the outgroup. Each contact zone was partitioned by lineages in contact then each individual was clustered into its lineage for EvoBiR, here delineated as P1, P2, P3, and Outgroup. For contact zone 1, *H. binoei* lineage clusters included were CYA6-N (P1), CYA6-S (P2), and EA6 (P3). Here I tested to see if admixture between CYA6-S and CYA6-N is accurate. Contact zone 3 clustering scheme had lineages NWQ (P1), SM6 (P2), and GULF (P3). Each topology in both contact zone schemes was verified from both mitochondrial (*ND*2) and nuclear (exon) phylogenies.

Principal Components

I conducted a principal component analysis (DAPC) to deduce population groupings from single nucleotide polymorphisms (SNPs) gathered from exon data. A principal component analysis (PCA) of genetic markers is useful in generating distinct population groupings since it makes no assumptions regarding Hardy Weinburg Equilibrium (HWE), diversity, linkage disequilibrium, or recombination. I ran a PCA using the R package *adegenet* (Jombart, 2008; Jombart & Ahmed, 2011).

Effective Estimation of Migration Surfaces

Estimation of migration and diversity between lineages of *H. binoei* was conducted using effective estimation of migration surfaces (EEMS) (Petkova *et al.*, 2016). Effective Estimation of Migration Surfaces (EEMS) is useful for visualizing migration and diversity along a Euclidean plane (here the geographic space of Queensland and Northern Territory,

Australia). Unlike traditional PCA and clustering approaches, EEMS explicitly represents genetic differentiation as a function of migration. EEMS uses a population genetic model involving migration on a unidirectional graph, $G = (V,E)$, where V is the vertices (demes) connecting edges (E) defined by polygons on a graph (G): the map defined here as the Australian Monsoonal Tropics. Two parameters are used $m = \{m_e: e \in E\}$, migration, and $q = \{q_v: v \in V\}$, diversity. m defines a migration estimation on each edge and diversity q is the estimation of genetic dissimilarity within each deme. These estimations are accomplished through a Bayesian framework using a likelihood to measure how well m and q explain the observed data. A *prior* describes the expectation of m and q. EEMS estimates migration and diversity on a map (Euclidian plane) to estimate the observed from the expected (*prior*). Expected migration and diversification are labeled as zero, while those below the expected model are negative and higher than expected are positive. Euclidean space was defined across all *H. binoei* lineages' ranges. Total demes were set to 1,000 tessellations. MCMC was run using 2,000,000 generations sampled every 1,000 generations with a 50% burning, and an MCMC thinner of 9999 was implemented across three chains.

Niche Models

I constructed a series of species distribution models using Maxent 3.41 (Phillips *et al.*, 2006; Phillips & Dudík, 2008; Phillips *et al.*, 2017) to estimate currently habitable niches among nine lineages of *H. binoei*. Locality information was collected from authors Fujita *et al.*, 2010; Moritz *et al.*, 2016; and (unpublished collection), from S. Zozoya (Supplementary Fig. 10). GIS layers were assembled in QGIS 3.2 (QGIS Development Team, 2009). We procured the base vector layer ('Australia administrative area') from the Global Administrative Database (GADM, 2017). Climate and environmental layers were taken from WorldClim 2 (Fick & Hijimans, 2017), current 1970-2000, at 30 arc-seconds (~1 km).

Landsat tree cover was parsed and stitched together to create a vegetative continuous field (VCF) in QGIS 3.2. The resolution was set to 30 m^2 per pixel. Vegetation was estimated as tree cover of horizontal wooded cover greater than 5 m in height. Map data was derived from Landsat-5 Thematic Mapper^{TM} and/or Landsat-t Enhanced Thematic Mapper Plus (ETM+) (Sexton *et al.*, 2013). Finally, a topographic raster group was derived from Shuttle Radar Topography Mission – February 2000 (SRTM) (USGS, 2006). The elevation layer was taken from SRTM at 30 m^2 resolution. Roughness, slope [using Horn & Goldburg (1995) algorithm (Fleming & Hoffer, 1979; and Ritter, 1987)], aspect, hillshade, terrain position index (TPI), and terrain ruggedness index (TRI) were derived from SRTM using R package raster 2.6-7 (Hijmans & Etten, 2012). All layers were trimmed and bound to the Australia monsoonal tropics in QGIS, then transformed to a raster in ASCII format for MaxEnt. Since locality cover for each species is high and broadly established within the putative range, no assumption was made for limiting presence data by arbitrary omission. Our niche model analysis was conducted with presence-only data (Fielding, 2002; Veloz, 2009) to predict regions where species distributions could extend without extraneous factors.

The final niche models were conducted on each species distribution independently to estimate possible range extension. 50 bootstrap replicates were performed on lineage groups with less than 50 recorded occurrences, while 30 cross-validation replicates were conducted on lineages with over 50 replicates, with each model using a random seed to estimate the robustness of prediction. A maximum number of background points were set to 10,000. Environmental variable importance used the Jackknife method in MaxEnt, and the output format was written in the Cloglog transformation.

Niche Overlap

To understand the ecological dynamics between *Heteronotia* lineages, I conducted a model analysis of the niche space overlap in NicheA (Qiao *et al.*, 2016). This analysis tested niche similarity between lineages to elucidate niche similarity. NicheA allows for visualization of Hutchinsonian duality (Pulliam, 2000; Colwell & Rangel, 2009) between environmental (E) and geographic (G) spaces across species distribution.

A background cloud (BC) was generated from a PCA of all raster groups independently (BioClim 2.0 layers, VCF layers, and topographic layers) from Australia monsoonal tropics, generating environmental spaces constrained to this region. Virtual species were generated from distributions of each lineage, independently, using the MVE (minimum volume ellipsoid) function in NicheA, creating ellipsoids of the fundamental niche of each species. Niche overlap was quantified between ellipsoids of *Heteronotia* lineages nearby or in contact. A virtual niche was simulated from the overlap quantification between convex polyhedrons (the E-space of each lineage the ellipsoids are derived from) for each overlapping lineage. Simulated niche space overlap was rasterized (.tiff) and overlaid Australia monsoonal tropics using QGIS 3.0 to visualize the model prediction for suitable shared niche space among species.

Results

Phylogenetic Analysis

Phylogenetic analysis (Fig. 12) is concordant with past mitochondrial analyses (Moritz *et al.*, 2016). Bootstrap support values are high, with most >95. Lineages cluster accordingly to their presumed lineage. Each lineage is monophyletic, holding up mitochondrial delineations.

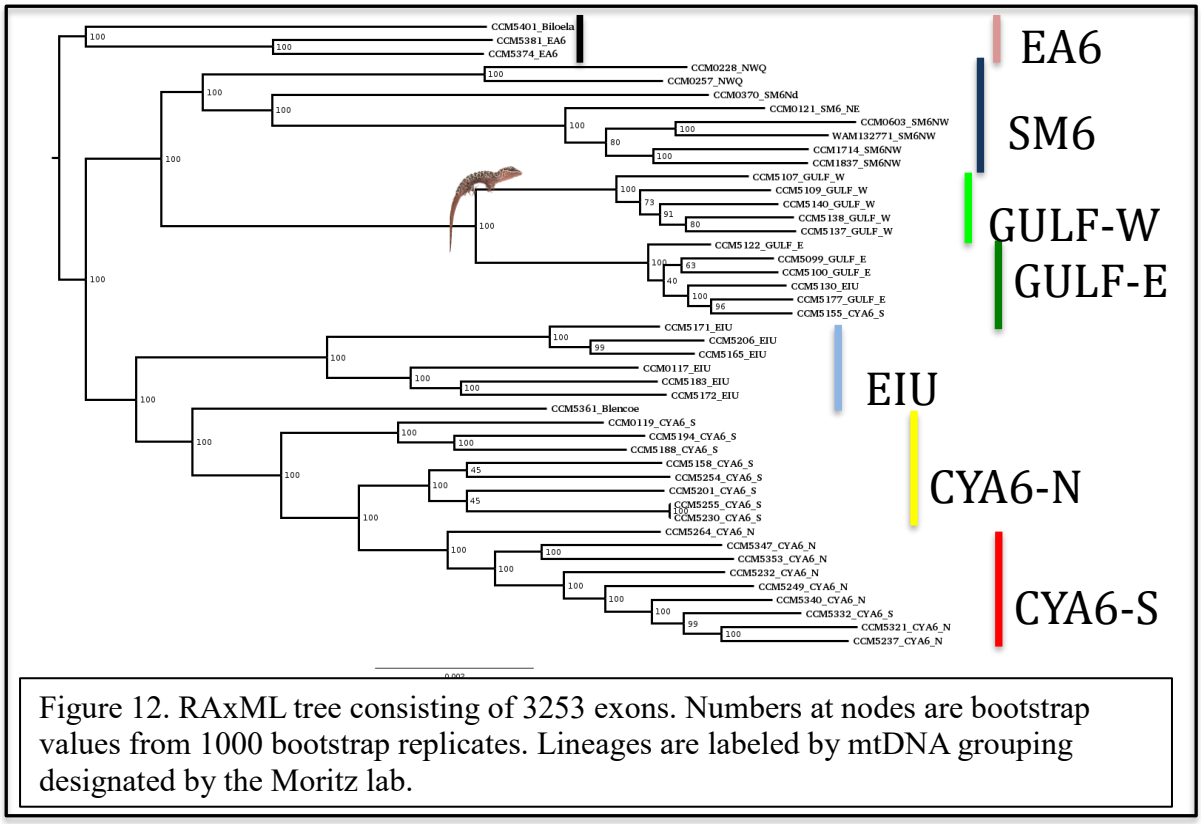


Figure 12. RAxML tree consisting of 3253 exons. Numbers at nodes are bootstrap values from 1000 bootstrap replicates. Lineages are labeled by mtDNA grouping designated by the Moritz lab.

See Supplementary Figure 7 to see threshold estimations to construct the best tree.

Bayesian analysis is concordant with maximum likelihood methods. DensiTree shows shared alleles between groups with two distinct radiations (Fig. 13). The monsoonal tropics have had expansions of *H. binoei* from two major biomes, the eastern mesic zone expanded into the east in Queensland, while the arid zone expanded into Northern Territory where they come into secondary contact with each other (Fujita *et al.*, 2010; Moritz *et al.*, 2016).

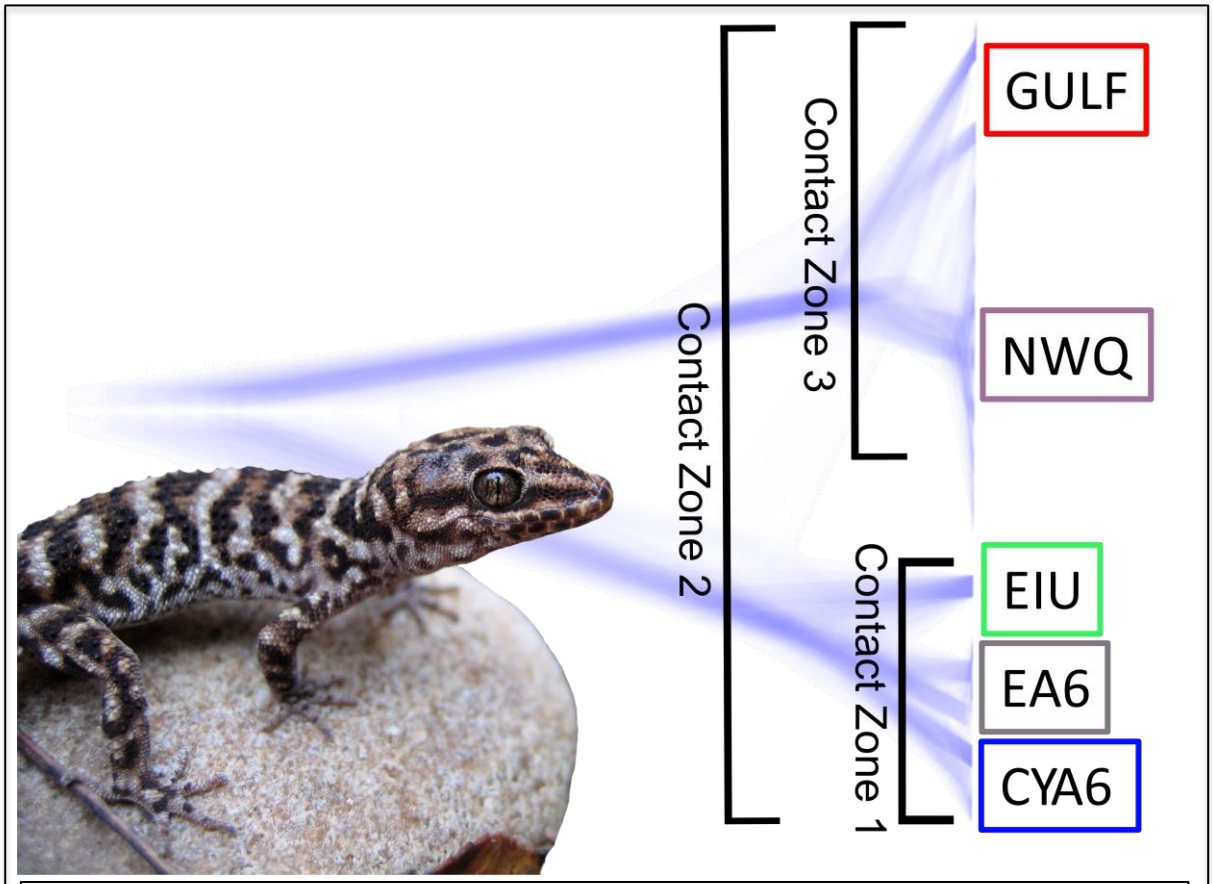


Figure 13. SNAPP tree consisting of 3176 SNPs. Individuals set to lineage. Contact zones label by bifurcation in the tree.

Gene Flow

The principal component analysis shows clustering within the major lineages of this study (Fig. 14). The proximity of groups to each other is determined by genetic similarity. Lineage clusters that are closer to each other show less genetic dissimilarity. The EA6 lineage is not genetically differentiated from CYA6. This result is expected given genetic studies using *Heteronotia* (Moritz, 1983; 1984; 1991; Moritz *et al.*, 1989; Moritz & Heideman, 1993; Fujita *et al.*, 2010; Pepper *et al.*, 2013; Moritz *et al.*, 2016). As the system continues to garner knowledge and data, the full family tree is becoming more fully understood. EIU lineage is not yet published in the literature. This study has EIU nested outside the CYA6 lineages and

EA6. PCA shows the same. Here the GULF lineages and the NWQ groups are not as genetically similar compared to some phylogenies generated from Fujita *et al.* (2010) or Moritz *et al.* (2016).

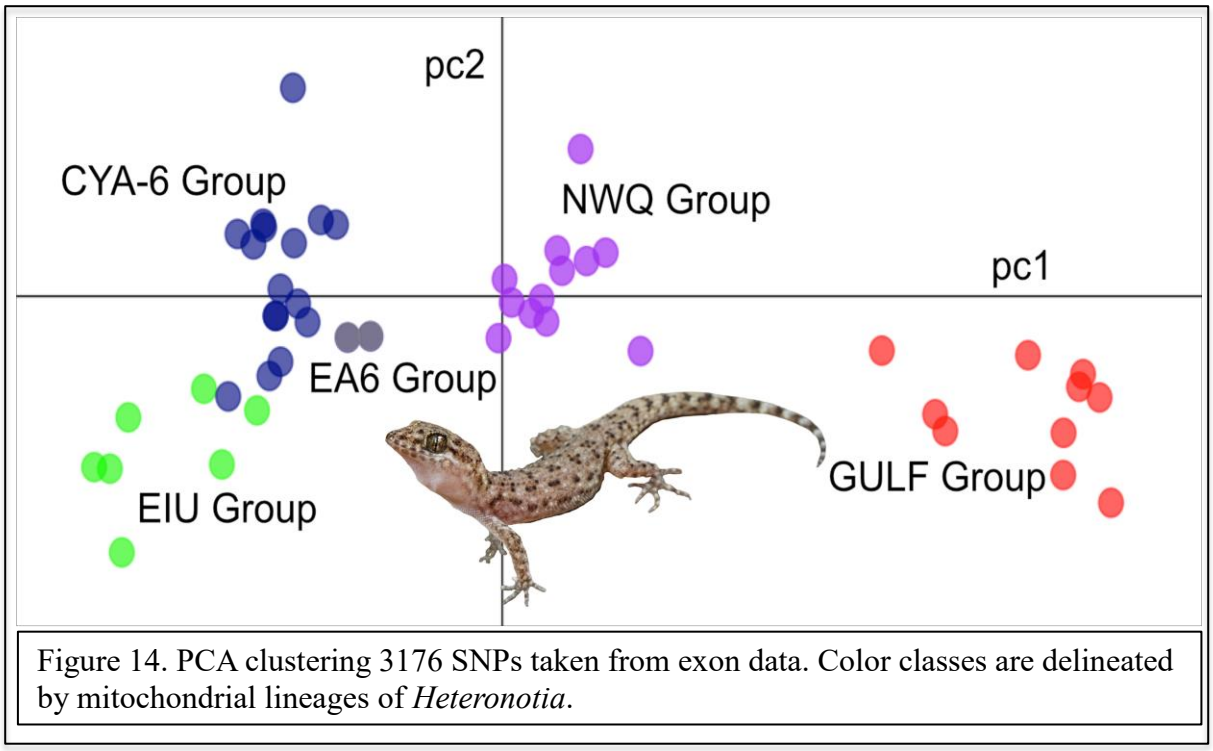


Figure 14. PCA clustering 3176 SNPs taken from exon data. Color classes are delineated by mitochondrial lineages of *Heteronotia*.

Each contact zone is at varying degrees of divergence, which corresponds with introgression (Fig. 15). Contact zone 1 is the least diverse group of lineages and have the most introgression. Contact zone 2 has overlapping ranges but has the deepest divergence between lineages without any introgression. Contact zone 3 is the oldest radiation in this study, moving west in the Northern Territory. This group does not overlap with sister lineages. There is potential introgression with NWQ and GULF lineages; however, this may be ancestral alleles and not true gene flow between lineages.

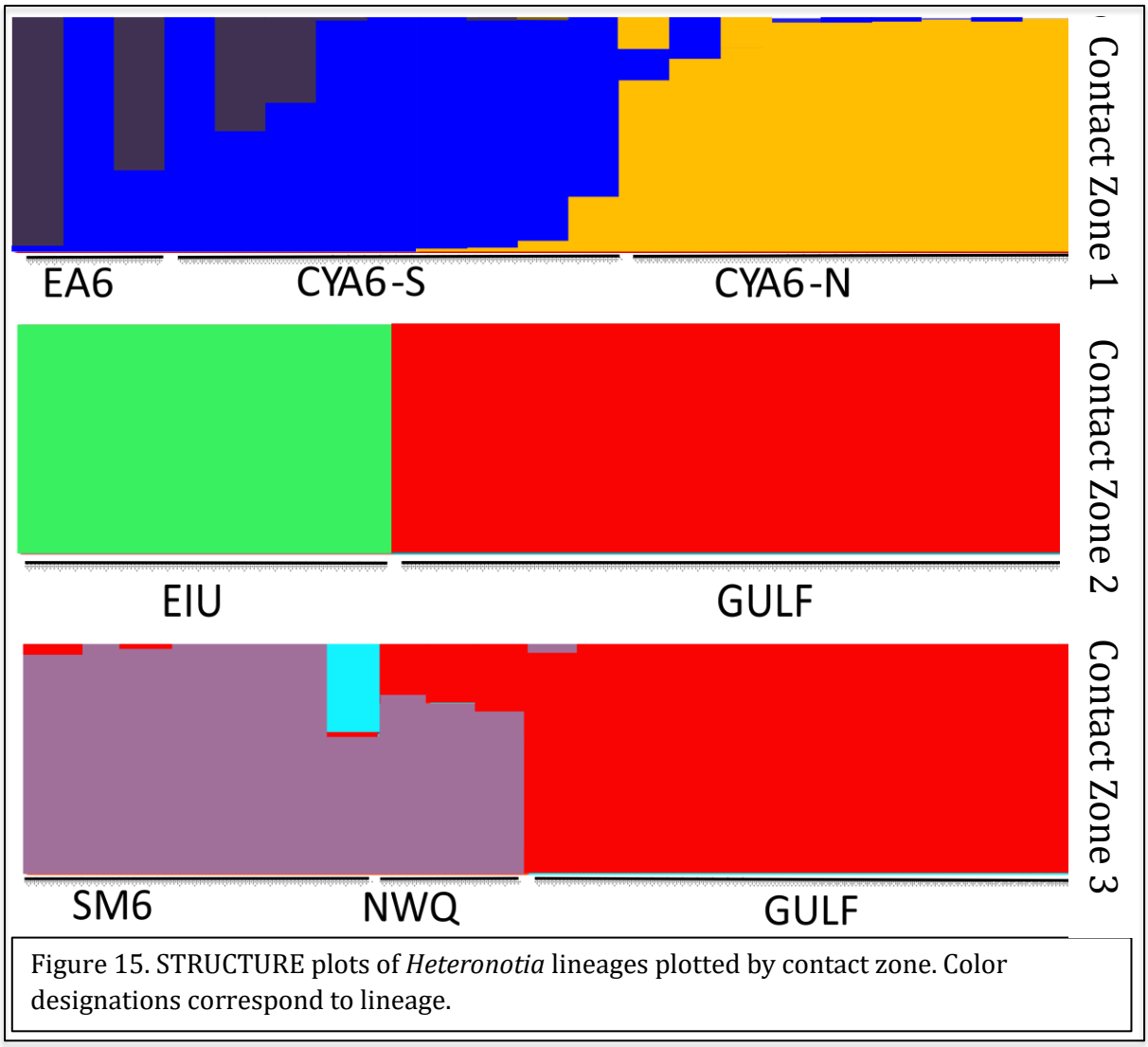
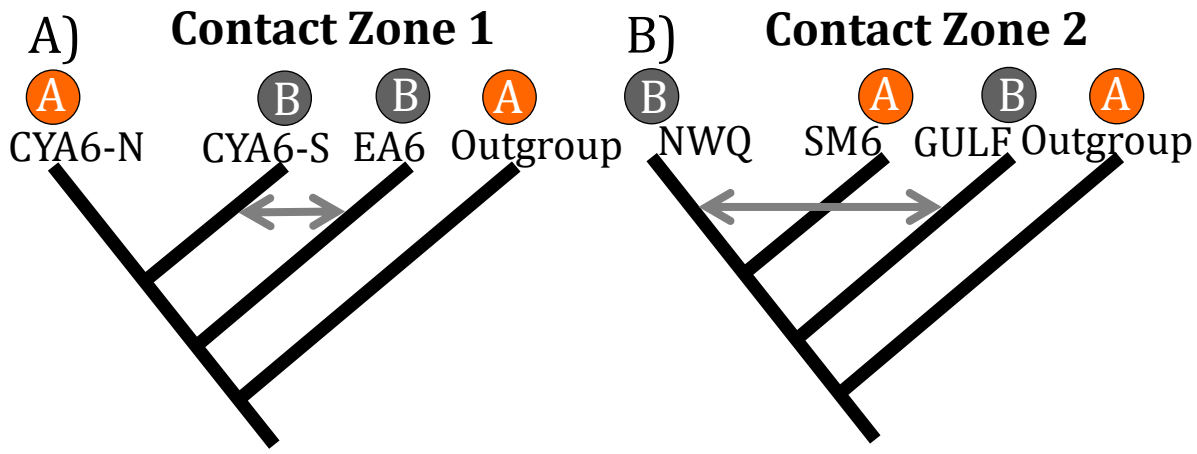


Figure 15. STRUCTURE plots of *Heteronotia* lineages plotted by contact zone. Color designations correspond to lineage.

Incomplete lineage sorting can give a Type I error in clustering analyses like *Structure*. The ABBA-BABA test in contact zone 1 has a positive Patterson's D statistic that is significant (Fig. 16A). Here the *null* is accepted, and introgression between recently diverged lineages is true. For contact zone 3, a significant negative Patterson's D statistic rejects the *null* and introgression is likely false and, thus, a victim of Type I error (Fig. 16B).



	Contact Zone 1	Contact Zone 2	Contact Zone 3
RAW Patterson's D	0.7566138	N/A	-0.1192661
SD D Statisic	0.004286304	N/A	0.007930389

Figure 16. ABBA-BABA and Patterson's D statistic used to test if introgression truly occurs, or if it is just a statistical error driven by incomplete lineage sorting. A) Contact zone 1 on the east side of Queensland. B) Contact zone 3 toward he southwester portion of Queensland.

Gene flow dynamics between *H. binoei* lineages are linked to both varying rates of diversity and the environment. Estimating effective migration surfaces illustrates lower than expected migration of alleles where the GULF lineage overlaps in contact zone 2 and 3 (Fig. 17). This result reinforces the *Structure* and ABBA-BABA tests. Diversity is low within the lineage but high in contact zones, which is expected, since contact zones are a source for diversity with multiple lineages overlapping. Some areas with low numbers of individuals within lineages have low diversity like in the EA6 lineage. Migration is slightly higher than expected (q = 0.5) in the CYA6 complex even across a ~600 km transect. However, diversity is lower than expected in the CYA6-N lineage, potentially representing a current bottleneck.

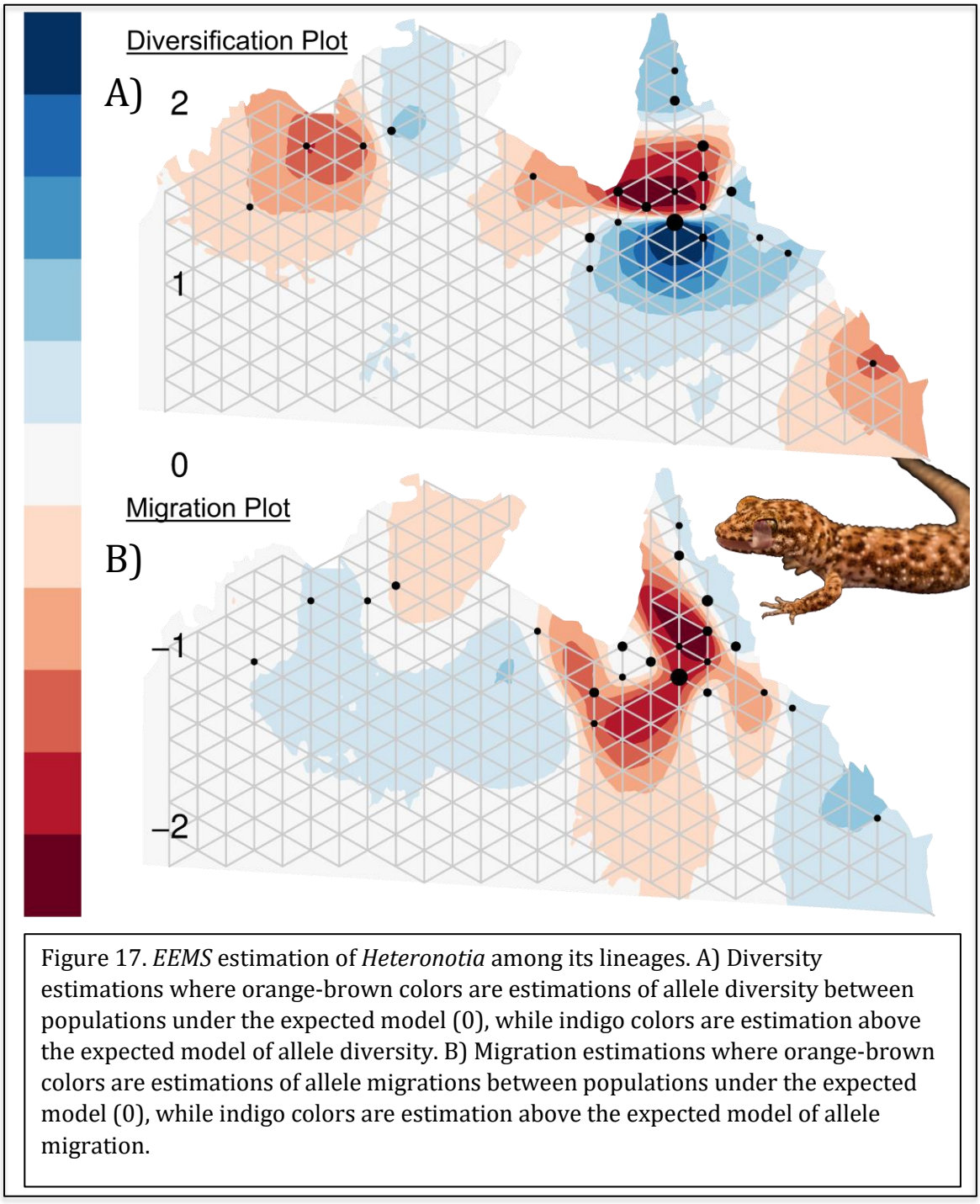
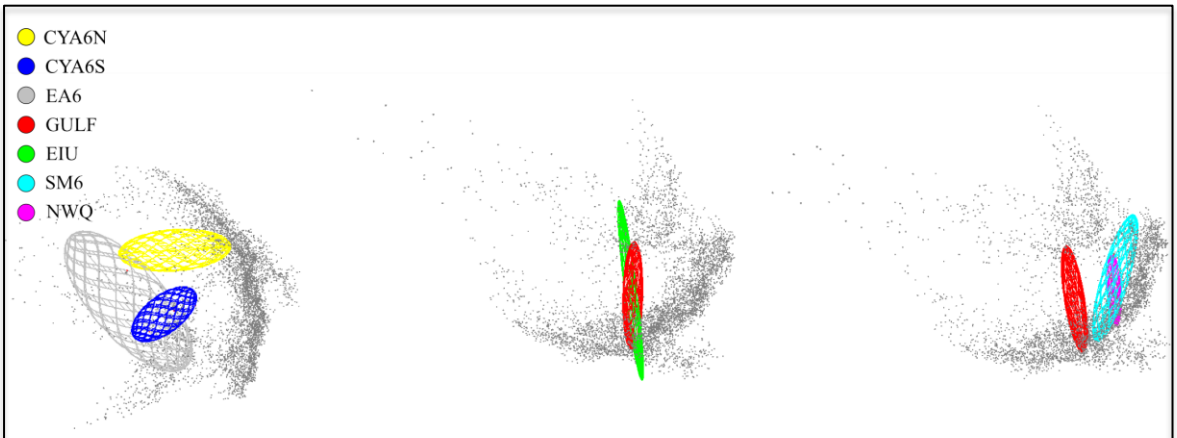


Figure 17. *EEMS* estimation of *Heteronotia* among its lineages. A) Diversity estimations where orange-brown colors are estimations of allele diversity between populations under the expected model (0), while indigo colors are estimation above the expected model of allele diversity. B) Migration estimations where orange-brown colors are estimations of allele migrations between populations under the expected model (0), while indigo colors are estimation above the expected model of allele migration.

Niche Barrier and Overlap

NicheA estimates of overlap illustrate a varied habitat and niche space across *H. binoei*'s distribution (Fig. 18). Contact zone 1 has limited overlap between the EA6 group

and the CYA6 lineages. Oddly, CYA6-N and CYA6-S occupy a different effective niche. This is telling, as there is a barrier to gene flow. Contact zone 2 is interesting, as these two lineages are found nearby, only meters apart, although niche overlap is small and only at the margins. Contact zone 3 has the GULF lineage not overlapping either the NWQ lineage or the SM6 lineage. GULF only comes in overlapping range after it penetrates the Selwyn Range. The GULF lineage occupies Queensland, while NWQ and SM6 are western ranges in Northern Territory.



Contact Zone 1 | Contact Zone 2 | Contact Zone 3

	Niche Overlap	Precision		Niche Overlap	Precision		Niche Overlap	Precision
CYA6S VS CYA6N	0.0	0.0092	GULF VS EIU	0.0752	0.0038	GULF VS NWQ	0.0	0.0013
CYA6S VS EA6	0.0077	0.0092				GULF VS SM6	0.0	0.0
CYA6N VS EA6	0.2534	0.0686				SM6 VS NWQ	0.1142	0.0013

Figure 18. *NicheA* overlap estimation. Background points are variables from clipped Australia raster layers. Each minimum volume ellipsoid (MVE) is represents the realized niche space occupied by a *H. binoei* lineage, differentiated by color. Niche overlap is the quantification of hypervolume overlap between two ellipsoids.

Discussion

Using the *Heteronotia binoei* complex, I identified geographic areas of potential gene flow to understand the mechanisms and processes that either facilitate or hamper speciation. Individuals were sampled across three explicit contact zones at varying degrees of divergence in Queensland, Australia. Lineages exhibit higher gene flow with more closely-related

lineages, while those more distantly related *i.e.*, divergent, lineages show limited or no gene flow in the case of contact zone 2.

Contact zone 1 are all sister taxa and are likely to be diversifying in parapatry through isolation and distance. Allopatric drivers are likely influencing isolation, especially since geneflow is present. One barrier right at CYA6-N and CYA6-S boundary is the Burdekin Gap (Keast, 1961; Galbraith, 1969). The Burdekin Gap is a lowland barrier for dry country habitats with more wet-mesic/rain forest habitats. CYA6-N has an enclave where there is a lack of diversity compared to individuals north and south within the lineage. This area of low diversity occurs at a specific portion of the Great Dividing Range called the Einasleigh Uplands (sometimes referred to as the Burdekin Lynd Divide) (Keast, 1961; Schodde & Mason, 1999). It is a transition zone between the Cape York lowland and the central Queensland mesic savannahs. The lack of diversity within CYA6-N individuals here is likely due to constituents now able to pass through this barrier easily and may have been sequestered here as it shrank during the Pleistocene. It may also be a founder's effect within the population. Being that the Burdekin Gap is a transition zone between dry and wet habitats, local adaption to these changes is at play. Niche overlap is nonexistent and driving diversity even with gene flow. There is a high probability that this is an early phase or a potential speciation event. Oddly enough, CYA6-N and EA6 have relatively high overlap because of the BioClim layer – "Mean Temperature of Driest Quarter." Being that both EA6's and CYA6-N's ranges are wetter; temperature is likely more constant during drier months than CYA6-S's range.

Contact zone 2 has two lineages from different evolutionary radiations coming together in secondary contact. These two lineages are easily found within meters of one

another. Fundamental niche overlap is relatively high given their respective MVE sizes. EIU is found mostly in rocks in the savannah woodland, while GULF lineages are grassland specialists. The major exception of EIU found outside of rocky areas is those found in lancewood (*Acacia shirleyi*) woodlands. They are micro-sympatric with GULF [GULF-E] residing in tree hollows and log piles. These two lineages do not exhibit any gene flow. Thus, there is another response driving genetic isolation between these lineages.

Contact zone 3 is an enigma within this study. SM6 occurs sympatrically with GULF and NWQ. This group has the same evolutionary radiation and is moderately derived since these lineages are not sister. Depending on what study has generated phylogenies, phylogenetic placement changes, depending on genetic data used and phylogenetic method. In this study, there are known missing lineages. The sampling is also low in this study compared to other contact zones. Moritz *et al.* (2016) had the most comprehensive sampling to date, using a multilocus approach. The mitochondrial phylogeny is discordant from the nuclear species tree, as is the mitochondrial and nuclear species tree from Fujita *et al.* (2010). This study utilizes genomic scale data but does not include many of the other lineages to the west. It is also inconclusive whether gene flow does occur. With ABBA-BABA tests uncovered signals of introgression as likely Type I error from incomplete lineage sorting. There are no defined allopatric barriers. Sampling of these lineages is along the Selwyn Range. SM6 and NWQ share a high proportion of their fundamental niche given the niche overlap quantifications. SM6 is a large-bodied generalist that can be rock-dwelling; however, it is typically replaced by other *Heteronotia species – H. fasciolatus, H. spelea,* and *H. planiceps* (Pepper *et al.*, 2013) on rocky outcrops, if present. This replacement also includes other *H. binoei* lineages *e.g.*, CA6, CQ, and including NWQ. NWQ is a rock-dweller, although it is occasionally observed in adjacent woodlands. Still, NWQ mostly inhabits semi-

arid areas. This habitat separation is the only inference which allows for diversification with the data presented in this study.

Chapter 4

Conclusion: Contrasting Diversifications of Two Australian Geckos Under Australia's Dynamic History

In this dissertation, two geckos were used to understand the processes of evolution in Australia's dynamic geological history. First, *Lialis burtonis* has limited gene flow between groups, only occurring when individuals expand out of their range. Populations of *L. burtonis* appear to associate with particular ecoregions in Australia driving its isolation. Migration is high within populations but not between populations. Consequently, this allows for populations to have large ranges and maintain cohesion within. From cohesion observed in Chapter 2, I hypothesize this phenomenon is due to *L. burtonis* being an active and highly mobile predator, facilitating more extensive home ranges and more chances to reproduce with individuals from greater distances away.

Heteronotia binoei, by contrast, is hyperdiverse across its range, with over 60 mitochondrial populations discovered and more found every field season. *H. binoei* has limits to its range since it is vulnerable outside of its shelter. Bustard (1968) found through a mark and recapture study that *H. binoei* is limited to around a 10 m home range. A maximum distance to a *Heteronotia* gecko ventured was 60 m. Over time, *Heteronotia binoei* does expand at the species scale, just not at the individual scale. However, as lineages expand, isolation begins among independent populations allowing for lineages to diverge independently and erect genetically distinct demes.

Australia's climatic and geologic history, after it split from Gondwana, has influenced the evolutionary history of its flora and fauna. Both *Lialis* and *Heteronotias'* origins diversified in isolation throughout the changing landscape and habitat of Australia. Both

lineages experienced the same biogeographic variables but have entirely different evolutionary histories.

Varying life histories between the species is presumably the root cause for incredibly different demographies. *Lialis* is, comparatively, a large, legless gecko and an active predator of secondary consumers (small squamates). *Heteronotia* does not cover relatively large distances to forage. Being insectivorous, they are confined to home ranges as secondary consumers. Their contrasting life histories are most parsimonious driver for their evolutionary histories being discordant under the same extrinsic variables.

References

Abdrakhmatov, K. Y., Aldazhanov, S. A., Hager, B. H., Hamburger, M. W., Herring, T. A., Kalabaev, K. B., ... & Reilinger, R. E. (1996). Relatively recent construction of the Tien Shan inferred from GPS measurements of present-day crustal deformation rates. *Nature*, *384*(6608), 450.

Aho, K., Derryberry, D., & Peterson, T. (2014). Model selection for ecologists: the worldviews of AIC and BIC. *Ecology*, *95*(3), 631-636.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199-213). Springer, New York, NY.

Ackerly, D. D. (2003). Community assembly, niche conservatism, and adaptive evolution in changing environments. *International Journal of Plant Sciences*, *164*(S3), S165-S184.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*.

Anderson, R. P., & Raza, A. (2010). The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela. *Journal of Biogeography*, *37*(7), 1378-1393.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Avise, J. C. (2000). *Phylogeography: the history and formation of species*. Harvard university press.

Bandelt et al, 1999: H. -J. Bandelt, P. Forster, and A. Röhl. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution, 16:37–48, 1999.

Barton, N. H. (1999). Clines in polygenic traits. *Genetics Research*, *74*(3), 223-236.

Beard, J. S. (1979). Phytogeographic regions. *Western landscapes*, 107-121.

Beolens, B., Watkins, M., & Grayson, M. (2011). *The eponym dictionary of reptiles*. JHU Press.

Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C., & Good, J. M. (2012). Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC genomics*, *13*(1), 403.

Blackmon, H., & Adams, R. (2015). EvobiR: tools for comparative analyses and teaching evolutionary biology.

Boden, R., & Given, D. R. (1995). Regional overview: Australia and New Zealand. *Centres of plant diversity: a guide and strategy for their conservation*, *2*, 433-518.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., ... & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology*, *10*(4), e1003537.

Bragg, J. G., Potter, S., Bi, K., & Moritz, C. (2016). Exon capture phylogenomics: efficacy across scales of divergence. *Molecular Ecology Resources*, *16*(5), 1059-1068.

Brennan, I. G., & Keogh, J. S. (2018). Miocene biome turnover drove conservative body size evolution across Australian vertebrates. *Proceedings of the Royal Society B*, *285*(1889), 20181474.

Brennan, I. G., & Oliver, P. M. (2017). Mass turnover and recovery dynamics of a diverse Australian continental radiation. *Evolution*, *71*(5), 1352-1365.

Bridgewater, P. B. (1987). The present Australian environment–terrestrial and freshwater. *Fauna of Australia*, *1*, 69-100.Bryant and Moulton 2004: D. Bryant and V. Moulton. Neighbor-net: An agglomerative method for the construction of phylogenetic networks. Molecular Biology and Evolution, 21(2):255– 265, 2004.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular biology and evolution*, *29*(8), 1917-1932.

Byrne, M., Yeates, D. K., Joseph, L., Kearney, M., Bowler, J., Williams, M. A. J., ... & Melville, J. (2008). Abstract. *Molecular Ecology*, *17*(20), 4398-4417.

Byrne, M., Steane, D. A., Joseph, L., Yeates, D. K., Jordan, G. J., Crayn, D., ... & Keogh, J. S. (2011). Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. *Journal of Biogeography*, *38*(9), 1635-1656.

Bustard, H. R. (1968). The ecology of the Australian gecko Heteronotia binoei in northern New South Wales. *Journal of Zoology*, *156*(4), 483-497.

Butlin, R. K., Galindo, J., & Grahame, J. W. (2008). Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1506), 2997-3007.

Camargo, A., Sinervo, B., & Sites Jr, J. W. (2010). Lizards as model organisms for linking phylogeographic and speciation studies. *Molecular Ecology*, *19*(16), 3250-3270.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, genomes, genetics*, *1*(3), 171-182.

Catullo, R. A., Lanfear, R., Doughty, P., & Keogh, J. S. (2014). The biogeographical boundaries of northern Australia: Evidence from ecological niche models and a multi-locus phylogeny of Uperoleia toadlets (Anura: Myobatrachidae). *Journal of Biogeography*, *41*(4), 659-672.

Chambers, L. E., Barnard, P., Poloczanska, E. S., Hobday, A. J., Keatley, M. R., Allsopp, N., & Underhill, L. G. (2017). Southern hemisphere biodiversity and global change: data gaps and strategies. *Austral ecology*, *42*(1), 20-30.

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, *30*(23), 3317-3324.

Clapperton, C. M. (1990). Quaternary glaciations in the Southern Hemisphere: an overview. *Quaternary Science Reviews*, *9*(2-3), 299-304.

Cogger, H. G. (1981). The Australian reptiles: origins, biogeography, distribution patterns and island evolution. *Ecological biogeography of Australia*.

Cogger, H. G. (2000). Reptiles and Amphibians of Australia. Reed New Holland, Sydney. *New South Wales, Australia*.

Colwell, R. K., & Rangel, T. F. (2009). Hutchinson's duality: the once and future niche. *Proceedings of the National Academy of Sciences*, *106*(Supplement 2), 19651-19658.

Cowling, R. M., Procheș, Ș., and Partridge, T. C.. 2009. Explaining the uniqueness of the Cape Flors: incorporating geomorphic evolution as a factor for explaining its diversification. *Mol. Phylogenet. Evol.* **51**: 64– 74.

Coyne, J. A., & Orr, H. A. (2004). Speciation. Sinauer. *Sunderland, MA*.

Crisp, M., Cook, L., & Steane, D. (2004). Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present–day communities?. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1450), 1551-1571.

Crisp, M. D., Arroyo, M. T., Cook, L. G., Gandolfo, M. A., Jordan, G. J., McGlone, M. S., ... & Linder, H. P. (2009). Phylogenetic biome conservatism on a global scale. *Nature*, *458*(7239), 754.

Cronin, L., & Oram, N. (2001). *Australian reptiles and amphibians*. Envirobook.

Daza, J. D., Alifanov, V. R., & Bauer, A. M. (2012). A redescription and phylogenetic reinterpretation of the fossil lizard Hoburogekko suchanovi Alifanov, 1989 (Squamata, Gekkota), from the Early Cretaceous of Mongolia. *Journal of Vertebrate Paleontology*, *32*(6), 1303-1312.

Daza, J. D., & Bauer, A. M. (2012). A new amber-embedded sphaerodactyl gecko from Hispaniola, with comments on morphological synapomorphies of the Sphaerodactylidae. *Breviora*, 1-28.

Daza, J. D., Bauer, A. M., & Snively, E. D. (2014). On the fossil record of the Gekkota. *The Anatomical Record*, *297*(3), 433-462.

Desjardins-Proulx, P., & Gravel, D. (2012). A complex speciation–richness relationship in a simple neutral model. *Ecology and evolution*, *2*(8), 1781-1790.

Dieckmann, U., & Doebeli, M. (1999). On the origin of species by sympatric speciation. *Nature*, *400*(6742), 354.

Dress & Huson. 2004: A.W.M. Dress and D.H. Huson, "Constructing splits graphs," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 1, no. 3, pp. 109-115, July-Sept. 2004.

Drummond, A. J., Ho, S. Y., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS biology*, *4*(5), e88.

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, *7*(1), 214.

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, *28*(8), 2239-2252.

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, *4*(2), 359-361.

Eaton, D. A., & Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic biology*, *62*(5), 689-706.

Eaton, D. A. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844-1849.

Eaton, D. A. R. (2018). iPyRAD: assembly of de novoRADseq loci for phylogenetic analyses. https://ipyrad.readthedocs.io/index.html

Eldridge, M. D., Potter, S., Johnson, C. N., & Ritchie, E. G. (2014). Differing impact of a major biogeographic barrier on genetic structure in two large kangaroos from the monsoon tropics of Northern Australia. *Ecology and Evolution*, *4*(5), 554-567.

Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., & Holzapfel, C. M. (2010). Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the national academy of sciences*, *107*(37), 16196-16200.

Endler, J. A. (1977). *Geographic variation, speciation, and clines* (No. 10). Princeton University Press.

Etter, P. D., Preston, J. L., Bassham, S., Cresko, W. A., & Johnson, E. A. (2011). Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PloS one*, *6*(4), e18561.

Evans, S. E. (2003). At the feet of the dinosaurs: the early history and radiation of lizards. *Biological Reviews*, *78*(4), 513-551.

Faircloth, B. C. (2015). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, *32*(5), 786-788.

Felsenstein, J. (1981). Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution*, *35*(1), 124-138.

Fielding, A. H. (2002). What are the appropriate characteristics of an accuracy measure? *Predicting species occurrences: issues of accuracy and scale*, 271-280.

Fisher, R. A. (1950). Gene frequencies in a cline determined by selection and diffusion. *Biometrics*, *6*(4), 353-361.

Fitzpatrick, B. M., Fordyce, J. A., & Gavrilets, S. (2008). What, if anything, is sympatric speciation?. *Journal of evolutionary biology*, *21*(6), 1452-1459.

Fitzpatrick, B. M., Fordyce, J. A., & Gavrilets, S. (2009). Pattern, process and geographic modes of speciation. *Journal of evolutionary biology*, *22*(11), 2342-2347.

Fleming, M. D., & Hoffer, R. M. (1979, January). Machine processing of Landsat MSS data and DMA topographic data for forest cover type mapping. In *LARS Symposia* (p. 302).

Ford, J. (1986). Avian hybridization and allopatry in the region of the Einasleigh Uplands and Burdekin-Lynd Divide, north-eastern Queensland. *Emu-Austral Ornithology*, *86*(2), 87-110.

Ford, J. (1987). Hybrid zones in Australian birds. *Emu-Austral Ornithology*, *87*(3), 158-178.

Galbraith, I. C. J. (1969). The Papuan and Little Cuckoo-Shrikes, Coracina papuensis and robusta, as races of a single species (Results of the Harold Hall Australian Expedition, No. 10). *Emu*, *69*(1), 9-29.

Gavrilets, S., & Hastings, A. (1996). Founder effect speciation: a theoretical reassessment. *The American Naturalist*, *147*(3), 466-491.

Gavrilets, S., Hai, L., & Vose, M. D. (1998). Rapid parapatric speciation on holey adaptive landscapes. *Proceedings of the Royal Society of London B: Biological Sciences*, *265*(1405), 1483-1489.

Gavrilets, S., Acton, R., & Gravner, J. (2000). Dynamics of speciation and diversification in a metapopulation. *Evolution*, *54*(5), 1493-1501.

Gavrilets, S., Li, H., & Vose, M. D. (2000). Patterns of parapatric speciation. *Evolution*, *54*(4), 1126-1134.

Gavrilets, S. (2003). Perspective: models of speciation: what have we learned in 40 years?. *Evolution*, *57*(10), 2197-2215.

Gavrilets, S. (2014). Models of speciation: Where are we now?. *Journal of heredity*, *105*(S1), 743-755.

Haldane, J. B. S. (1948). The theory of a cline. *Journal of genetics*, *48*(3), 277-284.

Hall, R. (2009). Southeast Asia's changing palaeogeography. *Blumea-Biodiversity, Evolution and Biogeography of Plants*, *54*(1-2), 148-161.

Halliday, T., & Adler, K. (2002). *new encyclopedia of reptiles and amphibians*. Oxford University Press.

Hamming 1950: Hamming, Richard W. "Error detecting and error correcting codes". Bell System Technical Journal. 29 (2): 147–160. MR 0035935, 1950.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1442), 183-195.

Hijmans, R. J., Cameron, S. E., & Parra, J. L. (2006). Worldclim global climate layers Version 1.4. *available from WorldClim database: http://www. worldclim. org [Verified July 2008]*.

Hijmans, R. J., & van Etten, J. (2012). raster: Geographic analysis and modeling with raster data. R package version 2.0–12.

Horn, J., & Goldberg, D. E. (1995). Genetic algorithm difficulty and the modality of fitness landscapes. In *Foundations of genetic algorithms* (Vol. 3, pp. 243-269). Elsevier.

Hoelzer, G. A., Drewes, R., Meier, J., & Doursat, R. (2008). Isolation-by-distance and outbreeding depression are sufficient to drive parapatric speciation in the absence of environmental influences. *PLoS Computational Biology*, *4*(7), e1000126.

Hubbell, S. P. (2005). Neutral theory in community ecology and the hypothesis of functional equivalence. *Functional ecology*, *19*(1), 166-172.

Huson, D. H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics (Oxford, England)*, *14*(1), 68-73.

Huson and Bryant 2006: D.H. Huson and D. Bryant. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution, 23:254–267, 2006.

Hutchinson, MN. 1997. The first fossil pygopod (Squamata, Gekkota), and a review of mandibular variation in living species. Memoirs of the Queensland Museum 41: 355–366.

Irwin, D. E. (2012a). A novel approach for finding ring species: look for barriers rather than rings. *BMC biology*, *10*(1), 21.

Irwin, D. E. (2012b). Local adaptation along smooth ecological gradients causes phylogeographic breaks and phenotypic clustering. *The American Naturalist*, *180*(1), 35-49.

Iturralde-Vinent, M. A., & MacPhee, R. D. E. (1996). Age and paleogeographical origin of Dominican amber. *Science*, *273*(5283), 1850-1852.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801-1806.

Jennings, W. B., Pianka, E. R., & Donnellan, S. (2003). Systematics of the lizard family Pygopodidae with implications for the diversification of Australian temperate biotas. *Systematic Biology*, *52*(6), 757-780.

Johansen, T. (2012). A Field Guide to the Geckos of Northern Territory. *AuthorHouse*

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403-1405.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070-3071.

Joseph, L., Edwards, S. V., & McLean, A. J. (2013). The Maluridae: inferring avian biology and evolutionary history from DNA sequences. *Emu-Austral Ornithology*, *113*(3), 195-207.

Kearney, M., & Shine, R. (2004). Morphological and physiological correlates of hybrid parthenogenesis. *The American Naturalist*, *164*(6), 803-813.

Kearns, A. M., Restani, M., Szabo, I., Schrøder-Nielsen, A., Kim, J. A., Richardson, H. M., ... & Omland, K. E. (2018). Genomic evidence of speciation reversal in ravens. *Nature communications*, *9*(1), 906.

Keast, A. (1961). Bird speciation on the Australian continent. *Bulletin of the Museum of Comparative Zoology, Harvard*, *123*, 403-495.

Keinan, A., Mullikin, J. C., Patterson, N., & Reich, D. (2007). Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature genetics*, *39*(10), 1251.

Kirkpatrick, M., & Barton, N. H. (1997). Evolution of a species' range. *The American Naturalist*, *150*(1), 1-23.

Kluge, AG. 1995. Cladistic relationships of sphaerodactyl lizards. *American Museum Novitates* 3139: 1–23.

Kondrashov, A. S., & Kondrashov, F. A. (1999). Interactions among quantitative traits in the course of sympatric speciation. *Nature*, *400*(6742), 351.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*(3), 772-773.

Lee, M. S. Y., Oliver, P. M., & Hutchinson, M. N. (2009). Phylogenetic uncertainty and molecular clock calibrations: a case study of legless lizards (Pygopodidae, Gekkota). *Molecular Phylogenetics and Evolution*, *50*(3), 661-666.

Levins, R. (1968). *Evolution in changing environments: some theoretical explorations* (No. 2). Princeton University Press.

Levins, R. (1968). *Evolution in changing environments: some theoretical explorations* (No. 2). Princeton University Press.

Mabbutt, J. A. (1988). Australian desert landscapes. *GeoJournal*, *16*(4), 355-369.

MacDonald, J. D. (1969). Notes on the Taxonomy of Neositta: Results of the Harold Hall Australian Expedition, No. 18. The previous number in this series appeared in Emu 69: 110–111. *Emu-Austral Ornithology*, *69*(3), 169-174.

Macey, J. R., Wang, Y., Ananjeva, N. B., Larson, A., & Papenfuss, T. J. (1999). Vicariant patterns of fragmentation among gekkonid lizards of the genus Teratoscincus produced by the Indian collision: a molecular phylogenetic perspective and an area cladogram for Central Asia. *Molecular Phylogenetics and Evolution*, *12*(3), 320-332.

Mallet, J., Meyer, A., Nosil, P., & FEDER, J. L. (2009). Space, sympatry and speciation. *Journal of evolutionary biology*, *22*(11), 2332-2341.

Markgraf, V., McGlone, M., & Hope, G. (1995). Neogene paleoenvironmental and paleoclimatic change in southern temperate ecosystems—a southern perspective. *Trends in Ecology & Evolution*, *10*(4), 143-147.

Martin, H. A. (2006). Cenozoic climatic change and the development of the arid vegetation in Australia. *Journal of Arid Environments*, *66*(3), 533-563.

Martins, A. B., de Aguiar, M. A., & Bar-Yam, Y. (2013). Evolution and stability of ring species. *Proceedings of the National Academy of Sciences*, *110*(13), 5080-5084.

Mayr, E. (1963). Animal Species and. *Evolution. Belknap Press, Cambridge Massacusetts*.

McLaren, S., & Wallace, M. W. (2010). Plio-Pleistocene climate change and the onset of aridity in southeastern Australia. *Global and Planetary Change*, *71*(1-2), 55-72.

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, *2010*(6), pdb-prot5448.

Monahan, W. B., Pereira, R. J., & Wake, D. B. (2012). Ring distributions leading to species formation: a global topographic analysis of geographic barriers associated with ring species. *BMC biology*, *10*(1), 20.

Moritz, C. (1983). Parthenogenesis in the endemic Australian lizard Heteronotia binoei (Gekkonidae). *Science*, *220*(4598), 735-737.

Moritz, C. (1984). The origin and evolution of parthenogenesis in Heteronotia binoei (Gekkonidae). *Chromosoma*, *89*(2), 151-162.

Moritz, C., Donnellan, S., Adams, M., & Baverstock, P. R. (1989). The origin and evolution of parthenogenesis in Heteronotia binoei (Gekkonidae): extensive genotypic diversity among parthenogens. *Evolution*, *43*(5), 994-1003.

Moritz, C. (1991). The origin and evolution of parthenogenesis in Heteronotia binoei (Gekkonidae): evidence for recent and localized origins of widespread clones. *Genetics*, *129*(1), 211-219.

Moritz, C., & Heideman, A. (1993). The origin and evolution of parthenogenesis in Heteronotia binoei (Gekkonidae): reciprocal origins and diverse mitochondrial DNA in western populations. *Systematic Biology*, *42*(3), 293-306.

Moritz, C., Fujita, M. K., Rosauer, D., Agudo, R., Bourke, G., Doughty, P., ... & Scott, M. (2016). Multilocus phylogeography reveals nested endemism in a gecko across the monsoonal tropics of Australia. *Molecular ecology*, *25*(6), 1354-1366.

Moro, D., & MacAulay, I. (2010). A guide to the reptiles and amphibians of Barrow Island.

Murray, B. A., Bradshaw, S. D., & Edward, D. H. (1991). Feeding behavior and the occurrence of caudal luring in Burton's pygopodid Lialis burtonis (Sauria: Pygopodidae). *Copeia*, 509-516.

Nydam, R. L. (2000). A new taxon of helodermatid-like lizard from the Albian–Cenomanian of Utah. *Journal of vertebrate Paleontology*, *20*(2), 285-294.

Orr, H. A. (1995). The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*, *139*(4), 1805-1813.

Orr, H. A., & Orr, L. H. (1996). Waiting for speciation: the effect of population subdivision on the time to speciation. *Evolution*, *50*(5), 1742-1749.

Ortiz-Jaureguizar, E., & Cladera, G. A. (2006). Paleoenvironmental evolution of southern South America during the Cenozoic. *Journal of Arid Environments*, *66*(3), 498-532.

Otto-Bliesner, B. L., Marshall, S. J., Overpeck, J. T., Miller, G. H., & Hu, A. (2006). Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *science*, *311*(5768), 1751-1753.

Patchell, F. C., & Shine, R. (1986). Hinged teeth for hard-bodied prey: a case of convergent evolution between snakes and legless lizards. *Journal of Zoology*, *208*(2), 269-275.

Pepper, M., Pagán, A. J., Igyártó, B. Z., Taylor, J. J., & Jenkins, M. K. (2011). Opposing signals from the Bcl6 transcription factor and the interleukin-2 receptor generate T helper 1 central and effector memory cells. *Immunity*, *35*(4), 583-595.

Pepper, M., Doughty, P., & Keogh, J. S. (2013). Geodiversity and endemism in the iconic Australian Pilbara region: a review of landscape evolution and biotic response in an ancient refugium. *Journal of Biogeography*, *40*(7), 1225-1239.

Pepper, M., Hamilton, D. G., Merkling, T., Svedin, N., Cser, B., Catullo, R. A., ... & Keogh, J. S. (2017). Phylogeographic structure across one of the largest intact tropical savannahs: molecular and morphological analysis of Australia's iconic frilled lizard Chlamydosaurus kingii. *Molecular phylogenetics and evolution*, *106*, 217-227.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, *7*(5), e37135.

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature genetics*, *48*(1), 94.

Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, *31*(2), 161-175.

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: an open-source release of Maxent. *Ecography*, *40*(7), 887-893.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS genetics*, *8*(11), e1002967.

Potter, S., Eldridge, M. D., Taggart, D. A., & Cooper, S. J. (2012). Multiple biogeographical barriers identified across the monsoon tropics of northern Australia: phylogeographic analysis of the brachyotis group of rock-wallabies. *Molecular Ecology*, *21*(9), 2254-2269.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Pulliam, H. R. (2000). On the relationship between niche and distribution. *Ecology letters*, *3*(4), 349-361.

QGIS Development Team (2019). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Qiao, H., Peterson, A. T., Campbell, L. P., Soberón, J., Ji, L., & Escobar, L. E. (2016). NicheA: creating virtual species and ecological niches in multivariate environmental scenarios. *Ecography*, *39*(8), 805-813.

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of biogeography*, *41*(4), 629-643.

Rabosky, D. L., Donnellan, S. C., Talaba, A. L., & Lovette, I. J. (2007). Exceptional among-lineage variation in diversification rates during the radiation of Australia's most diverse vertebrate clade. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1628), 2915-2923.

Reisz, R. R., & Müller, J. (2004). Molecular timescales and the fossil record: a paleontological perspective. *TRENDS in Genetics*, *20*(5), 237-241.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*(7263), 489.

Richardson, J. E., Weitz, F. M., Fay, M. F., Cronk, Q. C., Linder, H. P., Reeves, G., & Chase, M. W. (2001). Rapid and recent origin of species richness in the Cape flora of South Africa. *Nature*, *412*(6843), 181.

Ritter, P. (1987). A vector-based slope and aspect generation algorithm. *Photogrammetric Engineering and Remote Sensing*, *53*(8), 1109-1111.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572-1574.

Rosauer, D. F., Blom, M. P. K., Bourke, G., Catalano, S., Donnellan, S., Gillespie, G., ... & Rabosky, D. L. (2016). Phylogeography, hotspots and conservation priorities: an example from the Top End of Australia. *Biological Conservation*, *204*, 83-93.

Rosindell, J., Hubbell, S. P., & Etienne, R. S. (2011). The unified neutral theory of biodiversity and biogeography at age ten. *Trends in ecology & evolution*, *26*(7), 340-348.

Schodde, R., & Mason, I. J. (1999). *Directory of Australian birds: passerines: Passerines*. CSIRO publishing.

Singhal, S. (2013). De novo transcriptomic analyses for non-model organisms: an evaluation of methods across a multi-species data set. *Molecular Ecology Resources*, *13*(3), 403-416.

Slatkin, M. (1978). Spatial patterns in the distributions of polygenic characters. *Journal of Theoretical Biology*, *70*(2), 213-228.

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

Steven J. Phillips, Miroslav Dudík, Robert E. Schapire. (2018). [Internet] Maxent software for modeling species niches and distributions (Version 3.4.1). Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2019-2-5.

Strasburg, J. L., & Kearney, M. (2005). Phylogeography of sexual Heteronotia binoei (Gekkonidae) in the Australian arid zone: climatic cycling and repetitive hybridization. *Molecular Ecology*, *14*(9), 2755-2772.

Sunnucks, P., & Hales, D. F. (1996). Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus Sitobion (Hemiptera: Aphididae). *Molecular biology and evolution*, *13*(3), 510-524.

Swan, M., & Watharow, S. (2005). *Snakes, lizards and frogs of the Victorian Mallee*. csiro Publishing.

Tapponnier, P., Mattauer, M., Proust, F., & Cassaigneau, C. (1981). Mesozoic ophiolites, sutures, and arge-scale tectonic movements in Afghanistan. *Earth and Planetary Science Letters*, *52*(2), 355-371.

Veloz, S. D. (2009). Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, *36*(12), 2290-2299.

Voris, H. K. (2000). Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, *27*(5), 1153-1167.

Wacey, D., Kilburn, M. R., Saunders, M., Cliff, J., & Brasier, M. D. (2011). Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nature Geoscience*, *4*(10), 698.

Walsh, P. S., Metzger, D. A., & Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques, 10*(4), 506-513.

Wall, M., & Shine, R. (2008). Post-feeding thermophily in lizards (Lialis burtonis Gray, Pygopodidae): laboratory studies can provide misleading results. *Journal of Thermal Biology*, *33*(5), 274-279.

Wall, M., & Shine, R. (2009). The Relationship Between Foraging Ecology and Lizard Chemoreception: Can a Snake Analogue (Burton's Legless Lizard, Lialis burtonis) Detect Prey Scent? *Ethology*, *115*(3), 264-272.

Wall, M., & Shine, R. (2013). Ecology and behaviour of Burton's Legless Lizard (Lialis burtonis, Pygopodidae) in tropical Australia. *Asian Herpetological Research*, *4*(1), 9-21.

Weber, E., & Werner, Y. L. (1977). Vocalizations of two snake-lizards (Reptilia: Sauria: Pygopodidae). *Herpetologica*, 353-363.

Wilson, S. (2005). Field Guide to Reptiles of Queensland.

Wilson, S., & Swan, G. (2010). A Complete Guide to Reptiles of Australia., 3rd edn.(Reed New Holland: Sydney.).

Appendix I

Description of Living Systems in this Study

Lialis burtonis

Lialis burtonis is named after Edward Burton, a famous British army surgeon (Beolens, *et al.*, 2011). Its colloquial name is Burton's legless lizard or Burton's snake-lizard. The species is bi-typic with *Lialis jacari* and is part of the endemic gecko family *Pyogopodidae*. This species lack forelegs but has a vestigial pelvic girdle and rudimentary hind legs that resemble flaps (Cogger, 2000). *L. burtonis* is analogous to macrostomatan snakes in ecological niche with similar adaptations for swallowing prey whole (Wall & Shine, 2013). A unique morphological character that *Lialis* possesses is a kinetic symphasis, mesokinetic and hypokinetic, on the median of the orbital along the dorsal and ventral portions (Patchell & Shine, 1986; Shine, 2007). Its head is elongated, with recurved hinged teether curving from the anterior to the posterior (Patchell & Shine, 1986). The elongated head also may allow for binocular vision for use in striking (Shine, 2007). The eye muscles are attached to retractor muscles that pull the eye inside the eye socket, acting as a protective measure during hunting to protect the eye – important for a visual predator (Shine, 2007).

Lialis primariliy resides almost exclusively in Australia *sans* Tasmania, however, does expand its distrobution to the island of New Guinea (Cogger, 2000). The range distibution is isolated to the savannas along the southern coast of New Guinea. It is the only pygopodid genus that has a range which extends out of the Australian landmass into Papua New Guinea. The habitat spaces that are occupied by *Lialis burtonis* include deserts, woodland, savanna, and margins of rainforests (Cogger, 2000; Cronin & Oram, 2001), usually found in leaf litter and grasses (Wall & Shine, 2013).

The diet of *Lialis* is unique in that it feeds exclusively on lizards, predominantly skins as well as geckos (Cogger, 2000; Cronin & Oram, 2001; Wall & Shine, 2009). Patchell & Shine (1986) found that the species also ingests small snakes as part of its diet. In general, like many ectotherms, *L. burtonis* is does not feed very often (Wall & Shine, 2013).

L. burtonis is an oviparous squamate that has a breeding season in the southern hemisphere's Spring and lays eggs in mid-Summer (Patchell & Shine, 1986; Cogger, 2000; Cronin, 2001). Eggs are usually observed under leaf litter, under logs and rocks, and occasionally inside insect nests (Cronin & Oram, 2001). Clutch size is between 1-3 leathery eggs, which, although rare, nests can be communal with up to 20 eggs (Neill, 1957; Patchell & Shine, 1986). *Lialis* can also store sperm, a common adaptation in squamates (Cogger, 2000). Hatchlings on average are approximately 10-15 cm in length (Cronin & Oram, 2001).

The predominant circadian rhythms of *L. burtonis* is a diurnal habit, however, it can be found at night too, likely, to move between hunting sites without being predated on (Cogger, 2000). Vision is an important sense for being a hunter (Cogger, 2000). Usually shelter is used to both ambush prey and thermoregulate (Wall & Shine, 2008). Caudal luring is implemented to attract prey while in ambush position, however, not a behavior exclusively used (Murray *et al.*, 1991). Caudal luring uses three methods of approach: prey luring, prey distraction, or both (Murray *et al.*, 1991). Since the prey items of *Lialis* can cause harm, and are adapted to not be caught, strike precision is modified based on the size of the prey with larger organisms having strikes aimed at the head a neck to prevent bite back (Wall & Shine, 2007). Large prey is restrained till fully capacitated, while smaller prey is ingested whole while still struggling (Wall & Shine, 2007). If threatened, the species will vocalize (Weber & Werner, 1977). If captured the species can also utilize tail autonomy.

Heteronotia binoei

Heteronotia binoei, commonly referred to as Bynoe's Gecko, is a small (~50 mm SVL) slender gekkonid lizard from Australia with highly variable coloration (Wilson, 2005; Swan & Watharow, 2005; Moro & MacAulay, 2010; Johansen, 2012). Color pattern ranges from brown to reddish-brown, grey, yellowish-brown or black, and it is usually irregularly patterned possessing both light and dark bands, spots and flecks, or a combination of both (Swan & Watharow, 2005; Wilson, 2005; Moro & MacAulay, 2010; Johansen, 2012). *Heteronotia binoei* usually possesses a stripe that extends from the mouth through the eye and around the neck, which can be broken, solid, or missing altogether (Moro & MacAulay, 2010). This small lizard has a slender body along with a long, slender tail tapering to a point (Swan & Watharow, 2005). Its rostral and mental shields are rounded with enlarged labial scales compared to neighboring scales, which are granular (Swan & Watharow, 2005; Wilson, 2005). There are four enlarged postmental scales (Swan & Watharow, 2005; Wilson, 2005). Scales are small, keeled and spike-like, alternatively giving a colloquial name to *Heternotia binoie,* 'prickly gecko' (Swan & Watharow, 2005). Enlarged tubercles are arranged in a scattered (irregular) pattern over the dorsum (Johansen, 2012; Swan & Watharow, 2005; Wilson, 2005). It has long and slender digits with claws between three enlarged scales and no enlarged apical lamellae (Swan & Watharow, 2005; Wilson, 2005).

Heteronotia binoei is distinguishable from other described species in the genus by the uneven tubercle pattern where other species have tubercules arranged in rows (Johansen, 2012). Like other geckos, *Heteronotia binoei* has a relatively large head and a lack of eyelids (Halliday & Adler, 2002). *Heteronotia binoei* has well-developed vocal cords, common in most geckos species, and there are reports of high variation within *Heteronotia binoei* vocal

calls (Halliday & Adler, 2002). *Heteronotia binoei* is also a species complex comprised of many undescribed species (Johansen, 2012; Moritz personal communication; Pepper *et al.*, 2013).

The type species of the genus is *Heteronotia binoei* GRAY 1845. *Heteronotia* WERMUTH 1965 (nomen novum pro Heteronota GRAY non Heteronota GERMAR 1835 = nomen substitutum pro Heteronotus LAPORTE-CASTELNAU 1832 [Rhynchota] non Heteronota MEYRICK 1902 [Lepidoptera]). The holotype is accession number ZMB 5718 [Phyllodactylus anomalus PETERS 1867]. For many years there were only two species within the genus – *Heteronotia binoei* (generalist) and *Heteronotia splea* (saxicolous). In the 1970's karyotyping revealed a complex of chromosome races (King, 1979; King, 1983; Moritz, 1986). Later an intriguing find by Moritz (1984) discovered parthenogenetic populations. Little has changed taxonomically within *Heteronotia*, over the last two decades (Cogger, 2000; Storr, 1990; Wilson, 2010). *H. planiceps* (Storr, 1989) was the last described species within the genus until Pepper *et al.* (2013) erected two new species, *H. atra* and *H. fasciolatus*. *Heteronotia splea* was erected as a species by Kluge (1963) through a revision of *Heteronotia* GRAY. Both *Gehyra* and *Heteronotia* have been the subject of many multi-locus genetic studies, revealing evolutionary relationships and geographic distributions (Fujita *et al.*, 2010; Heinicke *et al.*, 2011; Sistrom *et al.*, 2009; 2012; Pepper *et al.*, 2011). Unlike *Gehyra*, *Heteronotia's* extreme cryptic diversity the taxomony of the genus has remained unchanged despite the increased knowledge to the genetic history within the lineages (Fujita *et al.*, 2010; Pepper *et al.*, 2011) until Pepper *et al.* (2013).

The environment that *Heteronotia* resides in is extreme. It consists of a two-biome complex with an arid portion in the center of the continent and a monsoonal biome to the

north and along the eastern coast where Australia's tropical rainforest is situated. The genus has an unusually large and cosmopolitan range throughout Australia, including lineages in both desert, arid scrub, and mosaic rock outcrops in the arid biome and north end dry tropics including savanna and southern rainforest in the monsoonal biomes (Fujita *et al.*, 2010; Pepper *et al.*, 2011). Rock-specialized habit of both *H. spelea* and *H. planiceps* highlights the relevance of an understanding of the geological setting around which these species have evolved. Regions where rock dwelling species, *H. spelea* and *H. planiceps*, are found (the Kimberley and Pilbara regions) are highly diverse in their geology and topography. The landforms of the Kimberley are rugged, characterized by extreme geology with faulted rocks, steeply scarped sandstone-capped hills, and mountains and limestone gorges (Bridgewater, 1987). To the south lie uniform expanses of linear sand ridges of inland deserts, which act as a barrier and factor of isolation between groups of *Heteronotia* (Beard, 1979; Fujita *et al.*, 2010; Pepper *et al.*, 2013). Another unique area is the Pilbara – where lineages of *Heteronotia* are in a mosaic distribution. The Pilbara is located in the northwestern part of Western Australia province. Two major mountain ranges, with deep gorges incised into the ironstone hills, lie to the south and central regions of the Pilbara. The northern Pilbara is less extreme where gently rolling granite hills have been weathered throughout earth's life history. This region has some of the oldest rocks on earth (Wacey *et al.*, 2011). The inland deserts of the continent bind the northern and eastern sides of the Pilbara. The central ranges are comprised of limestone and granite valleys with rugged ridges and dry riverbed system scattered throughout (Boden & Given, 1995; Mabbutt, 1988). These rugged inland valleys and ridges are completely encompassed by desert lowlands, giving the area extreme climate throughout the year (Pepper *et al.*, 2013).

Biogeographic patterns of *Heternotia* are influenced by aridification of the southern hemisphere brought about by glaciation periods (Avise, 2000).

Environmental change, especially through climate fluctuation, significantly drove biotic diversification through expanding distributions during interglacial periods (Ackerly, 2003; Avise, 2000; Levins, 1968). The Southern Hemisphere experienced periods of arid expansion during interglacial periods in the mid-Miocene and Pleistocene (Byrne *et al*., 2008; Martin, 2006). As with *Heteronotia,* much of the within-species biodiversity in Australia arose as a result of aridification cycles in the Pliocene (Clapperton, 1990; Markgraf, *et al.*, 1995). These aridification events are known to cause mass extinction (Cogger & Heatwole, 1981; Crisp *et al.*, 2009), thus opening large new adaptive zones allowing for new lineages to inhabit previously occupied niches. The Pilbara is an intermediate region between arid and tropical landscapes. These regions led to a distinct biogeography of *Heternotia*. (Fujita *et al.*, 2010; Pepper *et al.*, 2013). Paleoclimate fluctuations have resulted in both extensions and retreats of these two biomes, having influenced the demographic histories of populations, including population expansions and gene flow dynamics in *Heteronotia* (Fujita, *et* al., 2010; Pepper *et al.*, 2011; Pepper *et al.*, 2013). *H. planiceps* and *H. spelea* are entirely disjunct in distribution with each lineage confined to separate mountain ranges within the arid zone or in the periphery. Transitions between biomes are uncommon (Crisp et al., 2009), but this recent aridification (<15 mya) in the southern hemisphere opened new habitats that allowed invasion from older (already inhabited) mesic environments into younger arid lands (Rabosky *et al.*, 2007). These lineages are younger than their mesic-inhabiting ancestors, suggesting that adaptation to arid conditions is recent (Pepper *et al.*, 2011). The patterns seen in Heteronotia are consistent with those seen in these other Southern Hemisphere systems –

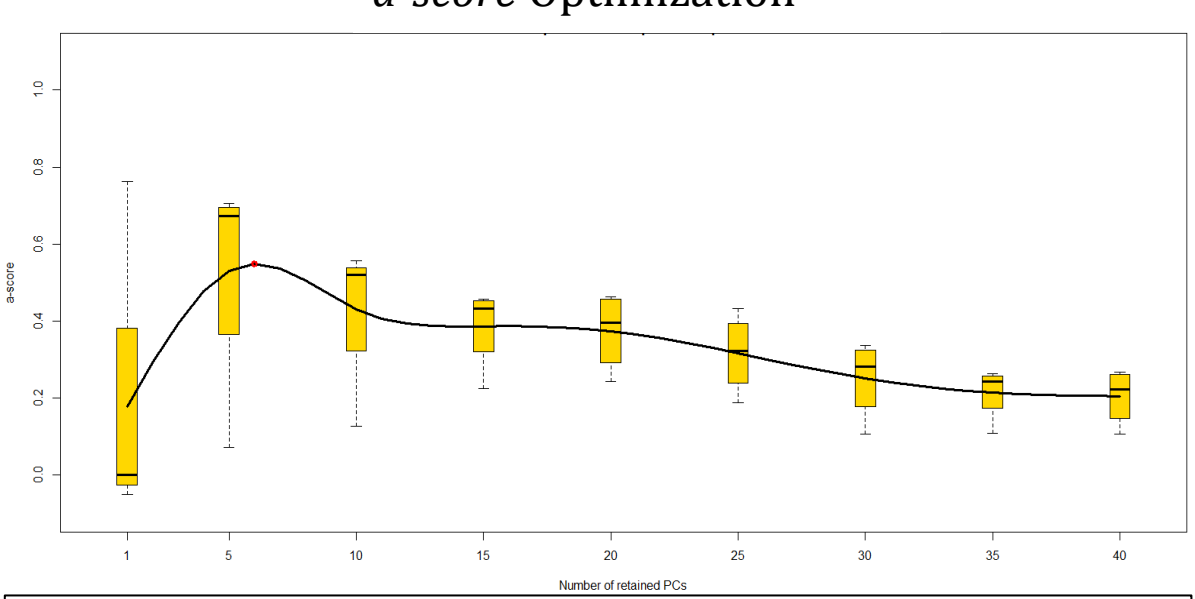
diversification and persistence in the monsoonal tropics, and range expansions in the arid zone.

Divergent time analysis from Pepper *et al.* (2011) and Fujita *et al.* (2010) suggest the divergence between *H. binoei*, *H. planiceps* and *H. spelea* conservatively around 5 mya to 2.4 mya, which corresponds to the Pliocene to the early Pleistocene. Time tree analysis siggest near simultaneous diversification (rapid radiation) because of short branch lengths and largely overlapping confidence intervals. This corresponds to interglacial periods of aridification and tropical refugia, where populations migrated to newly opened niches. This timing is much older than the formation of desert landforms (approximately 1 million years ago), which implies that desert landforms are not the imposing form of isolation causing speciation among *Heteronotia* species residing in desert isolated mountain ranges (Fujita *et al.*, 2010). Similar ages were estimated for divergences between major lineages of the *H. binoei* complex (Fujita et al. 2010). The divergence time estimates by Fujita *et al.* (2010) and Pepper *et al.* (2011) coincide with a time of extreme environmental change associated with deepening aridity and increased seasonality (Byrne et *al.*, 2008). The basin areas separating the uplands would have experienced termination of the warmer, wetter conditions – constituting in tropical refugia that the Australian continent experienced in the early Miocene. Evidence of palaeodrainage channels in central and western Australia corroborates aridification and loss of wetter climate because of unprecedented levels of erosion across the landscape (Martin, 2006). These conditions in central and western Australia previously considered favorable by *Heteronotia* ancestors would have changed sufficiently in the lowlands to fragment populations into the moister refuges of the rocky uplands of the Pilbara, Kimberley, and central Australia allowing time for isolation and further speciation (Pepper *et al.*, 2011). Pilbara and Central *H. spelea* lineages would have occurred recently around 2.9

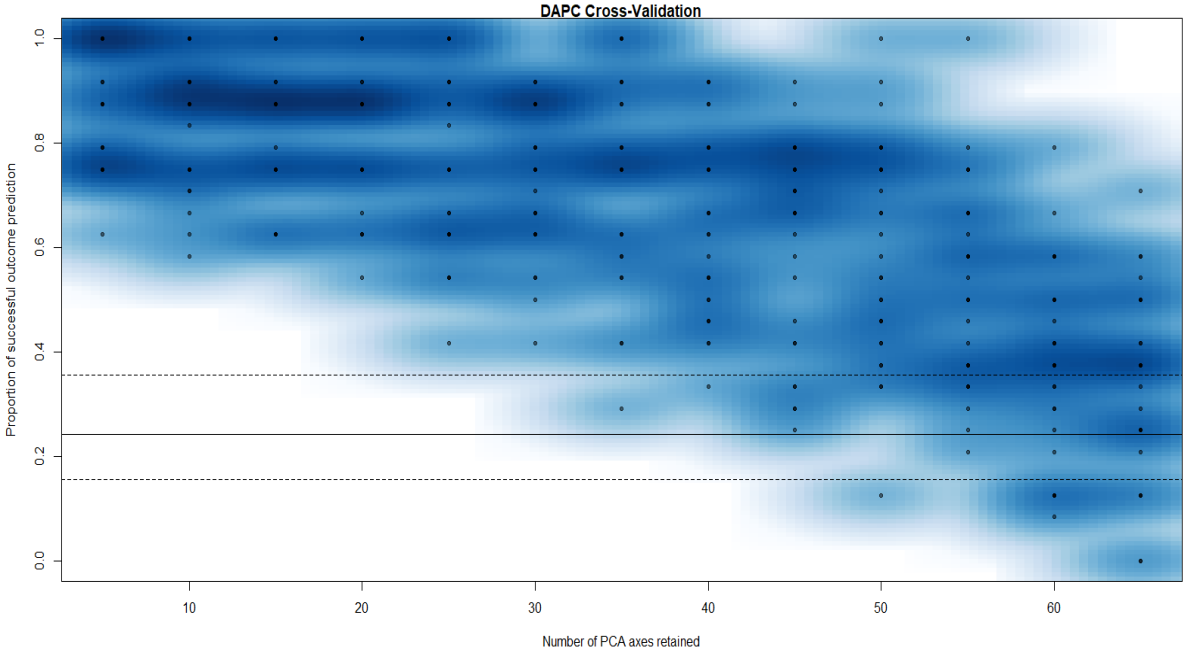
Ma, and possibly much earlier, which is also older than the geological age of desert formation, further reaffirming mobile deserts not existing and playing a role as an isolating barrier (Pepper *et al.*, 2011). Divergent times for non-saxicolous *Heteronotia*, (*H. spelea, H. fasciolatus, H. atra,* and *H. planiceps*) *H. binoei* diverges within 0.69 and 2.27 mya (north-central arid lineage – SM6) and 0.3 and 1.14 mya (central lineage below SM6 – CA6), with strong signatures of recent population expansion <1 Ma (Fujita *et al.*, 2010).

Appendix II

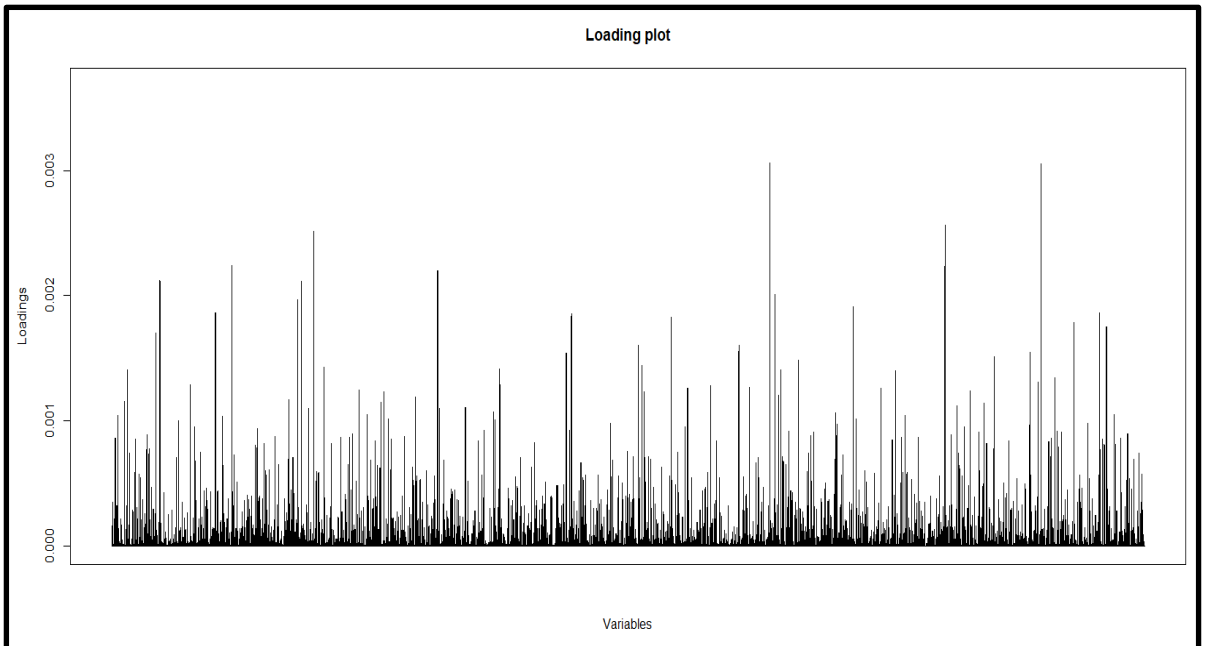
List of supplementary figures and tables.



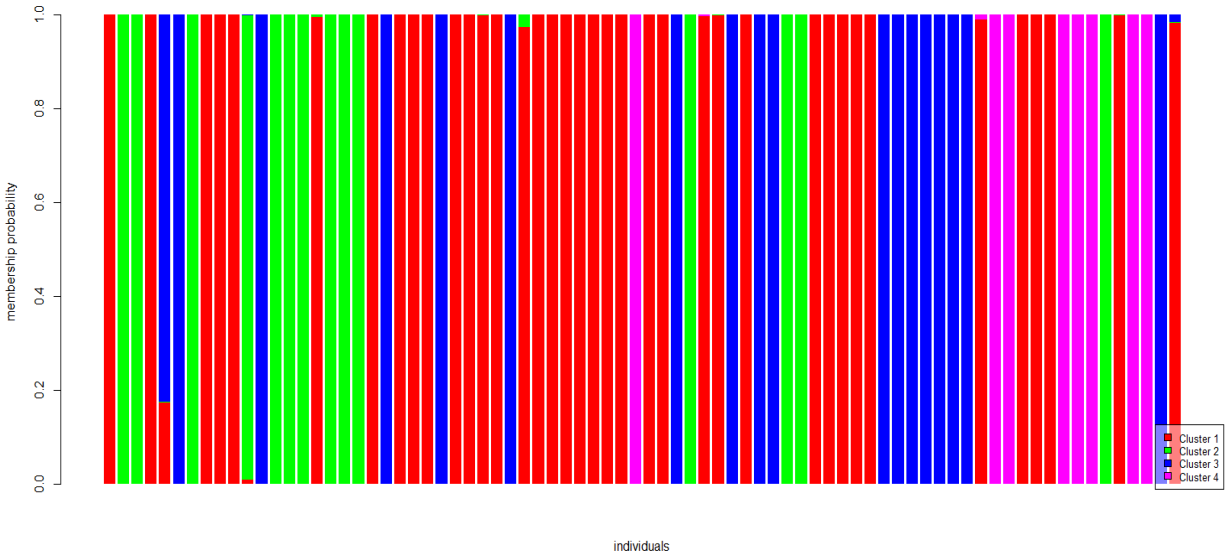
Supplementary Figure 1. *a-score* of randomized *L. burtonis* groups to assess principal comments retained in DAPC. The *a-scores* of PCs retained show a level of model significance at 40 PCs that do no overparameterize with little Type I error in group (lineage) assignments.



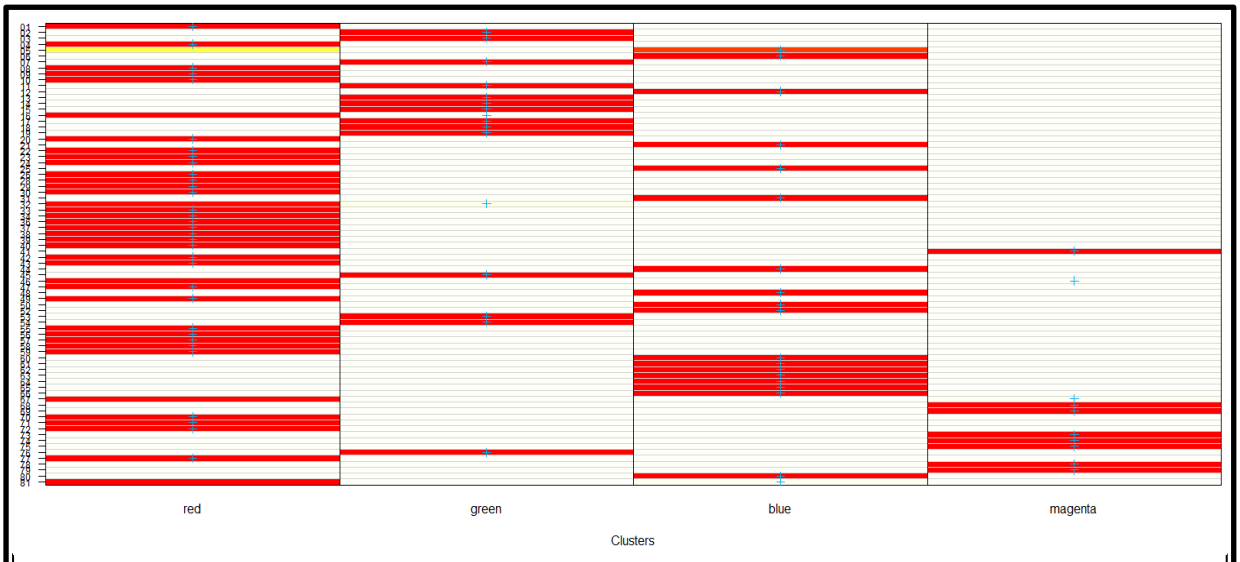
Supplementary Figure 2. Cross-validation of estimated success of k-value prediction. Darker color yield high success rate.



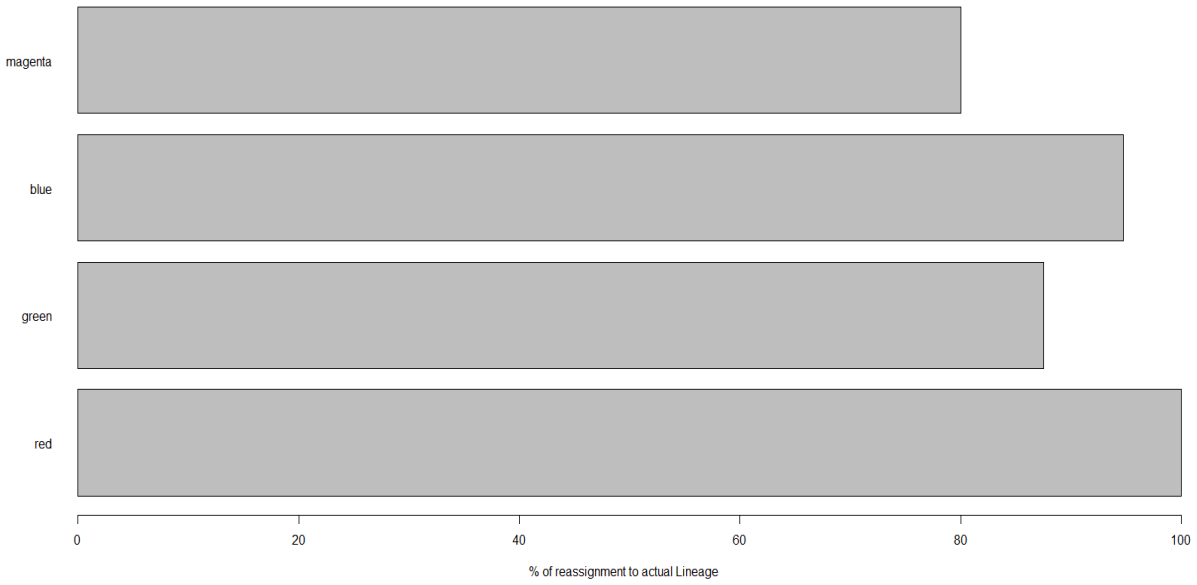
Supplementary Figure 3. Loadings from alleles (SNP variables) in their overall contribution to principal components.



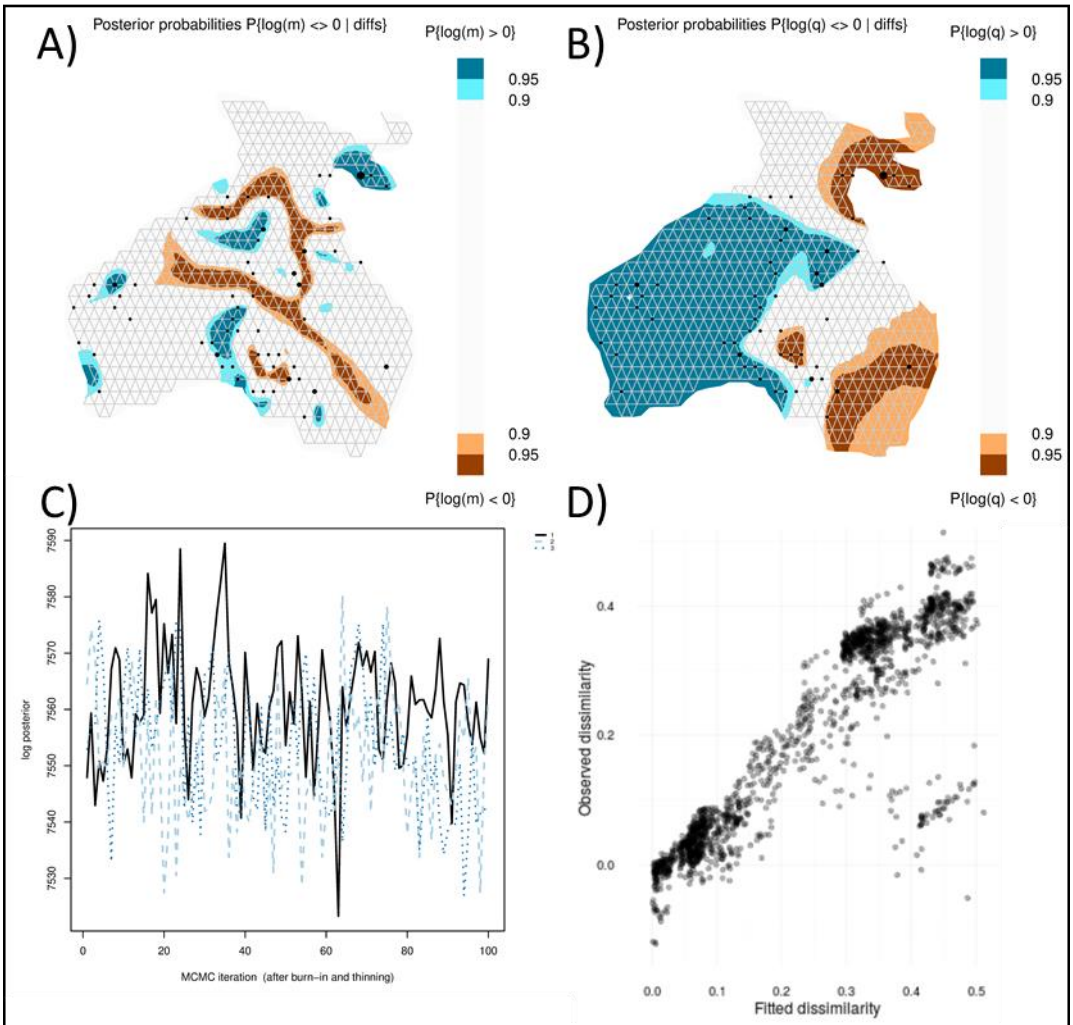
Supplementary Figure 4. Membership cluster plot from DAPC.



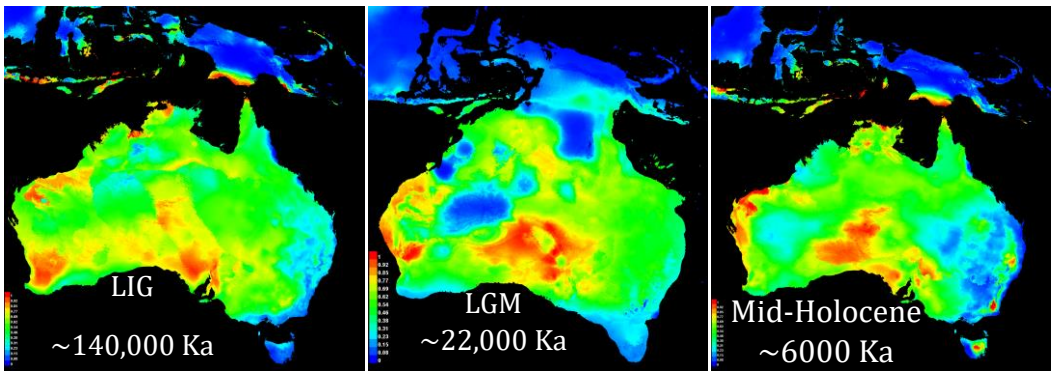
Supplementary Figure 5. Predictive plot to estimate Type I error of *a prior* population designations. Blue + marks symbolize the *a prior* population designations. Red symbolizes high probability, yellow low probability, and open cells show no probability.



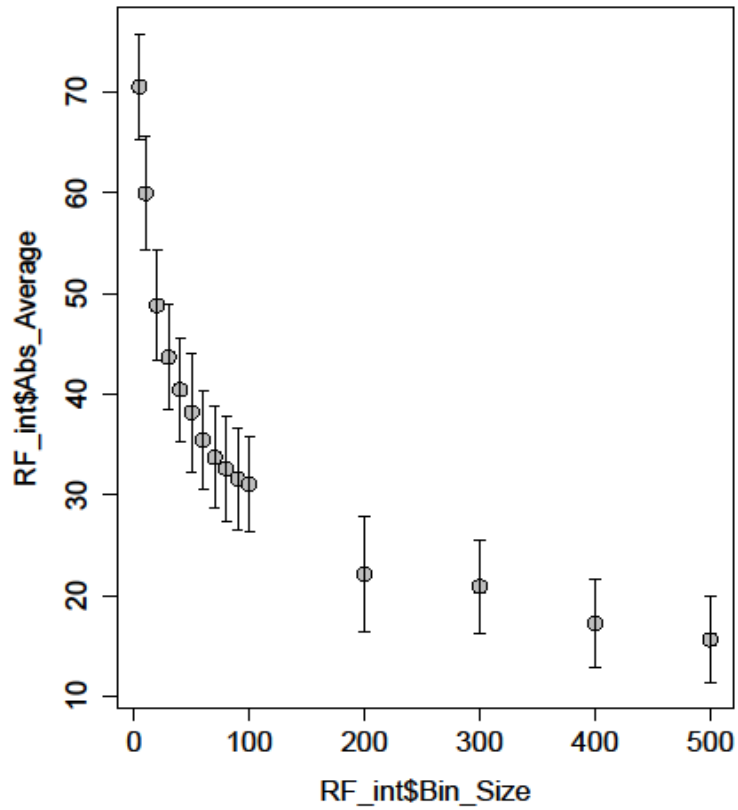
Supplementary Figure 6. Bar plot illustrating assignment of the number of individuals within the group as percentages.



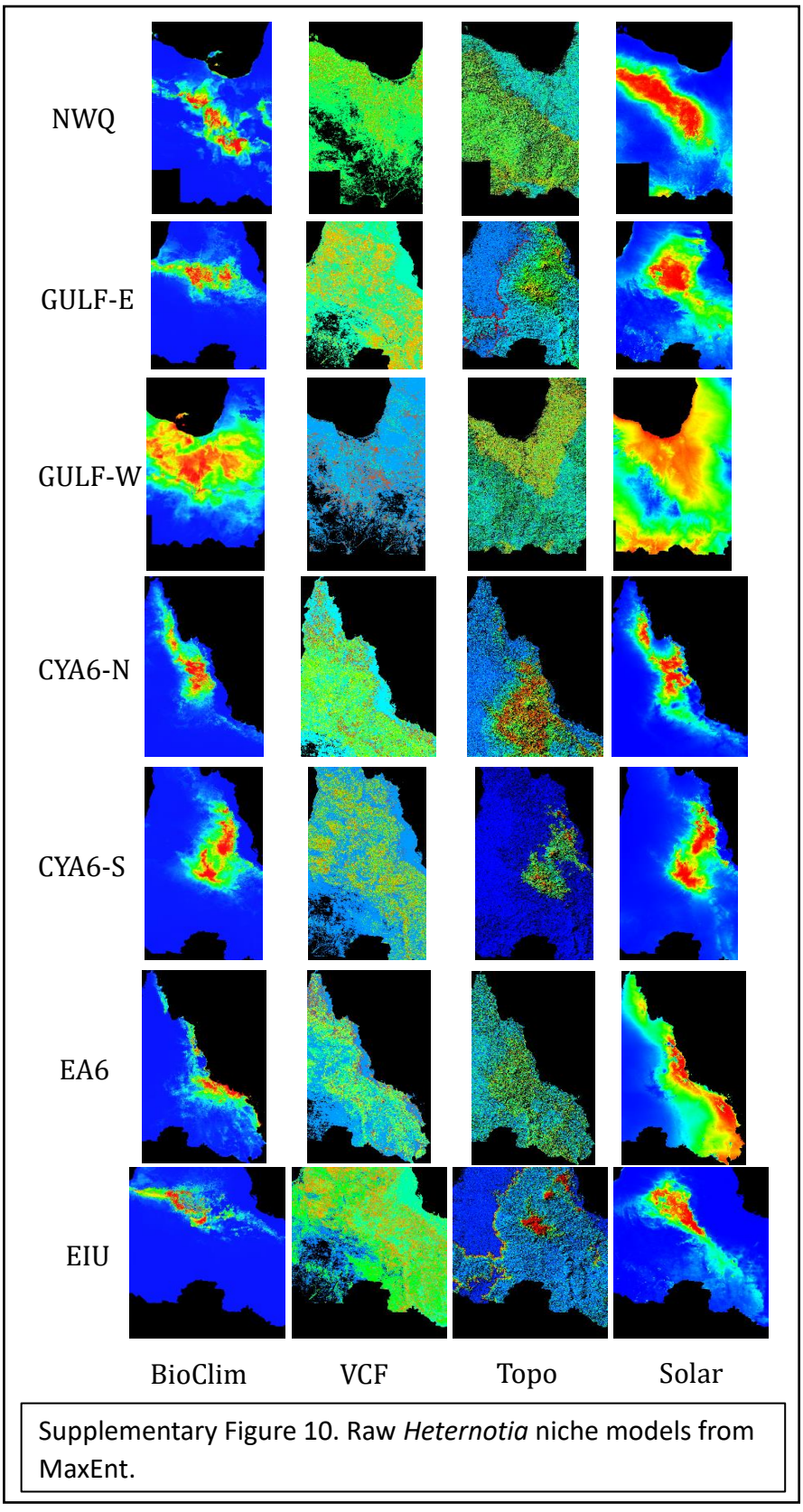
Supplementary Figure 7. A) Posterior probabilities from migration map. B) Posterior probabilities from diversity map. C) MCMC convergence plot. D) EEMS dissimilarity plot.



Supplementary Figure 8. Raw paleo-niche models from MaxEnt.



Supplementary Figure 9. Robinson-Fold Metric generated from Astral using 500 bootstrap bin method. Based on the bin and the amount of gene trees generated to estimate the ideal number of loci needed to generate an accurate topology consistently.



Supplementary Figure 10. Raw *Heternotia* niche models from MaxEnt.

Supplementary Table 1. Primers used in for *ND*2 in this study.

Gene	**Direction**	**BP**	**Annealing Temp**	**Primers (5'-3')**	**GenBank Accession**
nad2	Forward	1041	55	*tRNAI*: AAGGACTACTTTGATAGAGT4	GU388200
	Reverse			*tRNAA*: AAAGTGTTTGAGTTGCATTCAG4	GU388302
*From Strausburg and Kearney (2005)					

Supplementary Table 2. Summary statistics from RADseq initial assembly and filtering.

The number of loci caught by each filter.
ipyrad API location: [assembly].stats_dfs.s7_filters

	total_filters	**applied_order**	**retained_loci**
total_prefiltered_loci	103399	0	103399
filtered_by_rm_duplicates	7296	7296	96103
filtered_by_max_indels	466	466	95637
filtered_by_max_snps	3231	1159	94478
filtered_by_max_shared_het	269	42	94436
filtered_by_min_sample	90989	85728	8708
filtered_by_max_alleles	14899	1843	6865
total_filtered_loci	6865	0	6865

The number of loci recovered for each Sample.
ipyrad API location:
[assembly].stats_dfs.s7_samples

	sample_coverage
ABTC06583_red	5999
ABTC102707_green	4855
ABTC102713_green	5504
ABTC12043_red	6046
ABTC130658_blue	1078
ABTC14519_blue	5212
ABTC18024_green	5337
ABTC21863_red	6118
ABTC24061_red	6190
ABTC24154_red	5564
ABTC28071_green	5682
ABTC28237_blue	4243
ABTC28478_green	5387
ABTC29346_green	4945
ABTC29386_green	5229

ABTC29916_green	254
ABTC29938_green	5211
ABTC30297_green	5444
ABTC30417_green	5217
ABTC30915_red	5366
ABTC34933_blue	5567
ABTC35616_red	5880
ABTC35702_red	5511
ABTC36133_red	5978
ABTC3721_blue	5224
ABTC39401_red	5999
ABTC41731_red	5119
ABTC42217_red	45
ABTC42470_red	1302
ABTC49515_blue	5120
ABTC51378_green	636
ABTC52028_red	5579
ABTC57020_red	4645
ABTC58064_red	3560
ABTC58538_red	4140
ABTC58591_red	3657
ABTC59720_red	5648
ABTC59721_red	4665
ABTC62019_magenta	1974
ABTC64149_red	3737
ABTC64302_red	4308
ABTC6485_blue	4916
ABTC6491_green	4605
ABTC68796_magenta	5220
ABTC68902_red	10
ABTC70388_blue	5488
ABTC72801_red	5299
ABTC72844_blue	5323
ABTC77188_blue	3896
ABTC82421_green	5307
ABTC82436_green	3318
ABTC87339_red	6073
ABTC94340_red	6202
ABTC94768_red	5589
ABTC95330_red	5820
ABTC95758_red	6107
CCA2477_blue	4772
CCA2478_blue	3136
CCA2755_blue	1935
CCA3863_blue	4703

CCA3901_blue	3269
CCA5177_blue	5022
CCA5629_blue	1789
R102424_magenta	1099
R117118_magenta	3915
R123735_magenta	4961
R125975_red	4274
R127542_red	5444
R132003_red	5475
R139138_magenta	2169
R145508_magenta	2172
R145685_magenta	4317
R151829_green	3409
R154028_red	21
R156267_magenta	5378
R161046_magenta	4354
TG2108_blue	4280
TG2109_blue	507

The number of loci for which N taxa have data.
ipyrad API location: [assembly].stats_dfs.s7_loci

	locus_coverage	sum_coverage
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0

22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	376	376
37	367	743
38	364	1107
39	330	1437
40	332	1769
41	277	2046
42	264	2310
43	232	2542
44	247	2789
45	243	3032
46	231	3263
47	245	3508
48	200	3708
49	218	3926
50	208	4134
51	198	4332
52	184	4516
53	163	4679
54	182	4861
55	175	5036
56	159	5195
57	165	5360
58	146	5506
59	139	5645
60	126	5771
61	128	5899
62	103	6002
63	116	6118
64	95	6213
65	84	6297
66	86	6383

67	78	6461
68	82	6543
69	87	6630
70	71	6701
71	65	6766
72	50	6816
73	30	6846
74	14	6860
75	5	6865
76	0	6865
77	0	6865
78	0	6865
79	0	6865
80	0	6865
81	0	6865

The distribution of SNPs (var and pis) per locus.
var = Number of loci with n variable sites (pis + autapomorphies)
pis = Number of loci with n parsimony informative site (minor allele in >1 sample)
ipyrad API location: [assembly].stats_dfs.s7_snps

	var	sum_var	pis	sum_pis
0	27	0	194	0
1	83	83	407	407
2	144	371	539	1485
3	198	965	708	3609
4	277	2073	819	6885
5	349	3818	729	10530
6	418	6326	728	14898
7	406	9168	681	19665
8	492	13104	533	23929
9	572	18252	447	27952
10	540	23652	348	31432
11	499	29141	251	34193
12	467	34745	202	36617
13	431	40348	130	38307
14	412	46116	80	39427
15	347	51321	39	40012
16	337	56713	24	40396
17	284	61541	4	40464
18	228	65645	2	40500
19	193	69312	0	40500
20	161	72532	0	40500

Final Sample stats summary

	state	Reads raw	Reads passed filter	Clusters total	Clusters hidepth	Hetero est	Error est	Reads consensus	Loci in assembly
ABTC06583_red	7	4365037	4365037	294833	31691	0.023973	0.009192	27661	5999
ABTC102707_green	7	2242936	2242936	286165	26584	0.023586	0.014492	22739	4855
ABTC102713_green	7	6734344	6734344	376649	37800	0.028005	0.010279	32558	5504
ABTC12043_red	7	7254306	7254306	542084	37742	0.031158	0.010029	31488	6046
ABTC130658_blue	7	189541	189541	18986	4819	0.020197	0.00629	4427	1078
ABTC14519_blue	7	3541347	3541347	363586	38000	0.025936	0.014639	31307	5212
ABTC18024_green	7	5521337	5521337	341114	34055	0.025589	0.009446	29608	5337
ABTC21863_red	7	6562681	6562681	471996	42541	0.027142	0.009278	35730	6118
ABTC24061_red	7	8107446	8107446	464607	41120	0.029213	0.008043	35164	6190
ABTC24154_red	7	4158180	4158180	284816	28782	0.0254	0.011762	24669	5564
ABTC28071_green	7	1683895	1683895	68400	25387	0.025896	0.003951	23345	5682
ABTC28237_blue	7	4354465	4354465	264274	25792	0.026546	0.011848	21754	4243
ABTC28478_green	7	5060263	5060263	352149	34212	0.026366	0.010315	29404	5387
ABTC29346_green	7	4333815	4333815	410896	31216	0.026148	0.013401	26146	4945
ABTC29386_green	7	6141410	6141410	364994	32730	0.02592	0.009105	28336	5229
ABTC29916_green	7	54969	54969	21106	1339	0.030207	0.01601	1079	254
ABTC29938_green	7	7893424	7893424	531615	36704	0.031042	0.009515	30592	5211
ABTC30297_green	7	6274927	6274927	427113	37004	0.02738	0.00828	31356	5444
ABTC30417_green	7	6482287	6482287	463777	35596	0.029149	0.011371	29535	5217
ABTC30915_red	7	1303070	1303070	63967	21266	0.020558	0.003541	19710	5366
ABTC34933_blue	7	9904441	9904441	557008	43387	0.034254	0.00891	35721	5567
ABTC35616_red	7	6903447	6903447	241881	30048	0.022678	0.005269	26782	5880
ABTC35702_red	7	3257111	3257111	213347	26672	0.022475	0.008128	23546	5511
ABTC36133_red	7	6605969	6605969	342749	36025	0.02582	0.007331	31656	5978
ABTC3721_blue	7	2237289	2237289	62920	24636	0.015918	0.002349	23188	5224
ABTC39401_red	7	5273352	5273352	299940	32068	0.024531	0.008504	27934	5999
ABTC41583_red	7	10277	10277	6772	92	0.071009	0.025775	48	NaN

ABTC41731_red	7	1721871	1721871	78578	19465	0.019152	0.002581	18175	5119
ABTC42217_red	7	13474	13474	6804	233	0.040483	0.019195	174	45
ABTC42470_red	7	157033	157033	47031	4474	0.02313	0.008931	3981	1302
ABTC49515_blue	7	2568601	2568601	83316	27621	0.01514	0.00275	25887	5120
ABTC51378_green	7	307200	307200	102946	4612	0.031817	0.023034	3187	636
ABTC52028_red	7	3402136	3402136	161130	25833	0.021364	0.002255	23588	5579
ABTC57020_red	7	1838196	1838196	112083	17482	0.021556	0.002584	16019	4645
ABTC57182_red	7	534	534	446	5	0.04643	0.034114	3	NaN
ABTC58064_red	7	1200040	1200040	60181	12812	0.017537	0.002651	11869	3560
ABTC58538_red	7	1537254	1537254	90604	17275	0.022127	0.002724	15832	4140
ABTC58591_red	7	958275	958275	96472	12990	0.020584	0.00324	11868	3657
ABTC59720_red	7	1588207	1588207	57346	22622	0.014667	0.002897	21292	5648
ABTC59721_red	7	2087510	2087510	78274	17512	0.017426	0.002079	16281	4665
ABTC62019_magenta	7	234916	234916	54047	7200	0.021388	0.007316	6534	1974
ABTC64149_red	7	731385	731385	51255	13208	0.018624	0.003385	12343	3737
ABTC64302_red	7	792581	792581	55307	14707	0.018442	0.003408	13666	4308
ABTC6485_blue	7	1968542	1968542	58903	23567	0.015761	0.0033	22153	4916
ABTC6491_green	7	5655427	5655427	330417	28386	0.028907	0.010272	23970	4605
ABTC68796_magenta	7	1517478	1517478	60398	21023	0.017239	0.003439	19547	5220
ABTC68902_red	7	16899	16899	8506	193	0.058407	0.027335	103	10
ABTC70388_blue	7	9830894	9830894	515010	40084	0.033155	0.008456	33326	5488
ABTC72801_red	7	2251866	2251866	68842	27041	0.01824	0.002905	25319	5299
ABTC72844_blue	7	5545796	5545796	268701	29671	0.026145	0.008057	25839	5323
ABTC76457_red	7	5322	5322	3237	64	0.071872	0.029963	23	NaN
ABTC77188_blue	7	939258	939258	49272	15899	0.019198	0.004049	14735	3896
ABTC82421_green	7	2133750	2133750	65288	26679	0.019386	0.003057	24903	5307
ABTC82436_green	7	784654	784654	46338	13896	0.020665	0.00454	12789	3318
ABTC87339_red	7	5383046	5383046	303944	33294	0.024262	0.009148	29179	6073
ABTC94340_red	7	5852250	5852250	300163	37651	0.024561	0.0074	33313	6202
ABTC94768_red	7	3713038	3713038	175543	24212	0.019219	0.0072	21881	5589
ABTC95330_red	7	3935517	3935517	260765	28166	0.022022	0.011623	24838	5820

ABTC95758_red	7	6354531	6354531	309155	36087	0.024136	0.008265	31971	6107
CCA2477_blue	7	1935288	1935288	55975	21968	0.012931	0.003486	20609	4772
CCA2478_blue	7	847973	847973	42629	12313	0.013464	0.003638	11531	3136
CCA2755_blue	7	432198	432198	41685	8269	0.01538	0.005102	7618	1935
CCA3863_blue	7	2626186	2626186	45508	21005	0.012061	0.002696	19836	4703
CCA3901_blue	7	3092008	3092008	39602	15099	0.014768	0.002232	14128	3269
CCA5177_blue	7	3552179	3552179	64967	28928	0.012646	0.002748	27144	5022
CCA5629_blue	7	386062	386062	30208	7463	0.015117	0.005415	6891	1789
R102424_magenta	7	149018	149018	26982	4370	0.023022	0.006688	3976	1099
R117118_magenta	7	1132073	1132073	100027	14814	0.0219	0.002993	13588	3915
R123735_magenta	7	3188758	3188758	108260	24015	0.021471	0.001998	22093	4961
R125975_red	7	1136529	1136529	70099	17073	0.019677	0.002982	15896	4274
R127542_red	7	2633919	2633919	84343	25128	0.018026	0.00212	23563	5444
R132003_red	7	1853170	1853170	100048	25831	0.018677	0.002956	23960	5475
R139138_magenta	7	327545	327545	32127	8255	0.018937	0.005207	7640	2169
R145508_magenta	7	519316	519316	51750	9669	0.020324	0.003688	8941	2172
R145685_magenta	7	2233734	2233734	121604	19688	0.023783	0.00235	17896	4317
R151829_green	7	1642380	1642380	99039	19556	0.017955	0.002846	18013	3409
R154028_red	7	16341	16341	7714	238	0.055824	0.024735	140	21
R156267_magenta	7	3786083	3786083	73011	26484	0.015353	0.001827	24877	5378
R161046_magenta	7	920091	920091	122435	16678	0.021636	0.00393	15222	4354
TG2108_blue	7	2792914	2792914	110668	27154	0.017224	0.002873	25124	4280
TG2109_blue	7	333020	333020	84050	6391	0.028438	0.012164	5219	507

Supplementary Table 3. FourPop test from TreeMix. Number of populations tested were four with SNPs 69684. Estimating covariance matrix in 139 blocks of size 500. Estimating f_4 statistic in 139 blocks of size 500.

Population_combination	**f4_stat**	**Error**	**Z-Score**
red,blue;green,magenta	-0.05155	0.001089	-47.3291
red,green;blue,magenta	-0.05124	0.00108	-47.4415
red,magenta;green,blue	-0.00031	0.00018	-1.71493

Supplementary Table 4. Calibrations used in BEAST2 dating analysis.

Calibration Points				
Type	**Calibration**	**Age (*MYA*)**	**Node**	**Distribution**
Fossil	*Hoburogekko suchanovi*	112	Crown Gekkota	Log normal
Fossil	*Sphaerodactylus dommeli and S. ciguapa*	15-20	Crown *Sphaerodactylus*	Log normal
Fossil	*Pygopus hortulanus*	20	*Paradelma orientalis* + *Pygopus nigriceps*	Log normal
Fossil	*Primaderma nessovi*	99	Helodermatidae + Anguidae	Log normal
Fossil	*Polysphenodon* and *Brachyrhinodon*	225	Squamata + *Sphenodon*	Log normal
Biogeographic	Pamir uplift in western China	10	*Teratoscincus scincus* + *Teratoscincus roborowskii*	Log normal
Secondary	*Root*	252-257	Lepidosauria + Archosauria	Normal

Supplementary Table 5. Correlation matrix used in paleo-niche analyses.

Last inter-Glacial 140k

N = 11736091	LIG CCSM4bi10	LIG CCSM4bi11	LIG CCSM4bi15	LIG CCSM4bi18	LIG CCSM4bi19	LIG CCSM4bi2	LIG CCSM4bi3	LIG CCSM4bi4	LIG CCSM4bi8	LIG CCSM4bi9
LIG CCSM4bi10	1	0.724172	0.735088	-0.092573	-0.262656	0.378959	-0.066035	0.100989	0.789193	0.828281
LIG CCSM4bi11	0.724172	1	0.421286	0.49065	0.262241	-0.249153	0.600363	-0.612439	0.840762	0.667614
LIG CCSM4bi15	0.735088	0.421286	1	-0.336384	-0.505221	0.541832	-0.258644	0.233474	0.595264	0.628195
LIG CCSM4bi18	-0.092573	0.49065	-0.336384	1	0.779761	-0.754658	0.801256	-0.815749	0.264449	-0.034536
LIG CCSM4bi19	-0.262656	0.262241	-0.505221	0.779761	1	-0.76009	0.708885	-0.680086	-0.009533	-0.073606
LIG CCSM4bi2	0.378959	-0.249153	0.541832	-0.754658	-0.76009	1	-0.7384	0.794282	0.080096	0.197198
LIG CCSM4bi3	-0.066035	0.600363	-0.258644	0.801256	0.708885	-0.7384	1	-0.943843	0.296416	0.047441
LIG CCSM4bi4	0.100989	-0.612439	0.233474	-0.815749	-0.680086	0.794282	-0.943843	1	-0.31325	-0.008451
LIG CCSM4bi8	0.789193	0.840762	0.595264	0.264449	-0.009533	0.080096	0.296416	-0.31325	1	0.586882
LIG CCSM4bi9	0.828281	0.667614	0.628195	-0.034536	-0.073606	0.197198	0.047441	-0.008451	0.586882	1

Last Glacial Maximum 22k

N=665518	LGM mrlgmbi10	LGM mrlgmbi15	LGM mrlgmbi16	LGM mrlgmbi17	LGM mrlgmbi18	LGM mrlgmbi1	LGM mrlgmbi2	LGM mrlgmbi3	LGM mrlgmbi4	LGM mrlgmbi8	LGM mrlgmbi9
LGM mrlgmbi10	1	0.557477	0.010775	-0.321544	-0.113248	0.817977	0.492851	-0.082457	0.211618	0.829627	0.313717
LGM mrlgmbi15	0.557477	1	0.311029	-0.453908	-0.001052	0.639781	0.257579	0.040213	-0.177976	0.496599	0.340827

LGM mrlgmbi16	0.010775	0.311029	1	0.51167	0.786718	0.491601	-0.477451	0.771179	-0.833735	0.188555	0.534642
LGM mrlgmbi17	-0.321544	-0.453908	0.51167	1	0.744513	0.018088	-0.639571	0.629834	-0.545692	-0.103291	0.191116
LGM mrlgmbi18	-0.113248	-0.001052	0.786718	0.744513	1	0.302915	-0.579195	0.701469	-0.684792	0.15103	0.282595
LGM mrlgmbi1	0.817977	0.639781	0.491601	0.018088	0.302915	1	0.058005	0.454929	-0.385808	0.811385	0.631049
LGM mrlgmbi2	0.492851	0.257579	-0.477451	-0.639571	-0.579195	0.058005	1	-0.533511	0.693553	0.233915	-0.235638
LGM mrlgmbi3	-0.082457	0.040213	0.771179	0.629834	0.701469	0.454929	-0.533511	1	-0.909708	0.137992	0.584636
LGM mrlgmbi4	0.211618	-0.177976	-0.833735	-0.545692	-0.684792	-0.385808	0.693553	-0.909708	1	-0.027929	-0.59609
LGM mrlgmbi8	0.829627	0.496599	0.188555	-0.103291	0.15103	0.811385	0.233915	0.137992	-0.027929	1	0.17123
LGM mrlgmbi9	0.313717	0.340827	0.534642	0.191116	0.282595	0.631049	-0.235638	0.584636	-0.59609	0.17123	1

mH 6k

N=10546450	mH CCSM4bi15	mH CCSM4bi17	mH CCSM4bi18	mH CCSM4bi2	mH CCSM4bi3	mH CCSM4bi4	mH CCSM4bi6	mH CCSM4bi8	mH CCSM4bi9
mH CCSM4bi15	1	-0.375648	0.138972	0.073908	0.112774	-0.217708	0.566802	0.507654	0.564245
mH CCSM4bi17	-0.375648	1	0.736739	-0.58221	0.679209	-0.578214	0.323029	-0.202859	0.03031
mH CCSM4bi18	0.138972	0.736739	1	-0.624005	0.782127	-0.79111	0.671276	0.13336	0.226462
mH CCSM4bi2	0.073908	-0.58221	-0.624005	1	-0.49749	0.793199	-0.506252	0.383415	-0.209383
mH CCSM4bi3	0.112774	0.679209	0.782127	-0.49749	1	-0.845393	0.71481	0.057366	0.399877
mH CCSM4bi4	-0.217708	-0.578214	-0.79111	0.793199	-0.845393	1	-0.738461	0.149042	-0.403974
mH CCSM4bi6	0.566802	0.323029	0.671276	-0.506252	0.71481	-0.738461	1	0.388031	0.710768

mH CCSM4bi8	0.507654	-0.202859	0.13336	0.383415	0.057366	0.149042	0.388031	1	0.159645
mH CCSM4bi9	0.564245	0.03031	0.226462	-0.209383	0.399877	-0.403974	0.710768	0.159645	1

Supplementary Table 6. *Heteronotia* niche correlations.

N=6184	**Bio_01**	**Bio_02**	**Bio_03**	**Bio_04**	**Bio_05**	**Bio_09**	**Bio_12**	**Bio_15**	**Bio_17**	**Bio_18**
Bio_01	1.00	0.08	0.64	-0.46	0.54	0.83	0.22	0.84	-0.68	-0.03
Bio_02	0.08	1.00	-0.22	0.64	0.80	-0.28	-0.71	0.01	-0.31	-0.73
Bio_03	0.64	-0.22	1.00	-0.87	-0.10	0.85	0.66	0.66	-0.42	0.38
Bio_04	-0.46	0.64	-0.87	1.00	0.46	-0.81	-0.87	-0.53	0.15	-0.67
Bio_05	0.54	0.80	-0.10	0.46	1.00	0.08	-0.60	0.34	-0.56	-0.68
Bio_09	0.83	-0.28	0.85	-0.81	0.08	1.00	0.58	0.78	-0.45	0.36
Bio_12	0.22	-0.71	0.66	-0.87	-0.60	0.58	1.00	0.28	0.23	0.83
Bio_15	0.84	0.01	0.66	-0.53	0.34	0.78	0.28	1.00	-0.69	0.16
Bio_17	-0.68	-0.31	-0.42	0.15	-0.56	-0.45	0.23	-0.69	1.00	0.37
Bio_18	-0.03	-0.73	0.38	-0.67	-0.68	0.36	0.83	0.16	0.37	1.00

Appendix III

Lialis burtonis (RADseq code)

```
cd c_filtered

clone_filter -f ../ddRADseq_raw/INDX1_ATCACG/JT-ATCACG_S14
2_L004_R1_001.fastq.gz --inline_null --oligo_len_1 8 -o ./
 -i gzfastq -D
clone_filter -f ../ddRADseq_raw/INDX4_TGACCA/JT-TGACCA_S14
3_L004_R1_001.fastq.gz --inline_null --oligo_len_1 8 -o ./
 -i gzfastq -D
clone_filter -f ../ddRADseq_raw/INDX6_GCCAAT/JT-GCCAAT_S14
4_L004_R1_001.fastq.gz --inline_null --oligo_len_1 8 -o ./
 -i gzfastq -D
clone_filter -f ../ddRADseq_raw/INDX7_CAGATC/JT-CAGATC_S14
5_L004_R1_001.fastq.gz --inline_null --oligo_len_1 8 -o ./
 -i gzfastq -D
clone_filter -f ../ddRADseq_raw/INDX11_GGCTAC/JT-GGCTAC_S1
46_L004_R1_001.fastq.gz --inline_null --oligo_len_1 8 -o .
/ -i gzfastq -D
clone_filter -f ../ddRADseq_raw/INDX15_ATGTCA/JT-ATGTCA_S1
47_L004_R1_001.fastq.gz --inline_null --oligo_len_1 8 -o .
/ -i gzfastq -D
clone_filter -f ../ddRADseq_raw/INDX16_CCGTCC/JT-CCGTCC_S1
48_L004_R1_001.fastq.gz --inline_null --oligo_len_1 8 -o .
/ -i gzfastq -D

cd demultiplex_Lialis
```

```
process_radtags -f ../c_filtered/JT-ATCACG_S142_L004_R1_001.fq.gz -b ../Lialis_Barcodes/STACKS_INDX1.txt -r --inline_null -o ./ -c -q --renz_1 sbfI -i gzfastq
process_radtags -f ../c_filtered/JT-TGACCA_S143_L004_R1_001.fq.gz -b ../Lialis_Barcodes/STACKS_INDX4.txt -r --inline_null -o ./ -c -q --renz_1 sbfI -i gzfastq
process_radtags -f ../c_filtered/JT-GCCAAT_S144_L004_R1_001.fq.gz -b ../Lialis_Barcodes/STACKS_INDX6.txt -r --inline_null -o ./ -c -q --renz_1 sbfI -i gzfastq
process_radtags -f ../c_filtered/JT-CAGATC_S145_L004_R1_001.fq.gz -b ../Lialis_Barcodes/STACKS_INDX7.txt -r --inline_null -o ./ -c -q --renz_1 sbfI -i gzfastq
process_radtags -f ../c_filtered/JT-GGCTAC_S146_L004_R1_001.fq.gz -b ../Lialis_Barcodes/STACKS_INDX11.txt -r --inline_null -o ./ -c -q --renz_1 sbfI -i gzfastq
process_radtags -f ../c_filtered/JT-ATGTCA_S147_L004_R1_001.fq.gz -b ../Lialis_Barcodes/STACKS_INDX15.txt -r --inline_null -o ./ -c -q --renz_1 sbfI -i gzfastq
process_radtags -f ../c_filtered/JT-CCGTCC_S148_L004_R1_001.fq.gz -b ../Lialis_Barcodes/STACKS_INDX16.txt -r --inline_null -o ./ -c -q --renz_1 sbfI -i gzfastq

# RUN IPYRAD

ipyrad -p params-lialis_combine.txt -s 1234567
```

Lialis ADEGENET

```
# Set WD
setwd("C:/Users/JET/Desktop/Lialis_Ipyrad_combine/ADEGENTET")
# Load needed packages
library("ape")
library("pegas")
library("seqinr")
library("ggplot2")

# Load ADEDGENET
library("adegenet", lib.loc="~/R/win-library/3.5")

# Read files

Lialis_Test <- read.structure("../Lialis_refiltered/lialis_refilter.u.str", n.ind = 81, n.loc = 6838,
                col.lab = 0, col.pop = NULL, col.others = NULL,
                row.marknames = NULL, NA.char = "-9", pop = NULL, sep = NULL,
                ask = TRUE, quiet = FALSE)

Lialis_Genid <- import2genind("../Lialis_refiltered/lialis_refilter_popID.u.str", n.ind = 81, n.loc = 6838,
```

```
                                col.lab = 0, col.pop = 2, co
l.others = 0,
                                row.marknames = 0, NA.char =
 "-9", sep = NULL,
                                ask = TRUE, quiet = FALSE)

# DAPC with 6 retained PCs
Lialis_dapc_6PC <- dapc(Lialis_Genid)
# DAPC with 10 retained PCs
Lialis_dapc_10PC <- dapc(Lialis_Genid)
# DAPC with 20 retained PCs
Lialis_dapc_20PC <- dapc(Lialis_Genid)
# DAPC with 30 retained PCs
Lialis_dapc_30PC <- dapc(Lialis_Genid)
# DAPC with 40 retained PCs
Lialis_dapc_40PC <- dapc(Lialis_Genid)
# DAPC with 60 retained PCs
Lialis_dapc_60pc <- dapc(Lialis_Genid)
# DAPC with 77 retained PCs
Lialis_dapc77pc <- dapc(Lialis_Genid)


Lialis_PCA <- scatter(Lialis_dapc_6PC)
Lialis_PCA <- scatter(Lialis_dapc_10PC)
Lialis_PCA <- scatter(Lialis_dapc_20PC)
Lialis_PCA <- scatter(Lialis_dapc_30PC)
Lialis_PCA <- scatter(Lialis_dapc_40PC)
Lialis_PCA <- scatter(Lialis_dapc_60pc)
```

```
Lialis_PCA <- scatter(Lialis_dapc77pc)

# Web browers plot editor
adegenetServer(what = "DAPC")

myCol <- c("red","green","blue","magenta")
Lialis_PCA <- scatter(Lialis_dapc_40PC,posi.da="bottomrigh
t", bg="white",
                      col=myCol, scree.pca=TRUE,
                      posi.pca="bottomleft")

# PCA plot
Lialis_PCA <- scatter(Lialis_dapc_40PC,scree.pca=TRUE, scr
ee.da=TRUE,
                      bg="white", pch=20, cell=1.75, cstar
=1, col=myCol,
                      solid=.4, cex=3, clab=0, leg=TRUE, t
xt.leg=paste("Cluster",1:4))

# Discriminant function plot
Lialis_Discrim <- scatter(Lialis_dapc_40PC,1,1, col=myCol,
 bg="white",
        scree.da=FALSE, legend=TRUE, solid=.4)

# Plot alleles with the highest contribution
Lialis_contrib <- loadingplot(Lialis_dapc_40PC$var.contr,
axis=2,
                       thres=.07, lab.jitter=1)
```

```
## Group membership interpretation
# Posterior dimensions and matrix
class(Lialis_dapc_40PC$posterior)
dim(Lialis_dapc_40PC$posterior)
round(head(Lialis_dapc_40PC$posterior),4)
summary(Lialis_dapc_40PC)

# Membership proportion among lineages
assignplot(Lialis_dapc_40PC) # Blue pluses are concordant
with a priori asignments from ND2 lineages

# STRUCTURE-like visualazation in a Genotype composition p
lot
compoplot(Lialis_dapc_40PC, posi="bottomright", col.pal=my
Col,
          txt.leg=paste("Cluster", 1:4), n.col=1, xlab="In
dividuals")
### Group stability tests

temp <- summary(dapc(Lialis_Genid, n.da=40, n.pca = 40))$a
ssign.per.pop*100
par(mar=c(4.5,7.5,1,1))
barplot(temp, xlab="% of reassignment to actual Lineage",
horiz=TRUE, las=1)

#alpha scores
Lialis_dapc2 <- dapc(Lialis_Genid, n.da=100, n.pca=40)
```

```
Lialis_alpha <- a.score(Lialis_dapc2)
names(temp)
temp <- optim.a.score(Lialis_dapc2)

#Cross-validation
mat <- tab(Lialis_Genid, NA.method="mean")
grp <- pop(Lialis_Genid)
Lialis_xval <- xvalDapc(mat, grp, n.pca.max = 300, trainin
g.set = 0.9,
                  result = "groupMean", center = TRUE, scal
e = FALSE,
                  n.pca = NULL, n.rep = 30, xval.plot = TRU
E)
```

Lialis ADMIXTURE

```
#ADMIXTURE Commands
#Filter VCF
plink --vcf ../lialis_refilter.vcf --make-bed --allow-extr
a-chr --mind .5 --geno .1 --out lialis_refilter_loci_indv
plink --vcf ../lialis_refilter.vcf --make-bed --allow-extr
a-chr --mind .5 --geno .5 --out lialis_refilter_loci_indv_
50
plink --vcf ../lialis_refilter.vcf --make-bed --allow-extr
a-chr --mind .9 --geno .5 --out lialis_refilter_loci_50_in
```

```
dv_90


#ADMIXTURE Commands
admixture -B lialis_refilter_loci_indv.bed [K value] -j70

#for loop to run all K values in admxiture using bootstrap
s (You may only need to use the cv command below)
for K in 2 3 4 5 6 7 8 9 10; do admixture -B lialis_refilt
er_loci_indv.bed $K | tee log${K}.out; done

# For loop to run cross-validation in admixture
for K in 2 3 4 5 6 7 8 9 10; do admixture --cv lialis_refi
lter_loci_indv.bed $K | tee log${K}.out; done

# Grab CV values to plot, then redirect to a file
grep -h CV log*.out > K_Cross_Validation.txt

#Remove Indonesia sample with no coordinates
plink --vcf ../lialis_refilter_outfiles/lialis_refilter.vc
f --make-bed --allow-extra-chr --remove RemoveTG2108 --min
d .9 --geno .5 --out lialis_refilter_loci_50_indv_90_noTG2
108

#Options in effect:
#  --allow-extra-chr
#  --geno .5
#  --make-bed
```

```
#  --mind .9
#  --out lialis_refilter_loci_50_indv_90_noTG2108
#  --remove RemoveTG2108
#  --vcf ../lialis_refilter_outfiles/lialis_refilter.vcf
#
#120876 MB RAM detected; reserving 60438 MB for main works
pace.
#--vcf: lialis_refilter_loci_50_indv_90_noTG2108-temporary
.bed +
#lialis_refilter_loci_50_indv_90_noTG2108-temporary.bim +
#lialis_refilter_loci_50_indv_90_noTG2108-temporary.fam wr
itten.
#72532 variants loaded from .bim file.
#81 people (0 males, 0 females, 81 ambiguous) loaded from
.fam.
#Ambiguous sex IDs written to lialis_refilter_loci_50_indv
_90_noTG2108.nosex .
#--remove: 80 people remaining.
#9 people removed due to missing genotype data (--mind).
#IDs written to lialis_refilter_loci_50_indv_90_noTG2108.i
rem .
#Using 1 thread (no multithreaded calculations invoked).
#Before main variant filters, 71 founders and 0 nonfounder
s present.
#Calculating allele frequencies... done.
#Total genotyping rate in remaining samples is 0.671169.
#4053 variants removed due to missing genotype data (--gen
o).
```

```
#68479 variants and 71 people pass filters and QC.
#Note: No phenotypes present.
#--make-bed to lialis_refilter_loci_50_indv_90_noTG2108.be
d +
#lialis_refilter_loci_50_indv_90_noTG2108.bim +
#lialis_refilter_loci_50_indv_90_noTG2108.fam ... done.
```

```
# Convert Q-matrix into a rough dendrogram (coded by numer
ic order, must match order *script in the works to automat
e processes)
admix <- read.table('lialis_refilter_loci_50_indv_90.4.Q',
header=FALSE)
d <- dist(admix)
h <- hclust(d)
dend <- as.dendrogram(h)
plot(dend)
```

Lialis EEMS

```
/home/lepidodactylus/Desktop/Phylogenetic_tools/eems-maste
r/runeems_snps/src/runeems_snps --params EEMS_params/param
s-chain1.ini
/home/lepidodactylus/Desktop/Phylogenetic_tools/eems-maste
r/runeems_snps/src/runeems_snps --params EEMS_params/param
s-chain2.ini
/home/lepidodactylus/Desktop/Phylogenetic_tools/eems-maste
r/runeems_snps/src/runeems_snps --params EEMS_params/param
```

```
s-chain3.ini
```

Plot EEMS

```
# Set working directory
setwd("/Users/jamestmcquillan/Downloads/Liali/Users/jamestmcquillan/Downloads/Lialis_ADMIXTURE_Loose/Lialis_EEMS/mcmc3")

# Libraries

library(Rcpp)
library(RcppEigen)
library(raster)
library(rgeos)
library(sp)
library("rEEMSplots")
library("rgdal")
library("rworldmap")
library("rworldxtra")
library("RColorBrewer")
library("deldir")

# call files for plotting

Lialis_data_path <- ".."
#Lialis_files <- lapply(Lialis_data_path, read.delim)
```

```
eems_results <- file.path(Lialis_data_path, "lialis_eems")
name_figures <- file.path(path.expand("~"), "lialis_eems")

map_world <- getMap()
map_oceania <- map_world[which(map_world@data$continent ==
 "Oceania"), ]

map_oceania <- spTransform(map_oceania, CRSobj = CRS(proje
ction_mercator))
eems.plots(mcmcpath = c("~/Downloads/eems-master/plotting/
rEEMSplots/inst/All_Lineage/mcmc1","~/Downloads/eems-maste
r/plotting/rEEMSplots/inst/All_Lineage/mcmc2","~/Downloads
/eems-master/plotting/rEEMSplots/inst/All_Lineage/mcmc3"),
           plotpath = paste0(name_figures, "RBL-projected-
demes"),
           longlat = FALSE,
           projection.in = projection_none,
           projection.out = projection_mercator,
           add.grid = TRUE,
           add.demes = TRUE,
           min.cex.demes = 0.5,
           max.cex.demes = 1.5,
           eems.colors = brewer.pal(11, "RdBu"))


eems.plots(mcmcpath = "../lialis_eems", plotpath = paste0(
name_figures, "-default"), longlat = TRUE)
```

Lialis TETRAD

```
import ipyrad.analysis as ipa
import ipyparallel as ipp
import toytree

## connect to a cluster
ipyclient = ipp.Client()
print("connected to {} cores".format(len(ipyclient)))


# ### Run tetrad
## initiate a tetrad object
tet = ipa.tetrad(
    name="pedic-full",
    seqfile="~/Desktop/Lialis/ipyrad_analysis_combine/ipyr
ad_scripts/assembly.snps.phy",
    mapfile="~/Desktop/Lialis/ipyrad_analysis_combine/ipyr
ad_scripts/assembly.snps.map",
    nboots=100,
    )

## run tetrad on the cluster
tet.run(ipyclient=ipyclient)

# ### Plot the tree
## plot the resulting unrooted tree
```

```
import toytree
tre = toytree.tree(tet.trees.nhx)
tre.draw(
    width=350,
    node_labels=tre.get_node_values("support"),
    );

## save the tree as a pdf
import toyplot.pdf
toyplot.pdf.render(canvas, "analysis-tetrad/tetrad-tree.pd
f")


# ### What does *tetrad* do differently from *svdquartets*
?
#
# Not too much currently. But we have plans to expand it.
Importantly, however, the code is open source meaning that
 anybody can read it and contribute it, which is not the c
ase for Paup\*. *tetrad* is also easier to install using c
onda and therefore easier to setup on an HPC cluster or lo
cal machine, and it can be parallelized across an arbitrar
ily large number of compute nodes while retaining a super
small memory footprint.
```

Lialis TreeMix

```
treemix -i ../lialis_refilter_loci_50_indv_90.p.treemix.gz
 -noss -bootstrap 500 -k 500 -o MyTest_m4_noss_boot -m 64
```

This ploting function can be found at https://bitbucket.org/nygcresearch/treemix It was created by JK Pickrell and JK Pritchard.

```
source("~/Applications/treemix-1.13/src/plotting_funcs.R")

# From Prichard et al. 2014
# functions for plotting a tree
#

library(RColorBrewer)
set_y_coords = function(d){
    i = which(d[,3]=="ROOT")
    y = d[i,8]/ (d[i,8]+d[i,10])
    d[i,]$y = 1-y
    d[i,]$ymin = 0
    d[i,]$ymax = 1
    c1 = d[i,7]
    c2 = d[i,9]
    ni = which(d[,1]==c1)
    ny = d[ni,8]/ (d[ni,8]+d[ni,10])
        d[ni,]$ymin = 1-y
```

```
    d[ni,]$ymax = 1
    d[ni,]$y = 1- ny*(y)

        ni = which(d[,1]==c2)
        ny = d[ni,8]/ (d[ni,8]+d[ni,10])
        d[ni,]$ymin = 0
        d[ni,]$ymax = 1-y
        d[ni,]$y = (1-y)-ny*(1-y)

    for (j in 1:nrow(d)){
        d = set_y_coord(d, j)
    }
    return(d)
}

set_y_coord = function(d, i){
    index = d[i,1]
    parent = d[i,6]
    if (!is.na(d[i,]$y)){
        return(d)
    }
    tmp = d[d[,1] == parent,]
    if ( is.na(tmp[1,]$y)){
        d = set_y_coord(d, which(d[,1]==parent))
        tmp = d[d[,1]== parent,]
    }
    py = tmp[1,]$y
    pymin = tmp[1,]$ymin
```

```
    pymax = tmp[1,]$ymax
    f = d[i,8]/( d[i,8]+d[i,10])
    #print (paste(i, index, py, pymin, pymax, f))
    if (tmp[1,7] == index){
        d[i,]$ymin = py
        d[i,]$ymax = pymax
        d[i,]$y = pymax-f*(pymax-py)
        if (d[i,5]== "TIP"){
            d[i,]$y = (py+pymax)/2
        }
    }
    else{
        d[i,]$ymin = pymin
        d[i,]$ymax = py
        d[i,]$y = py-f*(py-pymin)
                if (d[i,5]== "TIP"){
                        d[i,]$y = (pymin+py)/2
                }

    }
    return(d)
}


set_x_coords = function(d, e){
    i = which(d[,3]=="ROOT")
    index = d[i,1]
        d[i,]$x = 0
```

```
        c1 = d[i,7]
        c2 = d[i,9]
        ni = which(d[,1]==c1)
    tmpx =  e[e[,1]==index & e[,2] == c1,3]
    if (length(tmpx) == 0){
        tmp = e[e[,1] == index,]
        tmpc1 = tmp[1,2]
        if ( d[d[,1]==tmpc1,4] != "MIG"){
            tmpc1 = tmp[2,2]
        }
        tmpx = get_dist_to_nmig(d, e, index, tmpc1)
    }
    if(tmpx < 0){
        tmpx = 0
    }
        d[ni,]$x = tmpx

        ni = which(d[,1]==c2)
        tmpx =  e[e[,1]==index & e[,2] == c2,3]
    if (length(tmpx) == 0){
            tmp = e[e[,1] == index,]
                tmpc2 = tmp[2,2]
                if ( d[d[,1]==tmpc2,4] != "MIG"){
                        tmpc2 = tmp[1,2]
                }
                tmpx = get_dist_to_nmig(d, e, index, tmpc2
)
    }
```

```
        if(tmpx < 0){
                tmpx = 0
        }
        d[ni,]$x = tmpx

        for (j in 1:nrow(d)){
                d = set_x_coord(d, e, j)
        }
        return(d)
    print(d)
}


set_x_coord = function(d, e, i){
    index = d[i,1]
        parent = d[i,6]
        if (!is.na(d[i,]$x)){
                return(d)
        }
        tmp = d[d[,1] == parent,]
        if ( is.na(tmp[1,]$x)){
                d = set_x_coord(d, e, which(d[,1]==parent)
)
                tmp = d[d[,1]== parent,]
        }
    #print (paste(parent, index))
    tmpx =  e[e[,1]==parent & e[,2] == index,3]
    if (length(tmpx) == 0){
```

```
                tmp2 = e[e[,1] == parent,]
                tmpc2 = tmp2[2,2]
        #print
                if ( d[d[,1]==tmpc2,4] != "MIG"){
                        tmpc2 = tmp2[1,2]
                }
        tmpx = get_dist_to_nmig(d, e, parent, tmpc2)
    }
        if(tmpx < 0){
                tmpx = 0
        }
    d[i,]$x = tmp[1,]$x+ tmpx
    return(d)
}

plot_tree_internal = function(d, e, o = NA, cex = 1, disp
= 0.005, plus = 0.005, arrow = 0.05, ybar = 0.01, scale =
T, mbar = T, mse = 0.01, plotmig = T, plotnames = T, xmin
= 0, lwd = 1, font = 1){
    plot(d$x, d$y, axes = F, ylab = "", xlab = "Drift para
meter", xlim = c(xmin, max(d$x)+plus), pch = "")
    axis(1)
    mw = max(e[e[,5]=="MIG",4])
    mcols = rev(heat.colors(150))
    for(i in 1:nrow(e)){
        col = "black"
        if (e[i,5] == "MIG"){
            w = floor(e[i,4]*200)+50
```

```
            if (mw > 0.5){
                w = floor(e[i,4]*100)+50
            }
            col = mcols[w]
            if (is.na(col)){
                col = "blue"
            }
        }
        v1 = d[d[,1] == e[i,1],]
        v2 = d[d[,1] == e[i,2],]
        if (e[i,5] == "MIG"){
            if (plotmig){
            arrows( v1[1,]$x, v1[1,]$y, v2[1,]$x, v2[1,]$y
, col = col, length = arrow)
            }
        }
        else{
            lines( c(v1[1,]$x, v2[1,]$x), c(v1[1,]$y, v2[1
,]$y), col = col, lwd = lwd)
        }
    }
    tmp = d[d[,5] == "TIP",]
    print(tmp$x)
    print(disp)
    if ( !is.na(o)){
        for(i in 1:nrow(tmp)){
            tcol = o[o[,1] == tmp[i,2],2]
            if(plotnames){
```

```
                #print(tmp[i,2])
                text(tmp[i,]$x+disp, tmp[i,]$y, labels = t
mp[i,2], adj = 0, cex = cex, col  = tcol, font = font)
            }
        }
    }
    else{
        if (plotnames){
        text(tmp$x+disp, tmp$y, labels = tmp[,2], adj = 0,
 cex = cex, font = font)
        }
    }
    if (scale){
    print (paste("mse", mse))
        lines(c(0, mse*10), c(ybar, ybar))
    text( 0, ybar - 0.04, lab = "10 s.e.", adj = 0, cex  =
 0.8)
    lines( c(0, 0), c( ybar - 0.01, ybar+0.01))
    lines( c(mse*10, mse*10), c(ybar- 0.01, ybar+ 0.01))
    }
        if (mbar){
                mcols = rev( heat.colors(150) )
                mcols = mcols[50:length(mcols)]
                ymi = ybar+0.15
                yma = ybar+0.35
                l = 0.2
                w = l/100
                xma = max(d$x/20)
```

```
                rect( rep(0, 100), ymi+(0:99)*w, rep(xma,
100), ymi+(1:100)*w, col = mcols, border = mcols)
                text(xma+disp, ymi, lab = "0", adj = 0, ce
x = 0.7)
        if ( mw >0.5){ text(xma+disp, yma, lab = "1", adj
= 0, cex = 0.7)}
                else{
            text(xma+disp, yma, lab = "0.5", adj = 0, cex
=0.7)
        }
        text(0, yma+0.06, lab = "Migration", adj = 0 , cex
 = 0.6)
        text(0, yma+0.03, lab = "weight", adj = 0 , cex =
0.6)
        }
}

set_mig_coords = function(d, e){
    for (j in 1:nrow(d)){
        if (d[j,4] == "MIG"){
            p = d[d[,1] == d[j,6],]
            c = d[d[,1] == d[j,7],]
            tmpe = e[e[,1] == d[j,1],]
            y1 = p[1,]$y
            y2 = c[1,]$y
            x1 = p[1,]$x
            x2 = c[1,]$x
```

```
            mf = tmpe[1,6]
            if (is.nan(mf)){
                mf = 0
            }
            #d[j,]$y = (y1+y2)* mf
                        #d[j,]$x = (x1+x2) *mf
                        d[j,]$y = y1+(y2-y1)* mf
            print(paste(mf, x1, x2))
                        d[j,]$x = x1+(x2-x1) *mf
        }

    }
    return(d)

}

get_f = function(stem){
    d = paste(stem, ".cov.gz", sep = "")
    d2 = paste(stem, ".modelcov.gz", sep = "")
    d = read.table(gzfile(d), as.is = T, comment.char = ""
, quote = "")
    d2 = read.table(gzfile(d2), as.is = T, comment.char =
"", quote = "")
    d = d[order(names(d)), order(names(d))]
    d2 = d2[order(names(d2)), order(names(d2))]
    tmpcf = vector()
        tmpmcf = vector()
        for (j in 1:nrow(d)){
```

```
                for (k in (j+1):nrow(d)){
                        tmpcf = append(tmpcf, d[j,k])
                        tmpmcf = append(tmpmcf, d[j,k] - d
2[j,k])
                }
        }
        tmpv = var(tmpmcf)/var(tmpcf)
    return(1-tmpv)

}

plot_tree = function(stem, o = NA, cex = 1, disp = 0.003,
plus = 0.01, flip = vector(), arrow = 0.05, scale = T, yba
r = 0.1, mbar = T, plotmig = T, plotnames = T, xmin = 0, l
wd = 1, font = 1){
    d = paste(stem, ".vertices.gz", sep = "")
    e = paste(stem, ".edges.gz", sep = "")
    se = paste(stem, ".covse.gz", sep = "")
    d = read.table(gzfile(d), as.is = T, comment.char = ""
, quote = "")
    e = read.table(gzfile(e), as.is  = T, comment.char = "
", quote = "")
    if (!is.na(o)){
        o = read.table(o, as.is = T, comment.char = "", qu
ote = "")
    }
    e[,3] = e[,3]*e[,4]
    e[,3] = e[,3]*e[,4]
```

```
    se = read.table(gzfile(se), as.is = T, comment.char =
"", quote = "")
    m1 = apply(se, 1, mean)
    m = mean(m1)
    #m = 0
    for(i in 1:length(flip)){
        d = flip_node(d, flip[i])
    }
    d$x = "NA"
    d$y = "NA"
    d$ymin = "NA"
    d$ymax = "NA"
    d$x = as.numeric(d$x)
    d$y = as.numeric(d$y)
    d$ymin = as.numeric(d$ymin)
    d$ymax = as.numeric(d$ymax)

    d = set_y_coords(d)
    d = set_x_coords(d, e)
    print(d)
    d = set_mig_coords(d, e)
    plot_tree_internal(d, e, o = o, cex = cex, xmin = xmin
, disp = disp, plus = plus, arrow = arrow, ybar = ybar, mb
ar = mbar, mse = m, scale = scale, plotmig = plotmig, plot
names = plotnames, lwd = lwd, font = font)
    return(list( d= d, e = e))
}
```

```
get_dist_to_nmig = function(d, e, n1, n2){
    toreturn = e[e[,1] == n1 & e[,2] == n2,3]
    #print(toreturn)
    while ( d[d[,1] ==n2,4] == "MIG"){
        tmp = e[e[,1] == n2 & e[,5] == "NOT_MIG",]
        toreturn = toreturn+tmp[1,3]
        n2 = tmp[1,2]
    }
    return(toreturn)
}

flip_node = function(d, n){
    i = which(d[,1] == n)
    t1 = d[i,7]
    t2 = d[i,8]
    d[i,7] = d[i,9]
    d[i,8] = d[i,10]
    d[i,9] = t1
    d[i,10] = t2
    return(d)

}

plot_modelcov = function(stem, pop_order, min = -0.009, ma
x = 0.009, cex = 1, usemax = T){
        c = read.table(gzfile(paste(stem, ".modelcov.gz",
sep = "")), as.is = T, head = T)
```

```
        o = read.table(pop_order, as.is = T, comment.char
= "", quote = "")


        toplot = data.frame(matrix(nrow = nrow(c), ncol =
ncol(c)))
        for(i in 1:nrow(o)){
                for( j in 1:nrow(o)){

                        toplot[i, j] = c[which(names(c)==o
[i,1]), which(names(c)==o[j,1])]
                }
        }
        if (usemax){
                m1 = max(abs(toplot))
                max = m1*1.1
                min = -(m1*1.1)
        }
        names(toplot) = o[,1]
        plot_resid_internal(toplot, max = max, min = min)
}



plot_cov = function(stem, pop_order, min = -0.009, max = 0
.009, cex = 1, usemax = T, wcols = ""){
        c = read.table(gzfile(paste(stem, ".cov.gz", sep =
 "")), as.is = T, head = T)
```

```
        o = read.table(pop_order, as.is = T)


        toplot = data.frame(matrix(nrow = nrow(c), ncol =
ncol(c)))
        for(i in 1:nrow(o)){
                for( j in 1:nrow(o)){

                        toplot[i, j] = c[which(names(c)==o
[i,1]), which(names(c)==o[j,1])]
                }
        }
        if (usemax){
                m1 = max(abs(toplot))
                max = m1*1.1
                min = 0
        }
        names(toplot) = o[,1]
        plot_cov_internal(toplot, max = max, min = min, wc
ols = wcols, o = o, cex = cex)
}


plot_resid = function(stem, pop_order, min = -0.009, max =
 0.009, cex = 1, usemax = T, wcols = "r"){
    c = read.table(gzfile(paste(stem, ".cov.gz", sep = "")
), as.is = T, head = T, quote = "", comment.char = "")
    m = read.table(gzfile(paste(stem, ".modelcov.gz", sep
```

```
= "")), as.is = T, head = T, quote = "", comment.char = ""
)
    names(c) = rownames(c)
    names(m) = rownames(m)
    o = read.table(pop_order, as.is = T, comment.char = ""
, quote = "")
    se = read.table(gzfile(paste(stem, ".covse.gz", sep =
"")), as.is = T, head = T, quote = "", comment.char = "")
    mse = apply(se, 1, mean)
    mse = mean(mse)
    print(mse)
    c = c[order(names(c)), order(names(c))]
    m = m[order(names(m)), order(names(m))]
    tmp = c -m
    #tmp = m - c
    #tmp = (m-c)/m
    #print(tmp)
    toplot = data.frame(matrix(nrow = nrow(tmp), ncol = nc
ol(tmp)))
    for(i in 1:nrow(o)){
            for( j in 1:nrow(o)){
            #print(paste(o[i,1], o[j,1]))
            if (o[i,1] %in% names(tmp) ==F){
                print(paste("not found", o[i,1]))
            }
            if (o[j,1] %in% names(tmp) ==F){
                print(paste("not found", o[j,1]))
            }
```

```
                    toplot[i, j] = tmp[which(names(tmp)==o
[i,1]), which(names(tmp)==o[j,1])]
            }
    }
    #print(toplot)
    if (usemax){
        m1 = max(abs(toplot), na.rm = T)
        max = m1*1.02
        min = -(m1*1.02)
    }
    print("here")
    names(toplot) = o[,1]
    toreturn = plot_resid_internal(toplot, max = max, min
= min, wcols = wcols, mse = mse, o = o, cex = cex)
    return(toreturn)
}

plot_cov_internal = function(d, o = NA, max = 0.009, min =
 -0.009, cex =0.5, wcols = "", mse = 5){
        npop = nrow(d)
        width = 1/npop
        height = 1/npop
        colors = brewer.pal(9, "Spectral")
        colors = c("red", "orange","yellow", "white", "gre
en", "blue", "black")
        pal = colorRampPalette(colors)
        ncol = 80
        cols = pal(ncol)
```

```
        plot("NA", xlim = c(0, 1), ylim = c(0, 1), axes =
F, xlab = "", ylab = "")
        for (i in 1:npop){
                for( j in 1:i){
                        v = d[i,j]
                        col= "white"
                        if (v < 0){
                                if (wcols == "rb"){
                                col = rgb(0, 0, 1, v/min)
                                }
                                else{
                                #col = rgb(0, 0, 1, 0.1+0.
9*(v/min))
                                col = cols[ncol/2-floor( (
v/min)*(ncol/2))]
                                }
                        }
                        else{
                                if (wcols == "rb"){
                                col = rgb(1, 0, 0, v/max)
                                }
                                else{
                                #col = rgb(1, 0, 0, 0.1+0.
9*(v/max))
                                col = cols[ceiling((v/max)
*(ncol))]
                                }
                        }
```

```
                        xmin = j/npop - 1/npop
                        xmax = j/npop
                        ymin = 1-(i/npop)
                        ymax = 1-(i/npop)+1/npop
            if (v == 0){ col = "white"}
                        rect(xmin, ymin, xmax, ymax, col =
 col, border = col)
                }
                tcol = "black"
                tmp = o[o[,1] == names(d)[i],]
                if (length(tmp) != 1){
                        tcol = tmp[1,2]
                }
                mtext(names(d)[i], side = 2, at = 1-i/npop
+0.5/npop, las = 1, cex = cex, col = tcol)
                mtext(names(d)[i], side = 1, at =  i/npop-
0.5/npop, las = 3, cex = cex, col = tcol)
        }
        if ( !is.na(mse)){
                ymi = 0.5
                yma = 0.9
                w = (yma-ymi)/ncol
                xma = 0.80
                lmi = round(min, digits = 1)
                lma = round(max, digits = 1)
                print(cols)
                print(ymi+(0:ncol)*w)
                rect( rep(0.75, ncol), ymi+(0:(ncol-1))*w,
```

```
 rep(xma, ncol), ymi+(1:ncol)*w, col = cols, border = cols
)
                text(xma+0.01, ymi, lab = paste(lmi),  adj
 = 0, cex = 0.8)
                text(xma+0.01, yma, lab = paste(lma, "(Var
iance)"), adj = 0, cex = 0.8)

        }
        return(d)
        #image(as.matrix(d), col = cols)
}

plot_resid_internal = function(d, o = NA, max = 0.009, min
 = -0.009, cex =0.5, wcols = "rb", mse = NA){
        npop = nrow(d)
        width = 1/npop
        height = 1/npop
    colors = brewer.pal(9, "Spectral")
    colors = c("red", "orange","yellow", "white", "green",
 "blue", "black")
    pal = colorRampPalette(colors)
    ncol = 80
    cols = pal(ncol)
        plot("NA", xlim = c(0, 1), ylim = c(0, 1), axes =
F, xlab = "", ylab = "")
        for (i in 1:npop){
                for( j in 1:i){
                        v = d[i,j]
```

```
            print(paste(i, j, v))
                        col= "white"
                        if (v < 0){
                if (wcols == "rb"){
                col = rgb(0, 0, 1, v/min)
                }
                else{
                                #col = rgb(0, 0, 1, 0.1+0.
9*(v/min))
                col = cols[ncol/2-floor( (v/min)*(ncol/2))
]
                #col = "white"
                }
                        }
                        else{
                if (wcols == "rb"){
                col = rgb(1, 0, 0, v/max)
                }
                else{
                                #col = rgb(1, 0, 0, 0.1+0.
9*(v/max))
                col = cols[ncol/2+ceiling((v/max)*(ncol/2)
)]
                }
                        }
                        xmin = j/npop - 1/npop
                        xmax = j/npop
                        ymin = 1-(i/npop)
```

```
                        ymax = 1-(i/npop)+1/npop
                        rect(xmin, ymin, xmax, ymax, col =
 col, border = col)
                }
        tcol = "black"
        tmp = o[o[,1] == names(d)[i],]
        if (length(tmp) != 1){
            tcol = tmp[1,2]
        }
                mtext(names(d)[i], side = 2, at = 1-i/npop
+0.5/npop, las = 1, cex = cex, col = tcol)
                mtext(names(d)[i], side = 1, at =  i/npop-
0.5/npop, las = 3, cex = cex, col = tcol)
        }
    if ( !is.na(mse)){
                ymi = 0.5
                yma = 0.9
                w = (yma-ymi)/ncol
                xma = 0.80
        lmi = round(min/mse, digits = 1)
        lma = round(max/mse, digits = 1)
        print(cols)
        print(ymi+(0:ncol)*w)
                rect( rep(0.75, ncol), ymi+(0:(ncol-1))*w,
 rep(xma, ncol), ymi+(1:ncol)*w, col = cols, border = cols
)
                text(xma+0.01, ymi, lab = paste(lmi, "SE")
,  adj = 0, cex = 0.8)
```

```
                text(xma+0.01, yma, lab = paste(lma, "SE")
, adj = 0, cex = 0.8)

    }
    return(d)
    #image(as.matrix(d), col = cols)
}
```

Heteronotia binoei (Seqence Capture Code)

Demultiplex, preclean fastq's, clean read, scrub contamination, generate dynamic clean data scripts, execute

```
 wget popgen.dk/software/NGSadmix/data/pop.info
 cd '/home/whiptail/Desktop/SeqCapTest/HET_test_2.15.2016'

 perl 1-PreCleanup /home/whiptail/Desktop/SeqCapTest/HET_t
est_2.15.2016/raw_data/raw fastqc
 cd..
 for i in raw
 clear
 perl 1-PreCleanup
```

```
 history > script1.txt
 perl 2-ScrubReads -f '/home/whiptail/Desktop/SeqCapTest/H
ET_test_2.15.2016/raw_data/raw/pre-clean'
    -o '/home/whiptail/Desktop/SeqCapTest/HET_test_2.15.20
16/cleaned_data'
    -a '/home/whiptail/Desktop/SeqCapTest/HET_test_2.15.20
16/Adapters.fasta'
    -b '/home/whiptail/Desktop/SeqCapTest/HET_test_2.15.20
16/libInfo.txt'
    -t '/home/whiptail/Desktop/Phylogenetic_Tools/trinityr
naseq-2.1.1/trinity-plugins/Trimmomatic-0.32/trimmomatic-0
.32.jar'
    -c '/home/whiptail/Desktop/SeqCapTest/HET_test_2.15.20
16/associated_files/ecoli/e_coli_K12.fasta'
    -e 200 -z -i D9S4KXP1
 cd raw_data
 cd raw
 ls
 sh RAW_indexChange_HET_Cap.sh
 cd ..
 perl 1-PreCleanup '/home/whiptail/Desktop/SeqCapTest/HET_
test_2.15.2016/raw_data/raw'
 perl 1-PreCleanup '/home/whiptail/Desktop/SeqCapTest/HET_
test_2.15.2016/raw_data/raw' fastqc
```

Create dynamic script to format

names for Phyluce

```
ls -1 Script2_data/* > files.txt
while read p;
    do acc=$(echo $p | cut -d "/" -f 2 | cut -d "_" -f 2);
    outdir=$(echo $(ls -1d $acc*/ | cut -d "/" -f 1));
    read=$(echo $p | cut -d "/" -f 2 | cut -d "_" -f 4);
    echo "mv $p $outdir/split-adapter-quality-trimmed/"$ou
tdir"-READ"$read".fastq";
    done < files.txt | sed 's/READu/READ-singleton/g' > sc
ript.sh
```

Run formating script

```
sh script.sh
```

Generate combine dataset between hyrbid zone Queensland (JET) and Kimberly (RA)

Move and merge both folders named trinity-assemblies together

Create a combined taxon-set.conf file

Copy over the targeted sequences targetsFilt.fasta

Run the phyluce_assembly_match_contigs_to_probes program

```
phyluce_assembly_match_contigs_to_probes
    --contigs trinity-assemblies/contigs
    --probes targetsFilt.fasta
    --output combine-target-exon-results
```

Create an output directory for this taxon set - this just keeps things neat

```
mkdir -p taxon-sets/all
```

Create the data matrix configuration file

```
phyluce_assembly_get_match_counts
    --locus-db combine-target-exon-results/probe.matches.s
qlite
    --taxon-list-config taxon-set.conf --taxon-group 'all'
    --incomplete-matrix
    --output taxon-sets/all/all-taxa-incomplete.conf
```

Change directory to taxon-sets/all

```
cd taxon-sets/all
```

Make a log directory to hold our log files - this keeps things neat

```
mkdir log
```

get FASTA data for taxa in our taxon set

```
phyluce_assembly_get_fastas_from_match_counts
    --contigs ../../trinity-assemblies/contigs
```

```
    --locus-db ../../combine-target-exon-results/probe.mat
ches.sqlite
    --match-count-output all-taxa-incomplete.conf
    --output all-taxa-incomplete.fasta
    --incomplete-matrix all-taxa-incomplete.incomplete
    --log-path log
```

Explode the monolithic FASTA by taxon (you can also do by locus)

```
phyluce_assembly_explode_get_fastas_file
    --input all-taxa-incomplete.fasta
    --output-dir exploded-fastas
    --by-taxon
```

align the data - turn off trimming and output FASTA

```
phyluce_align_seqcap_align
    --fasta all-taxa-incomplete.fasta
    --output mafft-nexus-internal-trimmed
    --taxa 104 --aligner mafft --cores 6
    --incomplete-matrix
    --output-format fasta
    --no-trim
    --log-path log
```

run gblocks trimming on the alignments

```
phyluce_align_get_gblocks_trimmed_alignments_from_untrimme
d
    --alignments mafft-nexus-internal-trimmed
    --output mafft-nexus-internal-trimmed-gblocks
    --cores 6
    --log log

phyluce_align_get_align_summary_data
    --alignments mafft-nexus-internal-trimmed-gblocks
    --cores 6
    --log-path log
```

align the data - turn off trimming and output FASTA

```
phyluce_align_remove_locus_name_from_nexus_lines
    --alignments mafft-nexus-internal-trimmed-gblocks
    --output mafft-nexus-internal-trimmed-gblocks-clean
    --cores 6
    --log-path log
```

the integer following --taxa is the number of TOTAL taxa

and I use “75p” to denote the 75% complete matrix

```
phyluce_align_get_only_loci_with_min_taxa
    --alignments mafft-nexus-internal-trimmed-gblocks-clea
n
    --taxa 104
    --percent 0.75
    --output mafft-nexus-internal-trimmed-gblocks-clean-75
p
    --cores 6
    --log-path log
```

Generate SNPs

Getting snps from Raw-data.

Create a new folder named JET_SNP.

Then make another directory called ra-data

```
mkdir raw-data
```

Move all R1 and R2 Illumina reads to raw-data

Next make the reference database. Here I use cDNA from Anolis, which is the database that derived the probes.

```
mkdir Anolis_DB
```

Download the Anolis_carolinensis.AnoCar2.0.dna.topleve from ftp://ftp.ensembl.org/pub/release-84/fasta/anolis_carolinensis/dna/ in the Anolis_DB directory

Create the index for the cDNA reference to make a reference database

```
bwa index -p Anolis_carolinensis.AnoCar2.0.dna.toplevel An
olis_DB/Anolis_carolinensis.AnoCar2.0.dna.toplevel.fa
```

Call bwa, -p = the prefex that will be used for the database. Needs to be the as name of your fasta file, next is the location of the cDNA Anolis fasta

Sorting bams

for i in BAMS/*; do samtools sort -m 5000000000 -o ${i/bwa/bwa.sort} -O bam -T tmp/$i -@ 4 $i; done

Mark Duplicates

for i in BAMS/*sort.bam; do java -Xmx64g -jar /home/whiptail/anaconda/pkgs/picard-1.106-0/jar/MarkDuplicates.jar INPUT=$i OUTPUT=${i/bwa.sort.bam/}dedup.bam METRICS_FILE=${i/bwa.sort.bam/}rmdup.metrics.txt

ASSUME_SORTED=TRUE TMP_DIR=/tmp
VALIDATION_STRINGENCY=SILENT; done

Make an Index with samtools

samtools faidx targetsFilt.fasta

Map Read with Bowtie2

parallel bowtie2 -x bowtie2_ref/targetsFilt -U {} -S {.}.sam ::: fastq-clean/*singleton.fastq.gz #for my clean fastq.gz files # Map singleton reads
for i in `ls fastq_clean/*.fastq.gz | cut -d "-" -f 1` ; do bowtie2 -x bowtie2_ref/targetsFilt -1 $i-READ1.fastq.gz -2 $i-READ2.fastq.gz -S $i.sam ; done # Map paired Read1 and READ2 fastqs

parallel bowtie2 -f -x /media/whiptail/Drive2/JET/JET_SNP/bowtie2_ref/targetsFilt -U {} -S {.}.sam ::: /home/whiptail/Desktop/JET_Cap/Phyluce_JET_RA_Combine_Het/taxon-sets/all/exploded-fastas/*.fasta # for final fastas, notice the -f flag
for i in *.sam; do mv $i ${i/.unaligned/}; done # Get rid of .unaligned in the sample name

For my cleaned fasta reads from here on

for i in *.sam; do samtools view -b -S -o ${i/.sam/}.bam $i; done # Convert to BAM format
for i in *bam; do samtools sort -m 5000000000 -o ${i/.bam/.sort.bam} -O bam -T tmp/$i -@ 4 $i; done # Sort BAMs

Make Varient bcf file

for i in *sort.bam; do samtools mpileup -u -f /media/whiptail/Drive2/JET/JET_SNP/bowtie2_ref/targetsFilt.fasta $i > ${i/.sort.bam/}.bcf; done

Make Multi sample bcf file

```
samtools mpileup -u -g -f /media/whiptail/Drive2/JET/JET_S
NP/bowtie2_ref/targetsFilt.fasta <all sorted bams> > JET_R
A_SNPs.bcf
```

My samples

CCM0117-EIU.sort.bam CCM0228-NWQ.sort.bam CCM0257-NWQ.sort.bam CCM0283-NWQ.sort.bam CCM0544-VRD.sort.bam CCM0546-VRD.sort.bam CCM0550-VRD.sort.bam CCM0573-VRD.sort.bam CCM0575-VRD.sort.bam CCM0580-VRD.sort.bam CCM0586-VRD.sort.bam CCM0604-VRD.sort.bam CCM1715-VRD.sort.bam CCM5130-EIU.sort.bam CCM5165-EIU.sort.bam CCM5171-EIU.sort.bam CCM5172-EIU.sort.bam CCM5183-EIU.sort.bam CCM5206-EIU.sort.bam CCM5374-EA6.sort.bam

CCM5381-EA6.sort.bam CMNT059-VRD.sort.bam CMNT060-VRD.sort.bam CMNT061-VRD.sort.bam CMNT078-VRD.sort.bam CMNT079-VRD.sort.bam CMNT089-VRD.sort.bam CMNT090-VRD.sort.bam CMNT096-VRD.sort.bam CMNT097-VRD.sort.bam CCM0147-CQLD.sort.bam CCM0185-NWQ1.sort.bam CCM0245-NWQ1.sort.bam CCM0251-NWQ1.sort.bam CCM0261-NWQ2.sort.bam CCM0274-NWQ2.sort.bam CCM0278-NWQ2.sort.bam CCM0304-NWQ2.sort.bam CCM0310-NWQ2.sort.bam CCM0319-NWQ2.sort.bam ABTC28047-VRD.sort.bam ABTC28048-VRD.sort.bam CCM0370-SM6Nd.sort.bam CCM0603-SM6NW.sort.bam CCM1714-SM6NW.sort.bam CCM1837-SM6NW.sort.bam MKF00555-NWQ1.sort.bam MKF00558-NWQ1.sort.bam MKF00579-NWQ1.sort.bam MKF00580-NWQ1.sort.bam MKF00581-NWQ1.sort.bam MKF00582-NWQ1.sort.bam MKF00584-NWQ1.sort.bam MKF00585-NWQ1.sort.bam MKF00611-NWQ2.sort.bam MKF00612-NWQ2.sort.bam MKF00619-NWQ2.sort.bam MKF00620-NWQ2.sort.bam MKF00622-NWQ2.sort.bam MKF00623-NWQ2.sort.bam MKF00624-NWQ2.sort.bam MKF00625-NWQ2.sort.bam MKF00626-NWQ2.sort.bam ABTC32884-NWQ1.sort.bam ABTC32942-NWQ2.sort.bam ABTC32946-NWQ1.sort.bam ABTC32947-NWQ2.sort.bam ABTC32989-NWQ2.sort.bam CCM0119-CYA6-S.sort.bam CCM0121-SM6-NE.sort.bam CCM3332-planN3.sort.bam CCM5099-GULF-E.sort.bam CCM5100-GULF-E.sort.bam CCM5107-GULF-W.sort.bam CCM5109-GULF-W.sort.bam CCM5122-GULF-E.sort.bam CCM5137-GULF-W.sort.bam CCM5138-GULF-W.sort.bam CCM5140-GULF-W.sort.bam CCM5155-CYA6-S.sort.bam CCM5158-CYA6-S.sort.bam CCM5177-GULF-E.sort.bam CCM5188-CYA6-

S.sort.bam CCM5194-CYA6-S.sort.bam CCM5201-CYA6-S.sort.bam CCM5230-CYA6-S.sort.bam CCM5232-CYA6-N.sort.bam CCM5237-CYA6-N.sort.bam CCM5249-CYA6-N.sort.bam CCM5254-CYA6-S.sort.bam CCM5255-CYA6-S.sort.bam CCM5264-CYA6-N.sort.bam CCM5321-CYA6-N.sort.bam CCM5332-CYA6-S.sort.bam CCM5340-CYA6-N.sort.bam CCM5347-CYA6-N.sort.bam CCM5353-CYA6-N.sort.bam NTM37234-planO.sort.bam ABTC11662-CA6WA.sort.bam CCM5361-Blencoe.sort.bam CCM5401-Biloela.sort.bam WAM132771-SM6NW.sort.bam ABTC72520-NA6-Ea.sort.bam CCM3226-outgroup.sort.bam

Per Hyrbid zone RA (Kimberly) and JET (Tropical Queensland)

JET Samples

```
ABTC11662-CA6WA.sort.bam ABTC72520-NA6-Ea.sort.bam CCM0117
-EIU.sort.bam CCM0119-CYA6-S.sort.bam CCM0121-SM6-NE.sort.
bam CCM0228-NWQ.sort.bam CCM0257-NWQ.sort.bam CCM0370-SM6N
d.sort.bam CCM0603-SM6NW.sort.bam CCM1714-SM6NW.sort.bam C
CM1837-SM6NW.sort.bam CCM3332-planN3.sort.bam CCM5099-GULF
-E.sort.bam CCM5100-GULF-E.sort.bam CCM5107-GULF-W.sort.ba
m CCM5109-GULF-W.sort.bam CCM5122-GULF-E.sort.bam CCM5130-
EIU.sort.bam CCM5137-GULF-W.sort.bam CCM5138-GULF-W.sort.b
am CCM5140-GULF-W.sort.bam CCM5155-CYA6-S.sort.bam CCM5158
-CYA6-S.sort.bam CCM5165-EIU.sort.bam CCM5171-EIU.sort.bam
 CCM5172-EIU.sort.bam CCM5177-GULF-E.sort.bam CCM5183-EIU.
sort.bam CCM5188-CYA6-S.sort.bam CCM5194-CYA6-S.sort.bam C
```

```
CM5201-CYA6-S.sort.bam CCM5206-EIU.sort.bam CCM5230-CYA6-S
.sort.bam CCM5232-CYA6-N.sort.bam CCM5237-CYA6-N.sort.bam
CCM5249-CYA6-N.sort.bam CCM5254-CYA6-S.sort.bam CCM5255-CY
A6-S.sort.bam CCM5264-CYA6-N.sort.bam CCM5321-CYA6-N.sort.
bam CCM5332-CYA6-S.sort.bam CCM5340-CYA6-N.sort.bam CCM534
7-CYA6-N.sort.bam CCM5353-CYA6-N.sort.bam CCM5361-Blencoe.
sort.bam CCM5374-EA6.sort.bam CCM5381-EA6.sort.bam CCM5401
-Biloela.sort.bam WAM132771-SM6NW.sort.bam
```

RA Samples

```
ABTC28047-VRD.sort.bam ABTC28048-VRD.sort.bam ABTC32884-NW
Q1.sort.bam ABTC32942-NWQ2.sort.bam ABTC32946-NWQ1.sort.ba
m ABTC32947-NWQ2.sort.bam ABTC32989-NWQ2.sort.bam CCM0185-
NWQ1.sort.bam CCM0245-NWQ1.sort.bam CCM0251-NWQ1.sort.bam
CCM0261-NWQ2.sort.bam CCM0274-NWQ2.sort.bam CCM0278-NWQ2.s
ort.bam CCM0283-NWQ.sort.bam CCM0304-NWQ2.sort.bam CCM0310
-NWQ2.sort.bam CCM0319-NWQ2.sort.bam CCM0544-VRD.sort.bam
CCM0546-VRD.sort.bam CCM0550-VRD.sort.bam CCM0573-VRD.sort
.bam CCM0575-VRD.sort.bam CCM0580-VRD.sort.bam CCM0586-VRD
.sort.bam CCM0604-VRD.sort.bam CCM1715-VRD.sort.bam CMNT05
9-VRD.sort.bam CMNT060-VRD.sort.bam CMNT061-VRD.sort.bam C
MNT078-VRD.sort.bam CMNT079-VRD.sort.bam CMNT089-VRD.sort.
bam CMNT090-VRD.sort.bam CMNT096-VRD.sort.bam CMNT097-VRD.
sort.bam MKF00555-NWQ1.sort.bam MKF00558-NWQ1.sort.bam MKF
00579-NWQ1.sort.bam MKF00580-NWQ1.sort.bam MKF00581-NWQ1.s
ort.bam MKF00582-NWQ1.sort.bam MKF00584-NWQ1.sort.bam MKF0
```

```
0585-NWQ1.sort.bam MKF00611-NWQ2.sort.bam MKF00612-NWQ2.so
rt.bam MKF00619-NWQ2.sort.bam MKF00620-NWQ2.sort.bam MKF00
622-NWQ2.sort.bam MKF00623-NWQ2.sort.bam MKF00624-NWQ2.sor
t.bam MKF00625-NWQ2.sort.bam MKF00626-NWQ2.sort.bam
```

Call Variants (original call algorithm)

bcftools call -O v -v -c JET_RA_SNPs.bcf -o JET_RA_SNPs_origCall.vcf

JET_Hets

bcftools call -O v -v -c JET_SNPs.bcf -o JET_SNPs_origCall.vcf

#RA_Hets

bcftools call -O v -v -c RA_SNPs.bcf -o RA_SNPs_origCall.vcf

Call Variants (new call algorithm)

bcftools call -O v -v -m JET_RA_SNPs.bcf -o JET_RA_SNPs.vcf

JET_Hets

bcftools call -O v -v -m JET_SNPs.bcf -o JET_SNPs.vcf

RA_Hets

bcftools call -O v -v -m RA_SNPs.bcf -o RA_SNPs.vcf

Convert VCF to STRUCTURE format with PGDSpider

Filter SNPs

1. A standard GATK raw variant filter

line 1 filters around indels and requires a variant quality score above 30 (if not, sets genotypes to missing data ./.)
line 2 performs some hard filters documented by GATK (if SNPs called outside GATK then this probably isn't possible)
see https://gatkforums.broadinstitute.org/gatk/discussion/2806/howto-apply-hard-filters-to-a-call-set
line 3 requires that the individual sample depth for each sample is at least 5 reads (if not, sets genotype at a sample to missing data ./.)
line 4 filters away variants that have a lot of coverage, which is indicative of reads mapping from paralogous loci
line 5 filters away variants that have very low coverage
lines 4 and 5 set non-passing variants as missing data and add an annotation to the FILTER column
commonly a quick and dirty filter for overall locus depth is variant must be > 0.5x average coverage (across sites) or < 2x average coverage (across sites)
line 6 filters based on the annotations from previous filtering steps, which add filter annotations to the FILTER column

so only variants that retain a PASS annotation (because they passed all other filters) are retained
line 7 only keep biallelic variants
re-zip your VCF with bgzip so it takes less space (BCFtools will require bgzipped files in many cases)

```
bcftools filter -g3 -G10 -s LowQual -e 'QUAL<30' -S . -m +
 <<input_file.vcf>> | \
bcftools filter -s GATKHardFilters -e 'QD<2 || FS>60.0 ||
MQ<40.0 || MQRankSum<-12.5 || ReadPosRankSum<-8' -S . -m +
 | \
bcftools filter -e 'FMT/DP<5' -S . -m + - | \
bcftools filter -s HighTotalDepth -e 'DP>500' -S . -m + -
| \
bcftools filter -s LowTotalDepth -e 'DP<100' -S . -m + - |
 \
bcftools filter -i 'FILTER="PASS"' - | \
bcftools filter -i 'N_ALT=1' - \
bgzip \
> <<output_file.vcf.gz>>
```

2. To determine average coverage across all raw variants (for coverage filtering) outputs number of records (variants), culmutive depth across all records, and everag depth

```
bcftools query -f '%DP\n' <<input_file.vcf>> | \
```

```
awk '{ sum += $1 } END { print NR, sum, sum / NR }'
```

3. You'll want to know the sample IDs in the VCF file for filtering purposes. Easy to get with:

```
bcftools query -l <<input_file.vcf>>
```

4. A common command to extract a VCF for a subset of samples in larger VCF

line 1 extract the samples. Sample IDs can be passed as comma separated list (like they are here)
or as a file using the -S flag instead. Prefixing the list or file with a "^" will do an inverse match
line 2 fiters based on missing data. The AN flag in VCF stores the number of called genotypes across
all samples at a variant. This is allele counts, so it is 2 x the number of samples. You can change this
to filter by missing data. Say you have 10 samples and only want 10% missing data: 10 samples = 20 alleles
20 alleles x 0.1 = 2 alleles. Since we are excluding with the '-e' flag, we want to exclude variants
where AN is <18.
line 3 filters by minor allele frequency. In this case 0.05. So with 20 alleles, if only 1 of them
varies then the variant is excluded. This matters more with large sample datasets, and if you have less than

10 samples then you don’t have the resolution to apply this filter at the 0.05 level

line 4 uses vcftools to keep only the first variant per 10kb, if you want to help ensure that variants

are not in LD. Not necessary if you don’t have a reference genome, but good to demo it.

```
bcftools view -s sample1,sample2 <<input_file.vcf>> | \
bcftools view -e 'AN<'N'' | \
bcftools view -c 2:minor -q 0.05 | \
vcftools --vcf - --thin 10000 --recode --recode-INFO-all -
-stdout | \
bgzip \
> <<output_file.vcf.gz>>
```

5. Ways you can automate some of the parameters in #4

A.

utomatically construct comma-delimited sample list to feed into:

```
bcftools view -s sample1,sample2 <<input_file.vcf>>
```

Requires a file called sample_key.txt that has the minimum format <sample>[tab]<population> you can add more columns if you want to do hierarchical filtering and modify the command with additional grep commands to do more complex filtering this command reads the sample

key file, filters to retain only lines of samples within the desired population, cuts out the first column containing sample numbers, and pastes those samples together into a list delimited by commas:

```
popluation=my_population; \
samples=`cat sample_key.txt | grep -e "${population}" | cu
t -f 1 | paste -s -d ","`;
```

Now you can run command as bcftools view -s ${samples} <<input_file.vcf>>

B.
Automatically determine the number of alleles we want for our missing data threshold requires the same sample_key.txt file as above and can also be made more complex this command reads the sample key file, filters to retain only lines of samples within the desired population, counts the number of total samples in a population, and then uses the proportion of missing data allowed (0.25 or 25% in this case) to calculate the number of alleles that must be present at a locus for it to get retained

```
miss=0.25; \
population=my_population; \
count=`cat sample_key.txt | grep -e "${population}" | \
wc -l | awk -v miss="${miss}" '{ printf "%1.0f\n", ((1 - m
iss) * ($1 * 2)) }'`
```

You can then specify this in the #4 command as follows:

```
bcftools view -e 'AN<'${count}''
```

6. If you were interested in filtering two populations independently and then combining them back together can do something like this. Most details are included above and demoed here. only difference is the ‘bcftools merge’ command, which merges the two population VCFs back together. Note that the samples in the two populations cannot overlap.

```
thin=10000; miss=0.25; \
pop1=my_population1; \
samples=`cat sample_key.txt | grep -e "${pop1}" | cut -f 1
 | paste -s -d ","`;
count=`cat sample_key.txt | grep -e "${pop1}" | \
wc -l | awk -v miss="${miss}" '{ printf "%1.0f\n", ((1 - m
iss) * ($1 * 2)) }'`; \
bcftools view -s ${samples} <<input_file.vcf>> | \
bcftools view -e 'AN<'${count}'' | \
bcftools view -c 2:minor -q 0.05 | \
vcftools --vcf - --thin ${thin} --recode --recode-INFO-all
 --stdout | \
```

```
bgzip > <<output_pop1.vcf.gz>>;
pop2=my_population2; \
samples=`cat sample_key.txt | grep -e "${pop2}" | cut -f 1
 | paste -s -d ","`;
count=`cat sample_key.txt | grep -e "${pop2}" | \
wc -l | awk -v miss="${miss}" '{ printf "%1.0f\n", ((1 - m
iss) * ($1 * 2)) }'`; \
bcftools view -s ${samples} <<input_file.vcf>> | \
bcftools view -e 'AN<'${count}'' | \
bcftools view -c 2:minor -q 0.05 | \
vcftools --vcf - --thin ${thin} --recode --recode-INFO-all
 --stdout | \
bgzip \
> <<output_pop2.vcf.gz>>;
bcftools merge <<output_pop1.vcf.gz>> <<output_pop2.vcf.gz
>> \
> <<combined_output.vcf.gz>>
```

7. bcftools query is really useful for getting custom formats out of VCF export a genotype matrix where the first two columns are the the reference chromosome/scaffold and position. Then there are N columns for N samples (same order as in bcftools query -l output) with genotype values reflecting the number of alternative alleles (0, 1, or 2), with

missing marked as ‘NA’.

```
bcftools query -f '%CHROM\t%POS[\t%GT]\n' | \
sed 's/\.\/\./NA/g' | sed 's/0\/0/0/g' | sed 's/1\/1/2/g'
| sed 's/0\/1/1/g' | sed 's/1\/0/1/g'
```

#Heteronotia Adegenet code

```
library(ade4)
library(adegenet)
library(ape)
library(pegas)
library(seqinr)
library(ggplot2)
library(RClone)

GULF_EIU <- read.genepop("GULF.gen")

GULF_EUI <- read.structure("GULF_EIU_SNPs.str", n.ind = 17
, n.loc = 2012, col.lab = 1, col.pop = 2, row.marknames =
1)



GULF_Hybrid <- hybridize(G_East, G_West, n = 11, pop = Hyb
rid_Gulf)
```

CCM5130 - check to confirm that mt is really EIU - in GULFE for mt

CCM5230 - check - probably CYA6-N mtDNA

CCM5332 - is CYA6-N (from repeat sequence)

CCM5155

```
All_Het_SNP_genlight <- fasta2genlight("HET_all_exon_cat.fasta", chunkSize = 3000, saveNbAlleles = TRUE, n.cores = 6)

temp <- density(position(All_Het_SNP_genlight), bw=10)
plot(temp, type="n", xlab="Position in the alignment",
     main="Location of the SNPs", xlim=c(0,1701))
polygon(c(temp$x,rev(temp$x)), c(temp$y, rep(0,length(temp$x))),
        col=transp("blue",.3))
points(position(All_Het_SNP_genlight), rep(0, nLoc(All_Het_SNP_genlight)), pch="|", col="blue")

100*mean(unlist(other(All_Het_SNP_genlight))>1)
```

About 6.354632% polymorphic

```
temp <- table(unlist(other(All_Het_SNP_genlight)))
barplot(temp, main="Distribution of the number \nof alleles per loci", xlab="Number of alleles", ylab="Number of sites", col=heat.colors(4))
```

```
glPlot(All_Het_SNP_genlight, posi="topleft")

x <- glSim(100, 1000, k=5, block.maxsize=200, ploidy=2, so
rt.pop=TRUE)
glPlot(x, col=bluepal(3))

myFreq <- glMean(All_Het_SNP_genlight)
hist(myFreq, proba=TRUE, col="gold", xlab="Allele frequenc
ies", main="Distribution of (second) allele frequencies")
hist(myFreq, proba=TRUE, col="darkseagreen3", xlab="Allele
 frequencies", main="Distribution of allele frequencies",
nclass=20)
temp <- density(myFreq)

head(glNA(All_Het_SNP_genlight),20)


pca1 <- glPca(All_Het_SNP_genlight)
scatter(pca1, posi="bottomright")
title("PCA Heteronotia")

SNP_tre <- nj(dist(as.matrix(All_Het_SNP_genlight)))
SNP_tre
plot(SNP_tre, typ="fan", cex=.25)
title("NJ tree Heteronotia")

myCol <- colorplot(pca1$scores,pca1$scores, transp=TRUE, c
ex=2)
```

```
abline(h=0,v=0, col="grey")
title("PCA Heteronotia")
add.scatter.eig(pca1$eig[1:40],2,1,2, posi="bottomright",
inset=.05, ratio=.3)

plot(SNP_tre, typ="fan", show.tip=FALSE)
tiplabels(pch=20, col=myCol, cex=4)
title("NJ tree Heteronotia")
```

IBD

```
GULF_2genpop <- genind2genpop(GULF_NA_Remove)

Dgen <- dist.genpop(GULF_2genpop,method=2)
Dgeo <- dist(GULF_NA_Remove$other$xy)
GULF_ibd <- mantel.randtest(Dgen,Dgeo)
GULF_ibd
```

ABBA-BABA Tests

```
CalcD(alignment = system.file("NWQ_SM6_GULF_EIU_CalcC.fast
a", package = "evobiR"), sig.test = "N", align.format='fas
ta')

CalcD(alignment = "NWQ_SM6_GULF_EIU_CalcC.fasta", sig.test
="J", ambig="D", block.size = 1000, replicate=100, align.f
ormat='fasta')
```

```
CalcD(alignment = "CYA6N_CYA6S_EA6_GULF_CalcD.fasta", sig.
test="J", ambig="D", block.size = 1000, replicate=100, ali
gn.format='fasta')
```