Title
A tale of three hotspots - the interplay of ecological and evolutionary processes in shaping arthropod communities of Sulawesi, Hawaii, and California

Author
Holmquist, Anna Jeanette

Publication Date
2023

Peer reviewed|Thesis/dissertation

A tale of three hotspots - the interplay of ecological and evolutionary processes in shaping arthropod communities of Sulawesi, Hawaii, and California

By

Anna Jeanette Holmquist

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy, and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Dr. Rosemary G. Gillespie, Chair
Dr. Jimmy A. McGuire
Dr. Damian O. Elias

Spring 2023

A tale of three hotspots - the interplay of ecological and evolutionary processes in shaping arthropod communities of Sulawesi, Hawaii, and California

ABSTRACT

A tale of three hotspots - the interplay of ecological and evolutionary processes in shaping arthropod communities of Sulawesi, Hawaii, and California

By

Anna Jeanette Holmquist

Doctor of Philosophy in Environmental Science, Policy and Management

University of California, Berkeley

Professor Rosemary G. Gillespie, Chair

Our planet is undergoing change at a rapid pace, with novel stressors including climate change, deforestation, and species invasion drastically altering ecosystems. Preservation of diverse biotic communities is vital toward the continued production of essential ecosystem services. However, developing effective management strategies is hindered by our limited knowledge of global biodiversity, the ways it was formed, and the ways it will respond to novel disturbance. My dissertation research begins to address these critical knowledge gaps using arthropod communities in three locations within biodiversity hotspots: Sulawesi, Hawaii, and California.

The first part of my dissertation focused on mountains across the island of Sulawesi, which is situated within the biodiversity hotspot of Wallacea. This region is known for its highly endemic fauna with affinities to both Asian and Australian taxa as well as its complex patterns of diversity structured into areas of endemism (AOEs). Despite its status as a biodiversity hotspot, there is a paucity of information for many taxonomic groups and particularly for arthropods, including order Araneae (the spiders). In my first chapter, titled "Using a COI mini-barcode to document novel spider diversity in the biodiversity hotspot of Wallacea", I employed a new methodology to expedite sample processing and increase species documentation in spiders. Through sequencing 2,263 spiders, I discovered a remarkably diverse spider fauna, identifying 514 operational taxonomic units (OTUs). Additionally, I detected high turnover across elevational gradients and between mountains as well as evidence of lineages related to both Asian and Australian taxa.

The results of my first chapter guided my second chapter, titled "Patterns of colonization and *in-situ* diversification of the tetragnathid spiders in Sulawesi, Indonesia". In this chapter, I explored diversification dynamics in this highly dispersive and diverse family. Using a multi-locus data set and the software BEAST, I estimated divergence dates to investigate the evolutionary history of tetragnathids in Sulawesi. Consistent with the results from the mini-barcode data, the family Tetragnathidae was very speciose, with 42 putative species belonging to five genera (*Tetragnatha, Dolichognatha, Leucauge, Tylorida* and *Mesida*). The genera *Mesida* and *Tylorida* displayed the highest diversity, with 20 and 11 putative species respectively. Each genus had

lowland clades across each mountain with deep divergence dates corresponding to the formation of Sulawesi. The mid-elevation zones on each mountain were predominantly filled by more recent colonizers rather than descendants from lowland taxa. *Mesida* was the sole genus found at the highest elevations. Despite their dispersive capabilities, the high elevations were not colonized by tetragnathids pre-adapted to high elevation but rather by diverged mid-elevation *Mesida* taxa.

In my third chapter, "The relative importance of stochastic and deterministic processes in the assemblage of leaf litter arthropod communities across an elevation gradient", I aimed to expand my research to a community-level perspective of biodiversity formation. I used metabarcoding techniques to sequence non-spider arthropod leaf litter communities across the elevation gradient of one mountain (Gunung Dako), focusing on the interactions between stochastic and deterministic processes in community assembly. By employing tests of phylogenetic dispersion and a null model analysis, I found evidence for heterogenous selection being important in shaping both the low- and high-elevation communities. In contrast, mid-elevation communities were exclusively shaped by stochastic processes, possibly reflecting communities in an equilibrium-like state. Two of the higher elevation sites showed phylogenetic over-dispersion, often associated with competitive exclusion dynamics. Contrasting this, the most disturbed lowland site showed a significant pattern of phylogenetic under-dispersion that may indicate a strong environmental filter. This observation suggests that habitat conversion could exert a cost on lowland native taxa, instead favoring closely related taxa that have ecological strategies suitable for disturbed habitats.

Anthropogenic stressors pose a significant threat on arthropod communities, and I delved into two additional critical aspects of this issue. In my fourth chapter, titled "Invasion by an ecosystem engineer changes biotic interactions between native and non-native taxa", I examined the effects of a plant invasion on biotic interactions. I did this by sequencing the gut content of an endemic spider, *Pagiopalus* spp., in both native forest and invaded habitat, analyzing changes in dietary composition between spiders. The results revealed that spiders from the invaded habitat were eating a less consistent diet with a higher proportion of non-native taxa. Furthermore, I observed a significant increase in parasite load, particularly entomopathogenic fungi, in the spiders from invaded habitat. These findings underscore the profound shifts in biotic interactions that occur following invasion, potentially imposing fitness costs on native spiders and other species.

Another significant consequence of global change is the alteration of disturbance regimes, exemplified by California wildfires. While fires are natural disturbances in many ecosystems in California, factors such as climate change, fire suppression, and human ignition events are changing where and how fires burn. In my fifth chapter titled "The importance of habitat type and historical fire regimes in arthropod community response following large-scale wildfires", I investigated how arthropod communities in different habitat types responded following the California lightening complex fires of 2020. The study revealed that habitat type is related to how arthropod communities respond after a fire event. Habitat types that historically experienced regular, cyclical fires exhibited a recovery trajectory that led to a similar community composition

as unburned sites by the following spring season. However, scrub habitats, which historically experienced wildfires every 100-500 years, displayed a deviation from the community composition of unburned habitats, indicating a negative effect of wildfires on less fire-adapted habitats.

My dissertation research highlights the extent of biodiversity left to be discovered in hyper-diverse locations, the complex processes involved in biodiversity formation and community assembly, and the consequences of various anthropogenic stressors on arthropod communities across different systems. This research underscores the importance of biodiversity documentation efforts in threatened locations and the need to conduct biodiversity research in an ecological and evolutionary context. Such endeavors will enable us to develop effective strategies to safeguard biodiversity in an era of global change. It is imperative that we act swiftly and decisively to protect our planet's remarkable biodiversity for the benefit of present and future generations.

For Evelyn and all the others
who leave us too soon.
The world is both so cruel and so beautiful.
I hold on to your memory
And chase the dreams
You never could.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

The Ph.D. process is a grueling one, defined by long nights, exasperation, and plenty of tears. Yet simultaneously, it is one of the most rewarding journeys. While the research itself is exciting, the journey is made most rewarding by the people you meet along the way. I have felt incredibly blessed by the mentors, colleagues and friends who have come into my life over my academic career thus far. My accomplishments would not have been possible without the support and guidance of so many voices.

I would like to begin by thanking those people who pushed me at the early stages of my career. Kelly Oten, a forest entomology professor at North Carolina State University (NCSU), mentored me through an Honors Research Project and was the first person to encourage me to pursue a research career. I then joined the Steven Frank lab of urban entomology as a research assistant. I worked with multiple graduate students on their projects, ranging from studying scale insects and leaf miners, to urban pollinators and parasitoid wasps. I began working with my second influential mentor, Emily Meineke, during my time in this lab. With samples she collected through her Ph.D. projects, I began studying communities of spiders across an urban warming gradient. Her guidance and support were instrumental in my progress. Kelly, Emily and other folks from the Steven Frank lab truly made me feel that science was a place I belonged, and a field in which I could succeed.

When I began at UC Berkeley, I was welcomed into the EvoLab with open arms. There were many Ph.D. students and early career scientists who helped me find my footing at the beginning of the program, including Natalie Graham, Ashley Adams, Susan Kennedy, Jun Ying Lim, and Henrik Krehenwinkel. During my time at UC Berkeley, we have gained many more students in our lab – Leke Hutchins, Emma Steigerwald, Freddy Gutierrez, Kathy Nagel and Ellen Hollstien. In addition, I bgean collaborating with Robert Markelz and Cierra Martinez at the Berkeley institute of Data Science (BIDS) on a wildfire project, expanding my network of fantastic colleagues and collaborators. I learned so much from every one of these people, and cannot thank them enough for their knowledge, kind words, and friendship over the years.

I was lucky enough to find another family of graduate students who became close friends and colleagues through my field seasons in Sulawesi, in particular Jeff Frederick, Ben Karin, Cynthia Wang and Rachael Joakim. There is nothing quite like midnight camp floods, leaches, ant invasions, and weeks without a proper bath to bond a group of people together! This graduate community went beyond folks from the USA, and I formed close friendships with many students from the MZB (Museum Zoologicum Bogoriense) and other institutions throughout Indonesia. The companionship of other early career scientists both in and out of the field has been incredibly important to me and has kept me motivated in periods of doubt, frustration and exhaustion.

Our team in Sulawesi consisted of many scientists from MZB who were key collaborators. They helped us overcome bureaucratic hurdles as well as organized the logistics of each field season. Beyond this, they were incredible scientists who taught me so much about Indonesia and the native biodiversity of Sulawesi. I am particularly thankful for Pak Anang Setiawan, Pak Sarino,

Pak Fahri Badjeber and Bu Pungki Lupiyaningdyah. Each was integral to the success of our field seasons and the success of my own research. Pak Fahri and Bu Pungki were additionally my main collaborators on a Fulbright project that would have taken me to Sulawesi for a year. Unfortunately, COVID-19 caused this project to prematurely end. I am still extremely disappointed that our research together was not possible, but I am excited for the continuation of our collaborations in the future.

Our field seasons would additionally not have been possible without support from multiple field guides and the astounding generosity of local people. Prior to each trip, we were welcomed into a village near each mountain where we were provided with home cooked meals and housing before we started our trips. Field guides and porters would escort us up the mountain, and help establish base camps as well as guide us through the forest and find ways to move higher up the mountain. There is absolutely no way our field seasons would have been accomplishable without these people.

Funding for our project was provided by the National Science Foundation (DEB 1457845) and the principal investigators involved in this grant were Jim McGuire, Rauri Bowie, Kevin Rowe, Susan Perkins, Rosemary Gillespie and Peter Oboyski. I spent time with Jim, Rauri, Kevin and Pete in the field and each was essential to the trips' functionality and success. In particular, I am incredibly thankful for the guidance and support I received from Jim, the primary investigator on our Sulawesi project and one of my dissertation committee members. Jim welcomed me onto the team and made my research endeavors possible. His deep understanding of the system and tenacity in the field was inspiring. I learned how to be a field biologist through watching Jim and the others during our trips. Jim's encouragement and advocacy through the years has fostered within me a greater sense of confidence and belonging in the sciences. I will always be grateful for the opportunities afforded to me by Jim, the amount I have learned from him about phylogenetics and Sulawesi, and the unwavering support he provided throughout my Ph.D. program.

I have had many other academic mentors throughout my career. Damian Elias, my additional dissertation committee member, has provided invaluable perspectives beyond the scope of biogeography and evolutionary biology. As a member of my qualifying committee, Damian additionally taught me an immense amount about fascinating aspects of spider behavior and web mechanics that have shaped many research ideas that I hope to pursue in the future. Another member of my qualifying committee, Ian Wang, expanded my understanding of landscape ecology, contributing significantly to my knowledge in that field. My final qualifying committee member, Charles Griswold, provided indispensable guidance on arachnid taxonomy. Without his expertise, my taxonomic skills would be practically non-existent. As a globally renowned arachnologist, Charles' continuous support and encouragement have been instrumental in shaping my research.

I am eternally indebted to my wonderful advisor, Rosemary Gillespie, who has played a vital role in shaping my career. Among the numerous accomplished female arachnologists and scientists I have had the pleasure of working with, Rosie has undoubtedly been the most influential. Since joining the EvoLab at UC Berkeley in 2016, she has supported me in

formulating research plans aligned with my passions. Despite my limited experience, she had faith in my abilities from the beginning and gave me the freedom to develop my own collection protocols and research plans rather than constraining me to a pre-determined path. This was a continued trend throughout my Ph.D. – Rosie was a pillar of support and encouragement while allowing me to explore projects that inspired me and show allowed me to progress at my own pace. Rosie's scientific prowess and broad expertise has propelled me beyond ecology to now work in the fields of biogeography, evolutionary biology, genomics and beyond. Rosie has had a profound influence on my development as a scientist and instilled the confidence in me to continue pursuing a career in academia. The independence I will be eternally grateful for the mentorship and guidance provided by Rosie.

Lastly, I cannot thank my friends and family enough for their love and presence through the years. My childhood spent catching spiders and snails has led me to this point, and I am so thankful for the freedom provided to me by my parents. My interests were never discouraged, which led me to fill the house with terrariums of creatures that are not usually the first pet choice of many children. Having a strong relationship with my parents and siblings has been such a blessing through this process and I am profoundly thankful for my family's support.

I have had many challenging years, but the consistent support and love I have received from such a wonderful community of people has propelled me forward. I am deeply and inexplicably grateful to be here today and that is a direct result of so many helping hands. Thank you all!

This research was supported by: the National Science Foundation (NSF) Graduate Research Fellowship Program; the NSF grant DEB 1457845; the US Fulbright Program; the Fulbright Critical Language Enhancement Award; the Environmental Sciences, Policy and Management Research Starter Grant; the Charles W. Woodworth Fellowship; the Walker Fund for Systematic Entomology; the Edward A. Steinhaus Memorial Fund; the Robert L. Usinger Memorial Fund; the Bob Land and Sandy Purcell Endowed Graduate Student Support Fund.

INTRODUCTION

Earth is currently experiencing a biodiversity crisis, considered by many to be the beginning of a sixth mass extinction event[1]. Several fundamental challenges impede our ability to assess and respond to this crisis. One major obstacle is our poor knowledge of global biodiversity; without quantification and documentation of species diversity, it is impossible to measure the extent of species loss and its effect on ecosystems. Along with species discovery, a deeper understanding of the processes involved in species formation and community assembly is instrumental to preserve global biodiversity. Furthermore, it is essential we quantify how biodiversity responds to various anthropogenic stressors in communities that are shaped by different processes in order to develop effective recovery strategies as well as identify ecosystems at the greatest risk. My dissertation expands our knowledge on these three challenges through studying arthropods in diverse systems across the globe.

PART 1:

A key goal in ecology is to understand the mechanisms involved in biodiversity formation. Biodiversity hotspots, which encompass only 1.4% of terrestrial land, are areas of concentrated terrestrial biodiversity that serve as ideal systems for studying the processes underlying biodiversity formation[2,3]. Speciation is influenced by both evolutionary and ecological processes that occur across broad spatial and temporal scales, making unraveling the formation of biodiversity a monumental feat. Islands provide simplified systems that serve as natural laboratories for investigating questions of ecology and evolution[4]. One such island is Sulawesi, Indonesia, the largest island in the biodiversity hotspot of Wallacea. Sulawesi exhibits remarkable endemic biodiversity with patterns of diversification tied to the regions' complex geological history[5,6]. Sulawesi formerly existed as a series of smaller islands then underwent extensive tectonic and volcanic activity in the last 1-4Ma[7], resulting in the connection of the paleoislands and formation of mountain ranges across the island. This unique history provides an ideal system to understand processes involved in speciation through time.

While the diversity and patterns of *in-situ* diversification for charismatic endemic vertebrate fauna of Sulawesi are relatively well-studied[5,8,9], the same cannot be said for phylum Arthropoda, which is the most abundant and speciose group of organisms. This includes order Araneae (the spiders), a key mid-trophic level group that is considered to be the most important predator of insects and other small animals[10] as well as an important source of prey for larger taxa, including birds and amphibians. While there are 238 species of spiders currently documented in Sulawesi[11], species diversity is estimated to be much higher. To establish a foundation for understanding spider biodiversity in Sulawesi, **CHAPTER 1** of my dissertation focuses on basic biodiversity exploration. I developed a protocol that increases the speed at which we can classify biodiversity by using a mini-barcode in combination with non-destructive whole body DNA extractions and metabarcoding of juvenile taxa, enabling increased throughput and species detection.

In **CHAPTER 2** of my dissertation, I aimed to understand the patterns of *in-situ* diversification and colonization in the tetragnathid spiders (Araneae: Tetrangathidae). In addition to the COI

mini-barcode data, I sequenced four additional loci as well as a longer region of COI for all adult specimens. By constructing a time-calibrated phylogeny using these data, I estimated divergence dates and investigated speciation processes across different mountains. This chapter focused on the "how" behind the formation of species-level biodiversity on Sulawesi.

Another component important in biodiversity formation is the process of community assembly, where deterministic processes such as competitive exclusion or environmental filters, combine with neutral processes such as drift to dictate a species membership in any given community and influence evolutionary trajectories[12,13]. In **CHAPTER 3** of my dissertation, I focused on the key processes involved in community assembly of arthropod communities through time and how communities today are responding to anthropogenic stressors, in particular habitat conversion. To achieve this, I used a metabarcoding approach to sequence whole arthropod leaf litter communities collected across the elevational gradient of one mountain, Gunung Dako. This elevational gradient was simultaneously a disturbance gradient, with the lowest elevation sites being agroecosystems. This chapter begins to bridge the gap between evolutionary and ecological processes acting at different temporal scales.

PART 2:

Anthropogenic stressors can have complex and diverse impacts on biotic communities shaped by different historical legacies[14,15]. The type of stressor will influence response as well, making it difficult to identify universal tipping points[15]. Therefore, it is essential we study multiple types of anthropogenic stressors in different systems to more broadly understand how major disruptions to the environment are changing the structure and distribution of biodiversity. By doing this, we can better predict how particular ecosystems may respond to disturbance.

In the second part of my dissertation, I investigated arthropod communities in Hawaii and California, each experiencing different types of anthropogenic disturbance, to gain more insight into response to different stressors. In **CHAPTER 4** of my dissertation, I assessed the effects of a recent invasion by a plant ecosystem engineer on biotic interactions between native and non-native taxa. By assessing the changes in diet and the interaction with parasites of an endemic generalist predator, *Pagiopalus* spp. (Araneae: Philodromidae) in native and invaded habitat, I explored the combined effects of ecology and evolution on the response to habitat alteration.

In addition, changes to disturbance regimes are a significant problem affecting biodiversity. Wildfires, for instance, are a natural, cyclical disturbance experienced by many ecosystems. However, wildfires are burning longer, hotter and larger than in historical fire regimes and entering new ecosystems that are not burn-adapted. In **CHAPTER 5** of my dissertation, I sequenced whole arthropod communities collected from burned and unburned sites across California to understand how biotic communities are responding to these changes. Using these data, I explored how environment and history shape community recovery by comparing burned and unburned communities across spatial scales and habitat type.

PART 1

Understanding arthropod biodiversity across three mountains in Sulawesi, Indonesia

Anna J. Holmquist, Pungki Lupiyaningdyah, Fahri Badjeber, Rosemary G. Gillespie

In the first part of my dissertation, I focus on the arthropods of Sulawesi. Using various molecular approaches, I assess spider biodiversity on three mountains, as well as study how biodiversity has formed both at a species and at a community level.

Chapter 1: Using a COI mini-barcode to document novel spider diversity in the biodiversity hotspot of Wallacea

1.1. Introduction

Global anthropogenic change has created a biodiversity crisis, characterized by extinction rates far above the background extinction rate and leading many in the scientific community to believe we have begun a sixth mass extinction event[1,16]. Close monitoring of ecosystems is of the utmost importance to detect tipping points that may lead to widespread species loss[14]. Research has predominantly focused on single species, especially keystone species, to track changes in ecosystems[17,18]. However, the keystone species does not accommodate for the complexity of true ecosystems[19]; a community-level monitoring is essential to detect broadscale shifts in species composition that may lead to ecosystem collapse. Effective conservation efforts require a holistic approach that incorporates knowledge across taxonomic groups and accounts for the ecological redundancy provided by species-rich systems [20].

A major challenge in studying biodiversity at the community scale is the high density of undescribed taxa, especially in biodiversity hotspots. Biodiversity hotspots, supporting over half of terrestrial biodiversity on just 1.4% of Earth's land area, additionally contain most of the globe's undocumented diversity, demonstrated by taxon-specific studies that can produce hundreds of new and endemic species [21–23]. Species in these hyper-diverse locations are being lost at a rate higher than species discovery, making it extremely difficult to measure the extent of global extinctions [24]. These taxa are at particular risk of extinction due to their limited ranges and dependence on rare habitats [25,26]. Focus on conservation of biodiversity hotspots has a high pay-off for species protection due to this concentration of diversity [2], but measuring the success of conservation efforts needs a community-level approach.

For rapid assessment of biodiversity on a broader scale, it is crucial we develop effective and efficient methods that allow exploring patterns at multiple taxonomic levels while maintaining specimens to enable species descriptions and further research. High-throughput methods using next generation sequencing (NGS) have emerged as a tool to rapidly assess unknown diversity at the community level [27–29]. Metabarcoding and eDNA approaches, using pooled DNA, allow us to quickly assess whole communities across taxonomic groups [30–32]. Here, we use whole community barcoding to understand general properties of novel biodiversity on Sulawesi, Indonesia in the hotspot Wallacea. Our study focuses on spiders (order Araneae), a hyperdiverse lineage consisting of over 50,000 described species. We use a COI "mini-barcode" to sequence thousands of spiders, both individually and using a pooled metabarcoding approach, to assess patterns of diversity across three mountains. A mini-barcode region was chosen to increase amplification and sequencing success across many specimens [33,34]. When paired with non-destructive whole body DNA extractions that allows both retention of voucher specimens and increases in throughput, these approaches are powerful in our efforts to document diversity [35].

Our first objective was to test the utility of the COI mini-barcode in classifying species-like operational taxonomic units (OTUs), and our second objective was to use the generated data to explore classic patterns of biodiversity on Sulawesi, including the elevational diversity gradient,

the presence of areas of endemism (AOEs) and the high prevalence of *in situ* diversification. We hypothesized that 1) species diversity will be highest at mid-elevations as expected under the elevational diversity gradient, 2) mountains will contain distinct pools of diversity, reflecting areas of endemism, and (3) patterns of divergence will be detectable based on finer or coarser clustering thresholds (99%, 97% and 95%). Here, high overlap across elevation or between mountains in 99% clusters could represent recent divergence occurring at a population-like level, divergence in 97% clusters but not 99% could represent species radiations on Sulawesi, and high overlap only at 95% could represent colonization of different areas by different superspecific lineages.

1.2. Methods

1.2.1. Sample collection

Three mountains of Sulawesi (Gunung Dako, Gunung Torompupu, and Gunung Ilomata; Figure 1) were sampled in coordination with an NSF-funded project (DEB 1457845) conducting a biotic survey of Sulawesi mountains. This project is in close collaboration with scientists at the Museum Zoologicum Bogoriense in Java, Indonesia as well as scientists from UNTAD in Palu, Sulawesi. Community sampling was conducted using standardized methods. The standardized protocol used sites of 20x20m, roughly 400m apart across an elevation gradient on three mountains, allowing thorough collection from distinct habitat types (Figure 2; Supplementary materials, Table 1). Each site was sampled using timed day hand collection (1 hour) and timed night hand collection (2 hours), leaf litter sorting (using a standardized volume of pre-sifted soil), timed beat sampling (20 seconds for distinct plant types, over 7 units), pitfall trapping (2 per site) and web documentation (1 hour). The range of elevation sampled and the number of sites at each elevation was dependent on local field conditions, accessibility, and time limitations. Specimens were stored in bulk in 95% EtOH, except for spiders collected from webs which were stored individually. Research was undertaken under permits issued by RISTEK-DIKTI, and samples were exported to the USA following protocols established in a memorandum of understanding (MoU) and material transfer agreement (MTA) with MZB and under export permits issued by LIPI. Upon importation to the USA, samples were stored in -20° until DNA extraction and then transferred to 70% EtOH and stored at room temperature.

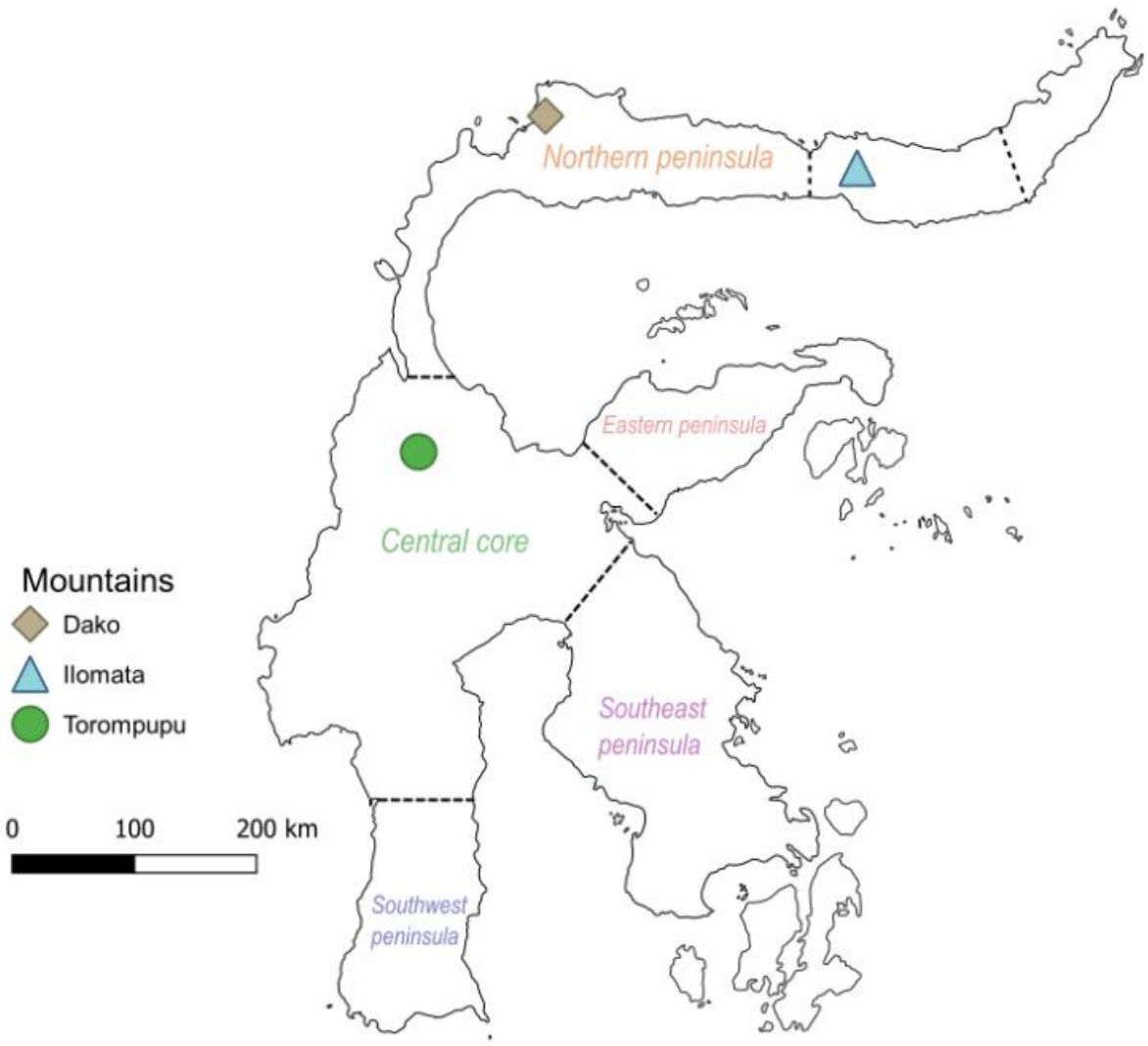


Figure 1. Map displaying mountains sampled for this project. Gunung Torompupu is located in the central core while Gunung Dako and Gunung Ilomata are located on the northern arm of Sulawesi.

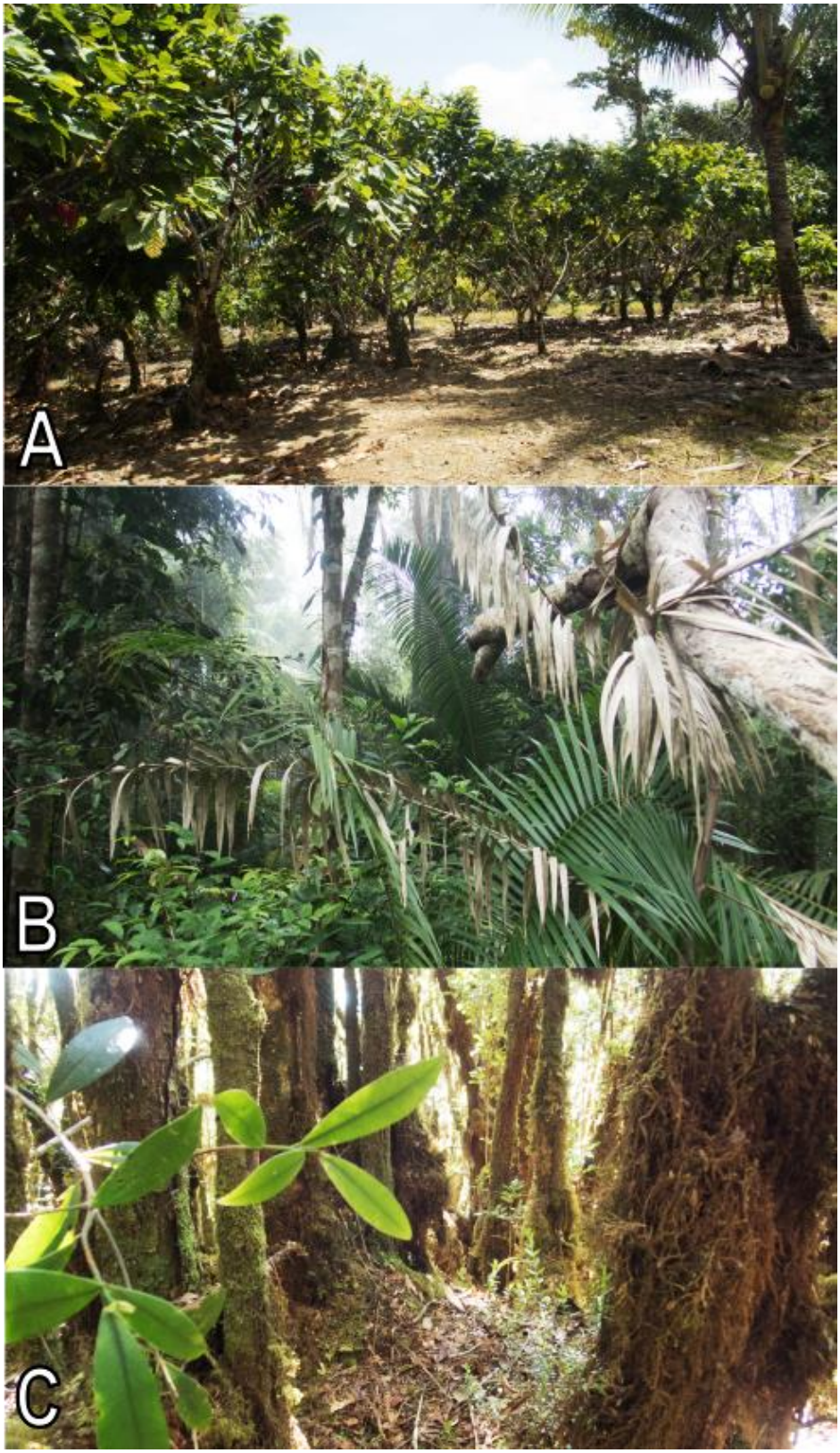


Figure 2. Examples of habitat types at different sites from Gunung Dako – lowland cacao agroecosystem (350m), mid-elevation forest (1400m), and high-elevation mossy forest (2200m).

1.2.2. Molecular Procedures

Samples were sorted by sex; juvenile specimens from each collection unit were combined into pools while adult taxa were grouped into morphospecies by site. A female and/or male of each morphospecies was sequenced for each site individually. DNA was extracted from pulverized leg tissue or non-destructively by soaking the entire specimen in cell lysis buffer prior to extraction. Volume of cell lysis buffer was varied based on tissue amount. Pooled juvenile samples were extracted non-destructively by adding 600uL of cell lysis buffer and proteinase K. Samples were then incubated for 3 hours to preserve morphological characters (Supplementary materials, Figure 1). Following lysing, additional extraction steps were performed using the Qiagen PureGene protocol (Qiagen, Hilden, Germany). To increase the throughput of the protocol, all extractions were performed in 96-well plates, aided by a repeater pipette and a bench smart 96-well pipette machine. The produced DNA pellets were eluted using Qiagen elution buffer with a volume of 20uL. To test success of extractions, DNA was spot checked across each plate via NanoDrop (ThermoFisher Scientific, Waltham, MA, USA).

Amplification was performed using the Qiagen multiplex kit (Qiagen, Hilden, Germany). The "mini-barcode" was amplified using the primer pair LCO1490 (5′-GGTCAACAAATCATAAAGATATTGG-3′)[36] and COI-CFMRb (5′-GGNACTAATCAATTHCCAAATCC-3′)[37] which additionally contained a 5` TruSeq tail. The widely used LCO1490 combined with the reverse COI-CFMRb amplifies a short fragment of COI (181bp) and was found very successful in amplifying spiders[37]. The PCR reaction consisted of 5ul of the Qiagen PCR MasterMix (MM) (Qiagen, Hilden, Germany), 3ul of H20, 0.5ul of each primer, and 1ul of template DNA, performed on a 96-well plate to increase throughput. Amplifications were conducted at an annealing temperature of 46°C for 30 cycles. A negative PCR control was included on each plate, consisting of MM, H20, and primers. PCR products were visualized using gel electrophoresis on a 3% agarose gel, ran at 140V. PCR products and associated DNA from specimens lacking bands were checked for quality then re-amplified if necessary. A dual indexing strategy was implemented using a second round of PCR to attach 8bp indexes. Annealing temperature for indexing PCR was 55°C and run for 6 cycles. PCR products were visualized once again using a 3% agarose gel, run at lower voltage (100V) to allow clear visualization of fragment length to confirm the addition of indexes to each amplicon. Final libraries were constructed by pooling PCR products proportionally based upon band strength; a separate library was constructed for individually-extracted adult specimens and bulk-extracted juvenile samples. Further processing was conducted at QB3 Genomics (QB3, Berkeley CA, USA). Library quantification was performed using both qPCR and a Qubit Fluoremeter (ThermoFisher Scientific, Waltham, MA, USA). Size selection was performed using Pippin Prep (Sage Science Inc, Beverly MA, USA) followed by Fragment Analyzer (Agilent, Santa Clara, CA USA)to confirm correct amplicon sizes. Pools were sequenced using Illumina MiSeq v3 300PE (Illumina Inc, San Diego, CA USA) on two lanes, combined with other libraries.

1.2.3. Bioinformatics

Samples were demultiplexed by QB3 staff (QB3, Berkeley CA, USA) using custom methods. CutAdapt was used for primer removal using the paired-end method[38]. Following primer removal, sequence data were denoised using DADA2 as implemented in R[39]; DADA2 relies on error models that incorporate all data produced during a sequencing run and has been shown to be accurate in producing accurate amplicon sequence variants (ASVs) which retain fine scale

nucleotide variation. Following DADA2 processing, sequences were filtered to the expected length of 181bp and chimeras were removed using the *removeBimeraDenovo* method implemented in the DADA2 package. The function *isContaminant* in package decontam was used to remove potential widespread contamination based on sequences detected in negative controls[40]. This method assesses prevalence in controls versus positive samples to identify contaminants. The threshold parameter was set to 0.5, which removes any contaminant more prevalent in negative controls than in true samples. The LULU curation algorithm was performed on the resulting sequence data to remove remaining erroneous ASVs[41].

Each of the above methods was performed by individual lane to account for run-specific errors. Following the above filtering, all data were combined. ASVs were then manually examined in Geneious Prime v. 2022.0.2 to remove any sequences with reading frame interruptions or stop codons. Individual sequence variants were clustered into operational taxonomic units (OTUs) using swarmv2[42] with the --fastidious option. Unlike other clustering methods, swarm does not rely on a universal threshold and instead uses an iterative single-linkage clustering algorithm to create units. The fastidious option reduces under-grouping by collapsing low-abundance OTUs into larger "parent" OTUs. This method was used to create species-like units while ASVs, with their fine-scale nucleotide variability, were treated as haplotypes. More aggressive clustering was performed using function *otu* in the package kmer[43] at 0.97 and 0.95 to represent coarser taxonomic units. Throughout this paper, these are referred to as 97% clusters and 95% clusters, while the swarmv2 species-like clustering results are referred to as OTUs and the products of DADA2 are referred to as ASVs.

The remaining curated data set was assigned taxonomy using megablast implemented in Geneious. Sequences belonging to Araneae were identified using the most up-to-date database on spider biodiversity, the World Spider Catalog[11], to the genus level. Non-spider sequences were removed. To assign samples to family, neighbor-joining trees were constructed in Geneious Prime using parent OTU sequences. A subset of adult samples were morphologically identified to assign taxonomy to clades. This allowed any OTUs in a well-supported clade to be assigned a confident family ID, enabling assignment of a taxonomic identification to juveniles.

Despite the multiple well-supported filtering methods we used, many samples still contained more than one OTU. The prevalence of low-abundance OTUs incorrectly appearing in a sample is a well-known problem in NGS approaches. Often, subjective read cut-offs are used to address this problem. To produce more quantitatively informed cutoffs to use as universal filtering methods, we compared morphological identifications of samples with the identity of OTUs found in a sample and assessed what were the strongest predictors of a mismatched OTU classification. The number of reads for OTUs across the entire run, the number of reads for an OTU in a sample, and the proportion of the reads found within a sample to the OTU reads found across all samples showed strong relationships to correct or incorrect identities (Supplementary Materials, Figure 2). Cutoffs were calculated using interquartile ranges, separated by values associated with incorrect and correct matches.

1.2.4. Statistical Analysis

To test if OTUs found across mountains were more likely to consist of multiple haplotypes (ASVs), we performed a student's T-test in R. We constructed haplotype networks for OTUs that consisted of multiple variants using packages ape[44] and pegas[45]. The OTU with the highest number of haplotypes and found across all three mountains is provided as an example. We

assessed alpha diversity using Hill numbers, calculated with function *renyi* in the package vegan[46]. We used three hill numbers: $q = 0$ which represents pure richness, $q = 1$ which is the exponential of Shannon's entropy index, and $q = 2$ which is the inverse of Simpson's concentration index. Differences in alpha diversity across mountains were tested for all Hill numbers using ANOVAs (Supplementary Materials, Table). To look at the relationship between alpha diversity and elevation, we constructed three linear models for each Hill number elevation as linear predictors. Venn diagrams were constructed using the package ggvenn[47].

A maximum likelihood phylogeny was constructed using IQTREE v1.6.12; the input alignment consisted of all OTU sequences as well as sequences downloaded from the BOLD database based on locality, ranging from South Asia to New Zealand. The appropriate substitution model was determined using ModelFinderPlus[48]. The general time reversible model was selected, using the FreeRate model of rate heterogeneity and empirical codon frequencies (GTR+F+R7). 440 total iterations were ran to reach parameter optimization. Phylogenetic alpha and beta diversity were calculated using the package BAT based on the produced tree[49]. We used NMDS as our ordination method to visualize group differences then tested for significant compositional differences by the categorical groupings of mountain and elevation using the ADONIS variant of PERMANOVA with *adonis2* implemented in vegan[46]. Differences in group dispersions can make PERMANOVA unreliable and so we additionally tested for dispersion differences using PERMDISP using the function *betadisper* (Supplementary Materials, Table). Well-defined clades were subset using ape[44] and visuals with color demarcation based on locality were constructed using ggtree[50] and treeio[51].

Analyses were ran in R v4.2.2 and Rstudio v2022.12.0. Figures were made using ggplot2 and ggpubr and tables were made using stargazer. Code was written using the tidyverse tidyverse. Further figure edits were made in Inkscape v1.2.

1.3. Results

1.3.1. Sequencing results

A total of 2,263 specimens were sequenced – 819 adults that were individually barcoded and 1,444 juveniles that were sequenced using a pooled approach. Following DADA2, decontamination and LULU filtering steps, 3,683,335 reads remained grouped into 986 ASVs and 676 OTUs using swarmv2. Juvenile samples produced an average 10.86 OTUs with a median of 11 OTUs per sample. Juvenile samples produced 116 unique OTUs not found in individually sequenced samples. Individual adult samples produced an average 4.07 OTUs and median 3 OTUs per sample with 382 OTUs not detected in juvenile samples. Because individual samples should produce a single OTU, additional filtering steps were implemented to select the most well-supported OTU for an individual. The most common method adopted method is selecting the OTU with the highest reads. However, this method produced 523 taxonomic matches and 201 mismatches. Visual exploration of the samples with identifications showed incorrect matches were 1) those in low read abundance, 2) those that made up a small fraction of total reads for an individual, and 3) those that were a small proportion of the total reads produced for a given OTU (Supplementary Materials, Figure 2). The distributions of each metric for correct and incorrect identifications were used to decide on an appropriate threshold. The OTU size threshold was chosen at 45 reads or greater. This step alone reduced the number of OTUs per individual to a median value of 1 and mean value of 2.14. The proportion individual reads to

total reads for an OTU was chosen as 0.025 or greater; a mean of 1.52 OTUs per individual resulted. For the samples which remained with multiple OTUs, a final filtering step was applied where only OTUs that made up over 0.35 of the total reads for an individual were kept. This resulted in a mean of 1.06 OTUs per individual, with 42 of the 820 retained samples remaining with two OTUs. These were assessed manually based on phylogenetic placement and morphological ID and then the most likely OTU chosen.

1.3.2. Biodiversity discovery

A total of 26 spider families were detected in the final data set, including three not yet known from Sulawesi (Figure 3; Supplementary Materials, Table 2). As anticipated due to the short region of COI, family monophyly did not emerge on the phylogeny; however, clear clades within families did form (Supplementary Materials, Figure 3). Following all filtering and cleaning procedures, 514 OTUs consisting of 650 ASVs were produced. These were further clustered into 333 97% clusters and 142 95% clusters (Table 1). The majority of OTUs (453 of 515) returned matches below 95% and required manual assignment to family using the phylogeny. Only six OTUs produced above a 99% BLAST threshold, suitable for species assignment. These species were *Steatoda cingulata, Leucauge decorata, Leucauge celebesiana,* and *Cyclosa bifida*. Two other OTUs returned matches above 99% but produced only a genus-level match (*Herennia sp.* and *Neoscona sp.*). The six OTUs were documented only in specimens collected from lowland sites, below 500m.

Clustering Threshold	Number of produced clustering units
Amplicon sequence variants (ASVs)	650
Swarmv2 clustering (OTUs)	514
97% kmer clustering	337
95% kmer clustering	142

Table 1. Number of clusters produced when using a) amplicon sequence variants from DADA2, b) operational taxonomic units from swarmv2, c) 97% otu clustering using kmer and d) 95% otu clustering using kmer.

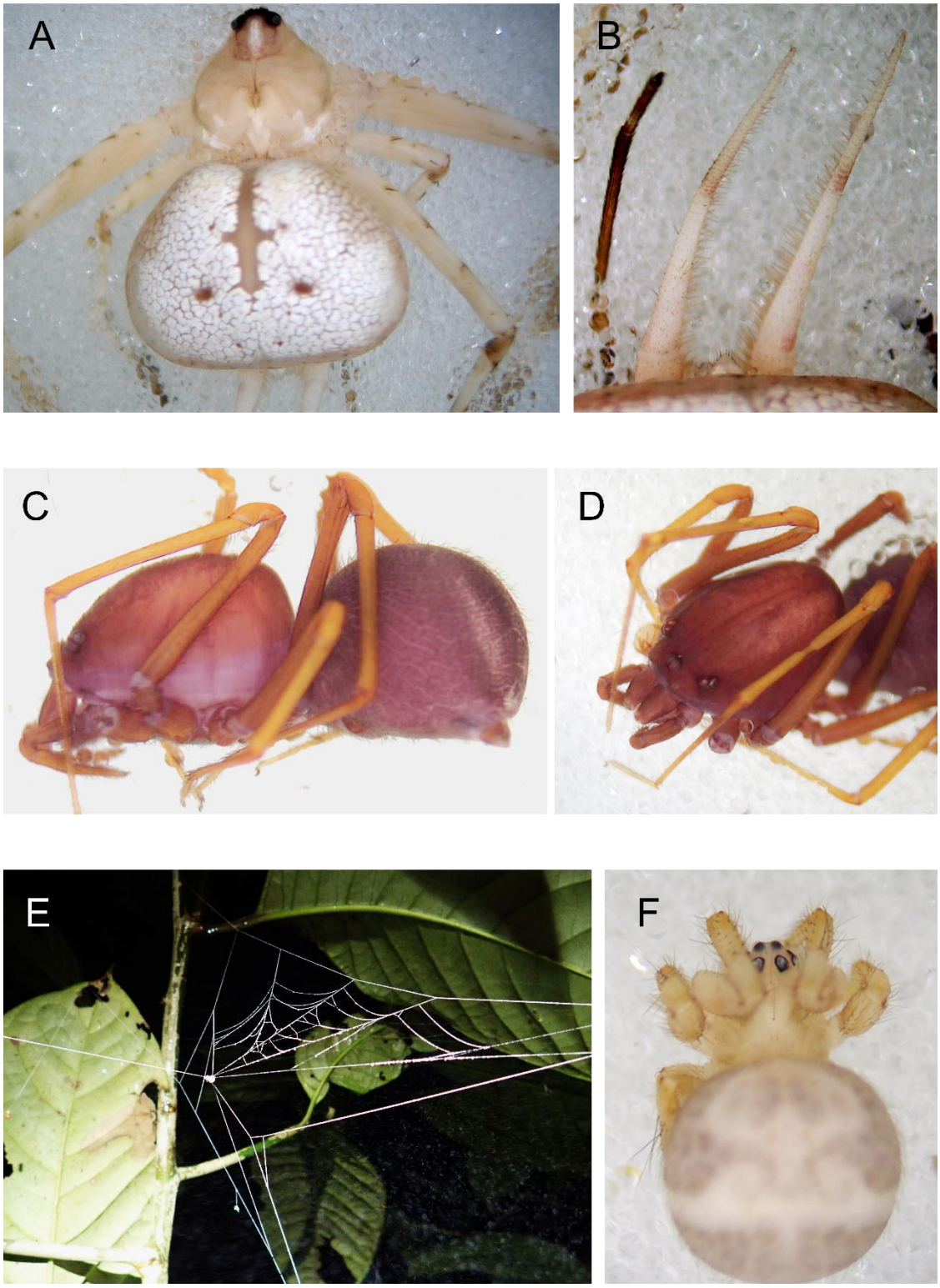


Figure 3. Three families (Hersiliidae, Scytodidae, and Theridiosomatidae) were not yet known from Sulawesi. Image A) shows a hersiliid specimen and B) the characteristic spinnerets of this family. Images C) and D) show a scytodid specimen, with the defining C) domed carapace and D) six eyes of this family. Image E) shows the distinctive cone-shaped ray web belonging to a F) theridiosomatid specimen.

98 of the 515 OTUs consisted of multiple ASVs, which can be considered haplotypes; 6 OTUs consisted of 4 or more haplotypes. OTUs found across multiple mountains were significantly more likely to consist of multiple haplotypes (t = 7.5069, p-value = 7.345e-11) with a mean value of 1.97 ASV haplotypes per OTU compared to 1.13 ASV haplotypes per OTU found on a single mountain (Figure 4). All OTUs consisting of 4 or more ASVs were detected on multiple mountains. A haplotype network was constructed for OTU-1257, which consisted of 5 ASVs found across 10 samples (Figure 5). The most common haplotype was detected only on Dako. Haplotypes found on Ilomata and Torompupu stemmed from this central haplotype and did not show connection to one another. Ilomata haplotypes were linked to a smaller haplotype found on Dako and additionally one haplotype was shared between the two mountains.

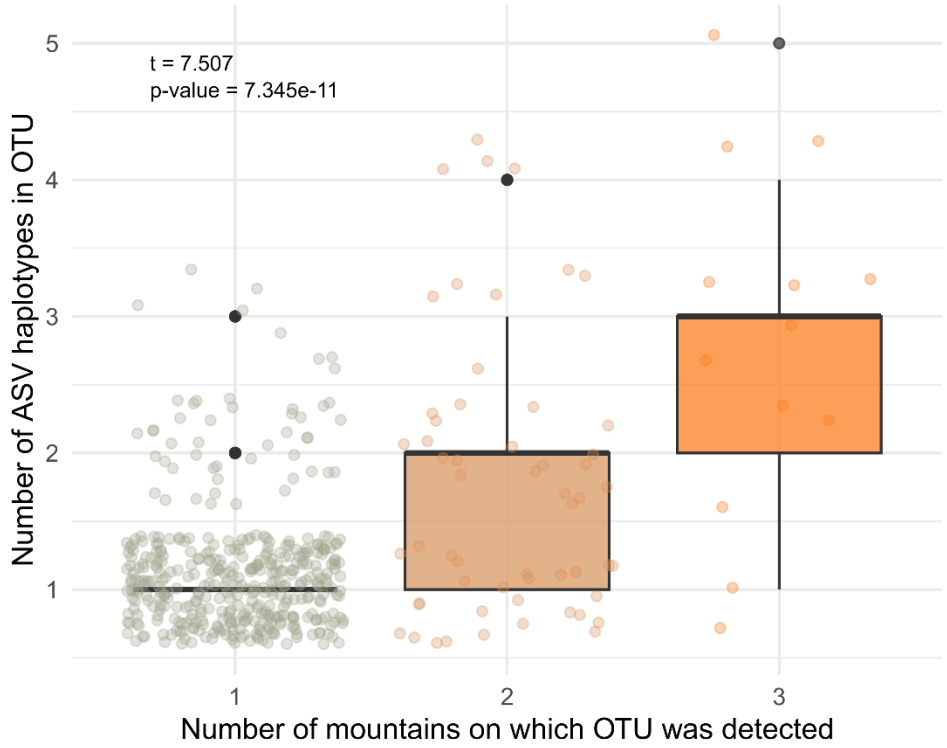


Figure 4. Number of mountains on which an OTU was found, with the y-axis representing number of haplotypes (ASVs) within each OTU. OTUs with more than one haplotype were more likely to be found on multiple mountains.

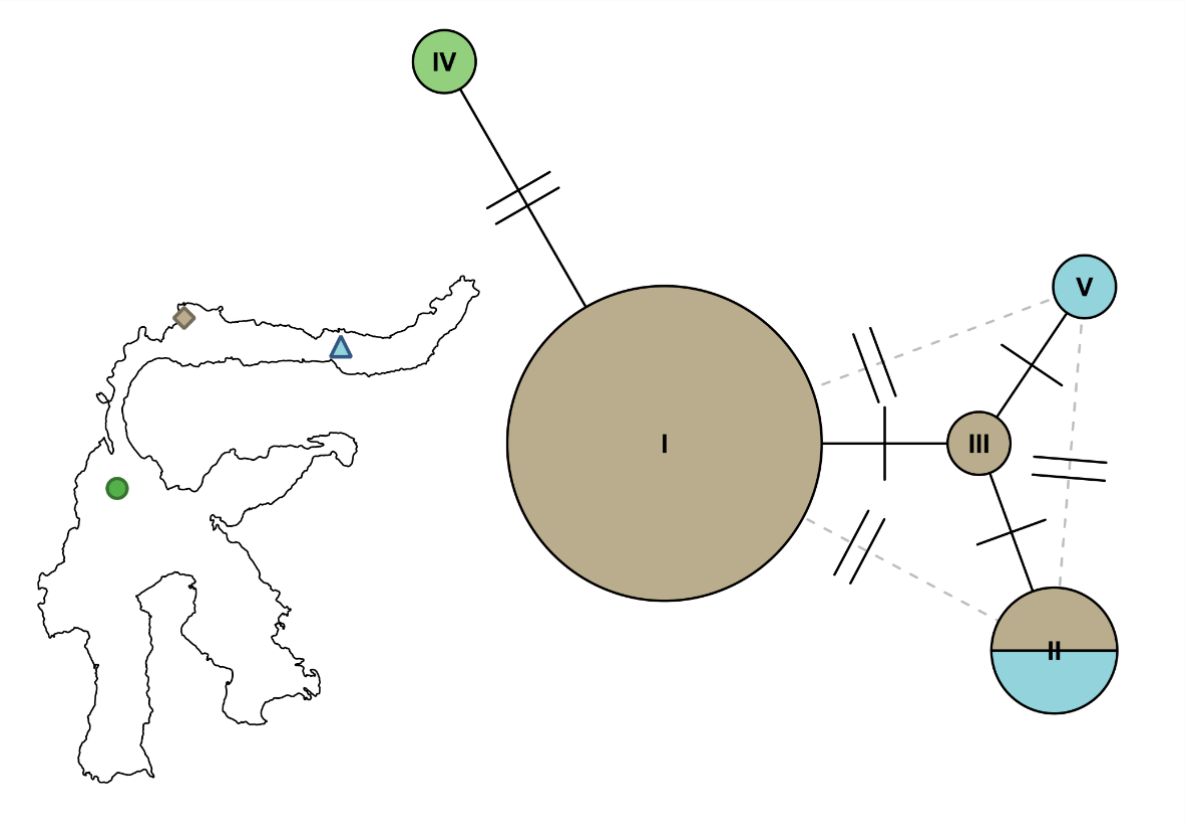


Figure 5. Haplotype network of the OTU containing the most ASVs, showing Dako in brown, Ilomata in blue, and Toromopupu in green.

1.3.3. Patterns of diversity across elevational gradients and geographic locations

Alpha diversity decreased slightly as elevation decreased; this pattern was reflected in all Hill numbers but was significant for $q = 2$ (Adj. $R^2 = 0.209$, Pr = 0.016) when abundance data was a major component (Figure 6; Supplementary Materials, Table 3). The highest diversity values were detected at mid-elevation sites (Figure 7). Alpha diversity varied significantly by mountain for q = 0 and q = 1, with Dako showing the lowest mean diversity (Figure 8; Supplementary Materials, Table 4).

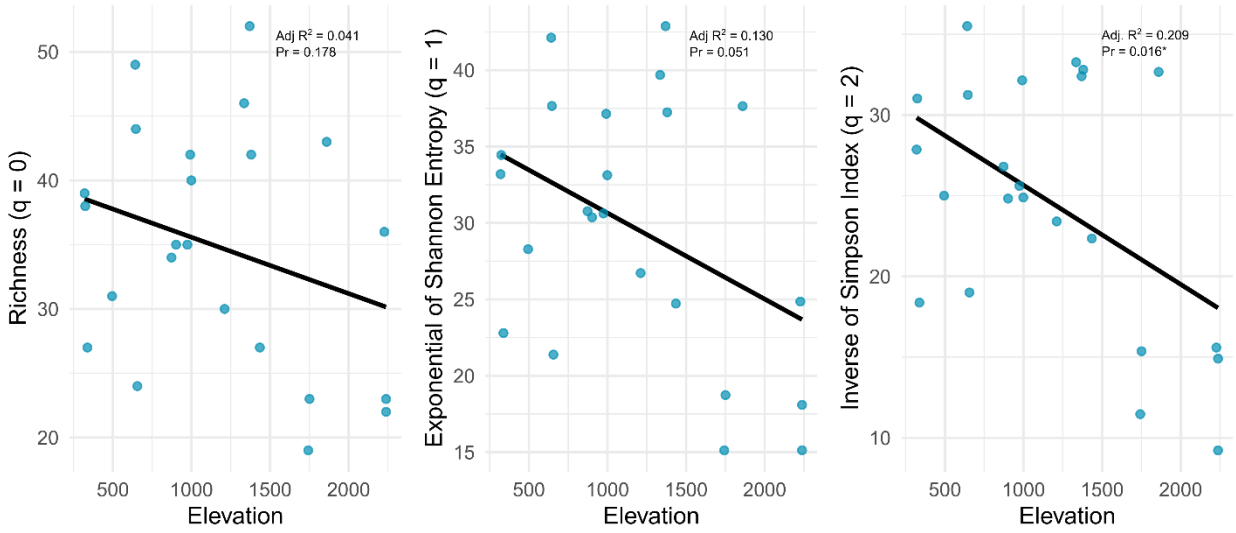


Figure 6. Fitted regression lines for Hill numbers (y) against elevation (x). All Hill values decreased slightly over elevation, and the association was significant when using q = 2.

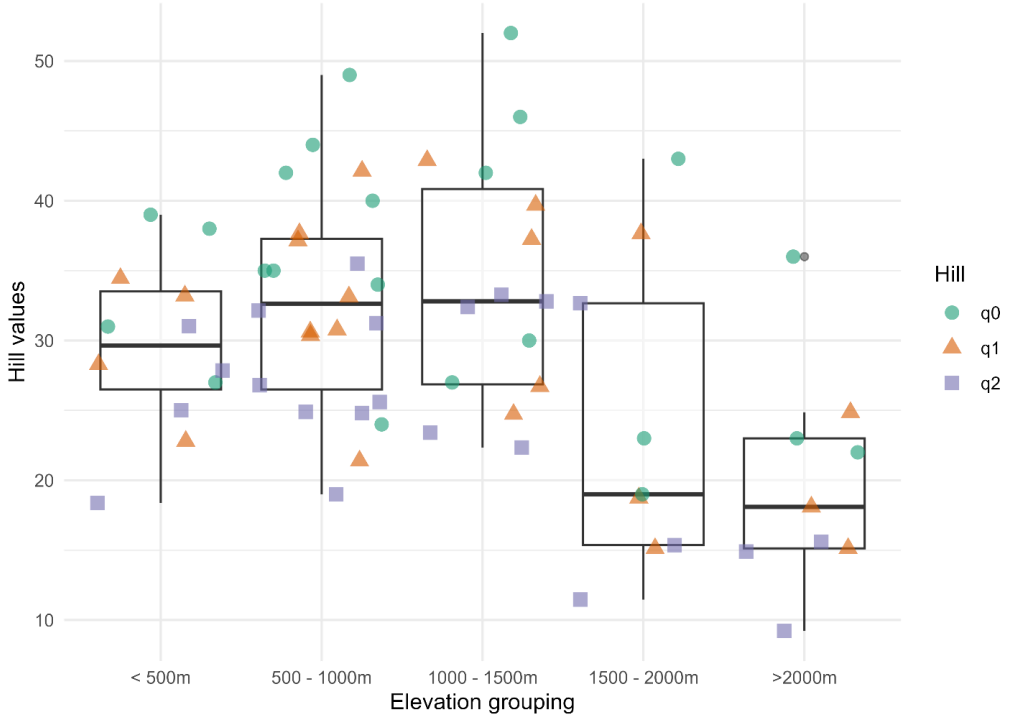


Figure 7. Hill values using q = 0, q = 1, and q = 2, grouped by elevation. The mid-elevation bands (500-1500m) had the highest diversity values.

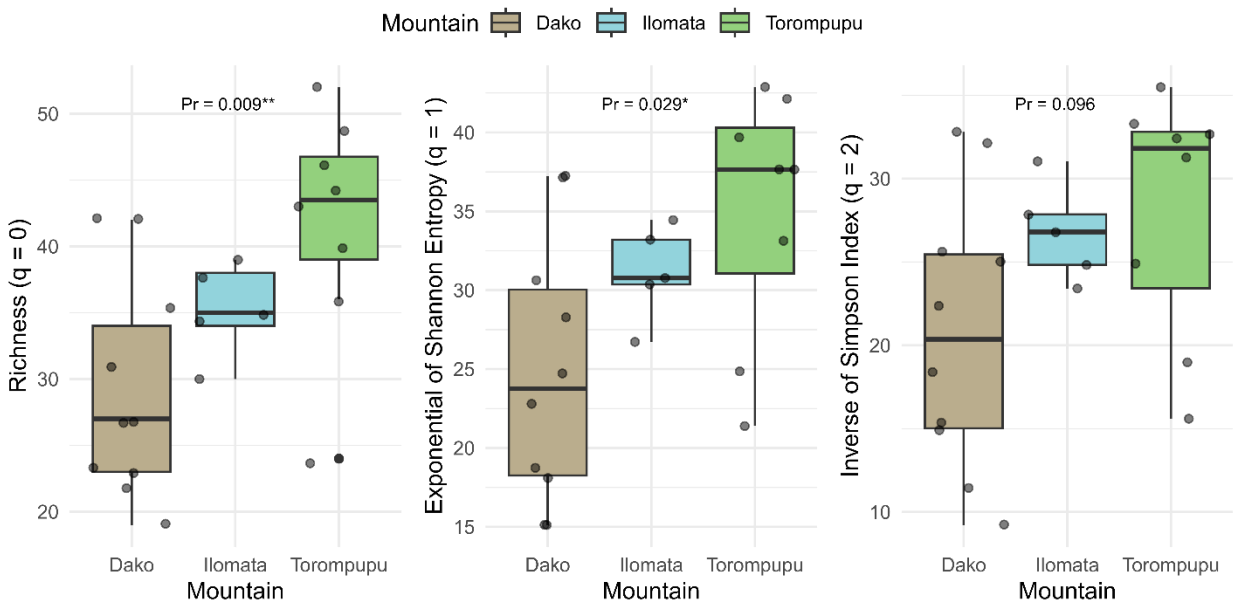


Figure 8. Hill values by mountain using q = 0, q = 1, and q = 2. Dako had the lowest Hill values while Torompupu had the highest.

15 OTUs were shared between all three mountains. Ilomata and Torompupu shared the most OTUs followed closely by Dako and Ilomata and lastly Dako and Torompupu (Figure 9). Communities clustered in ordination space predominantly by mountain (Figure 10). Using PERMANOVA, mountain, elevation, and an interaction between mountain and elevation proved significant (Pr < 0.001; Supplementary Materials, Table 5). The interaction between mountain and elevation was the strongest explanatory variable (R^2 = 0.249, F = 1.429), followed by elevation (R^2 = 0.215, F = 1.544) and mountain (R^2 = 0.153, F = 2.193). No group dispersion differences were identified using either elevation group or mountain (Pr = 0.448, Pr = 0.427).

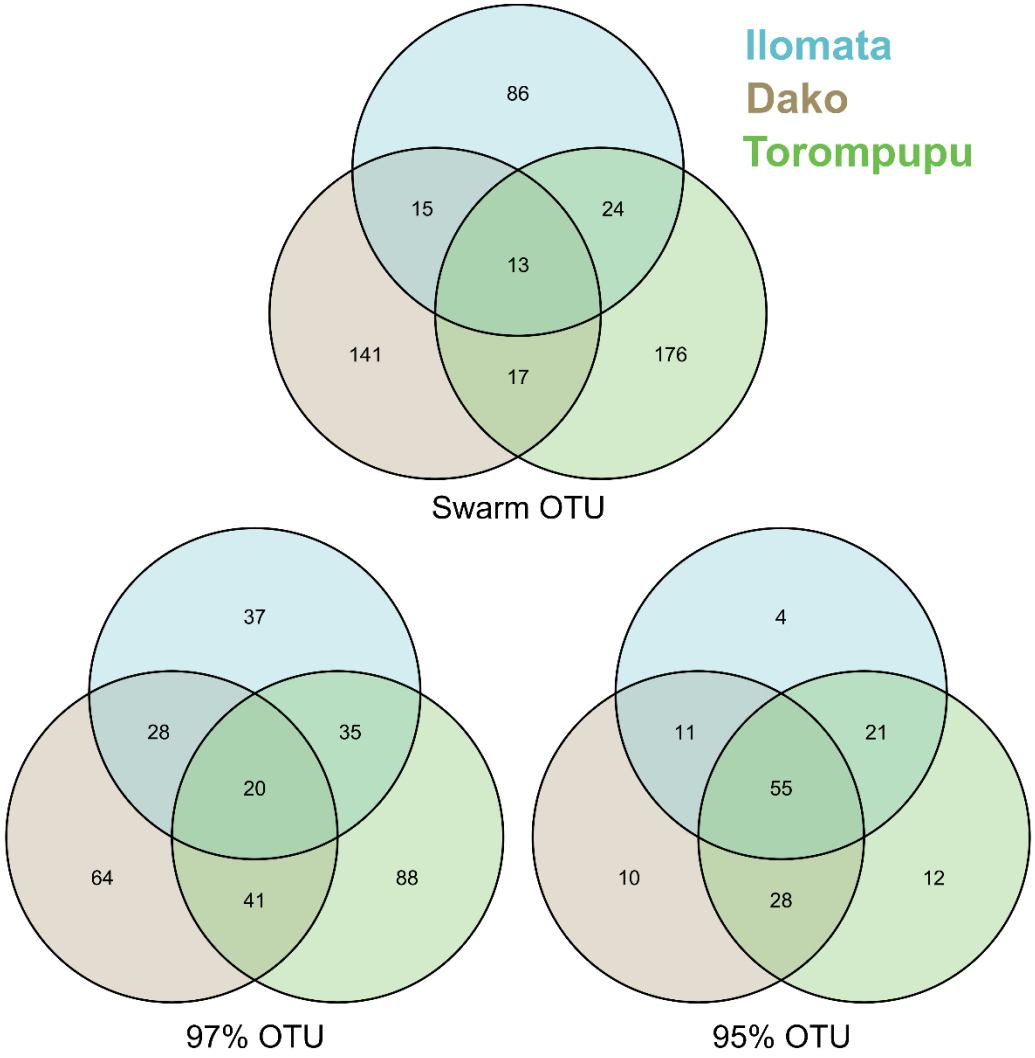


Figure 9. Venn diagrams showing the number of clusters at different thresholds shared between mountains. The fewest shared clusters were when using swarmv2, at 13 shared OTUs. This number of shared clusters increased with coarseness, with 95% threshold resulting in 55 shared clusters.

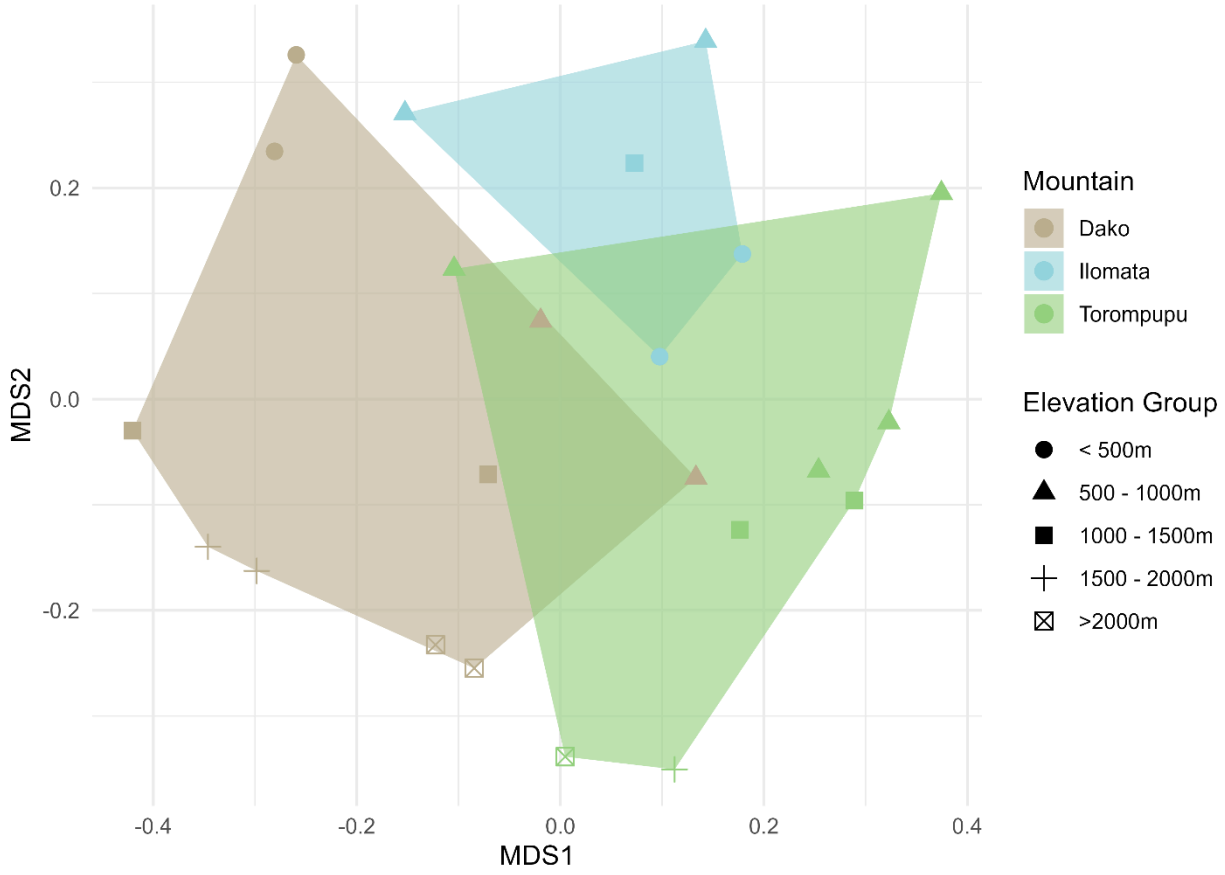


Figure 10. NMDS plot based on phylogenetic distance. Sites clustered by both mountain and elevation, with the most similar communities across the mountains occurring at the low-mid elevations.

1.3.4. Estimating diversification using different clustering thresholds

Patterns across mountains

As discussed above, few OTUs were shared across mountains (Figure 9). At 97%, a more generous species-like grouping, the number of shared clusters increases to 24 of a total 337, still a small percentage of 7.12% clusters shared across mountains. The most clusters are shared between Dako and Torompupu, a total of 49 compared to the 32 clusters shared between Dako and Ilomata and the 33 shared between Ilomata and Torompupu. Dako has 85 distinct clusters (25%) and Torompupu has 75 distinct clusters (22%). Clustering at 95%, representing genera and/or subfamilies, greatly condenses sequences, producing 133 clusters. Nearly one half of clusters are shared across all mountains (62 of 133, 47%). Each mountain contains few unique 95% clusters, Ilomata with the fewest at 4 clusters (3%) followed by Torompupu with 11 clusters (8%) and Dako with 13 clusters (10%). Dako and Torompupu share 26 clusters, 20% of the 95% clusters. The majority of Ilomata's clusters are found across all mountains, but 11 distinct clusters are found on only Ilomata and Torompupu, compared to the 6 shared only between Dako and Ilomata. This differs from the other clustering thresholds, in which Ilomata shares a similar number between both mountains.

Patterns across elevation groups

Each elevation band hosted a high number of unique OTUs (Figure 11). Only 1 OTU was found across all elevations. The low to mid-elevation groups shared the most OTUs, in particular the 500-1000m and 1000-1500m bands with 32 OTUs shared. The >2000m sites shared the fewest OTUs with any other elevation. At 97%, the number of shared clusters across elevation increases only by one and patterns remain largely similar. More clusters are shared within neighboring elevation bands, most notably for the highest elevation grouping that shows more shared 97% clusters with 1000-1500m and 1500-2000m. The 500-1000m site showed a high number of shared clusters between <500m and 1000-1500m, 21 and 24 clusters respectively. The most notable changes occur using a 95% clustering threshold. More clusters are shared across multiple elevation groupings, with a total of 14 shared across all elevation groups. However, this represents only 11% of all clusters and so still a low number of lineages shared broadly across elevation groups. As in the other Venn diagrams, 500-1000m sites contain the highest number of unique clusters, while other groupings lose a large number of unique clusters and share clusters across multiple elevation groups.

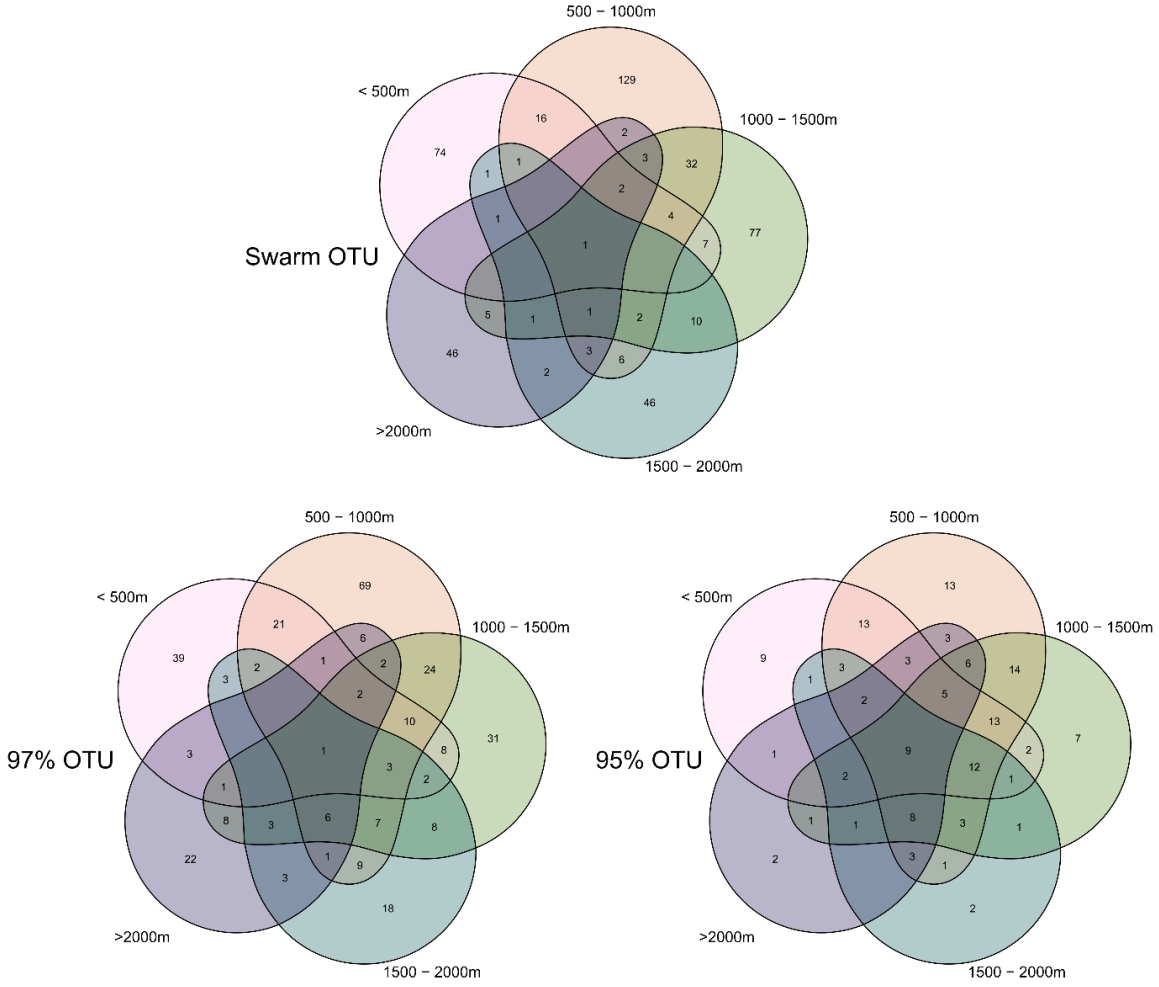


Figure 11. Venn diagrams showing number of shared clusters across elevation groups using different clustering thresholds.

1.3.5. Exploration of evolutionary relationships

A phylogeny constructed using OTUs, sequences queried from BOLD, and an arachnid outgroup showed multiple clearly-defined clades (Supplementary Materials, Figure 3). Distinct clades with multiple OTUs from Sulawesi and sequences from specimens collected in Asia and Australia were examined more carefully; examples from Linyphiidae, Oonopidae, Thomisidae and Araneidae are displayed (Figure 12). The clade within family Linyphiidae shows close affinities to Asian taxa. Araneidae, Thomisidae, and Oonopidae have clades both closely related to Asian taxa and related to Australian taxa. There were many OTU clusters that did not form clear affinities and clustered independently from all other BOLD sequences. Theridiidae had one distinct clade sister to spiders from Asia; however, a single sequence of Asian origins did occur in this clade.

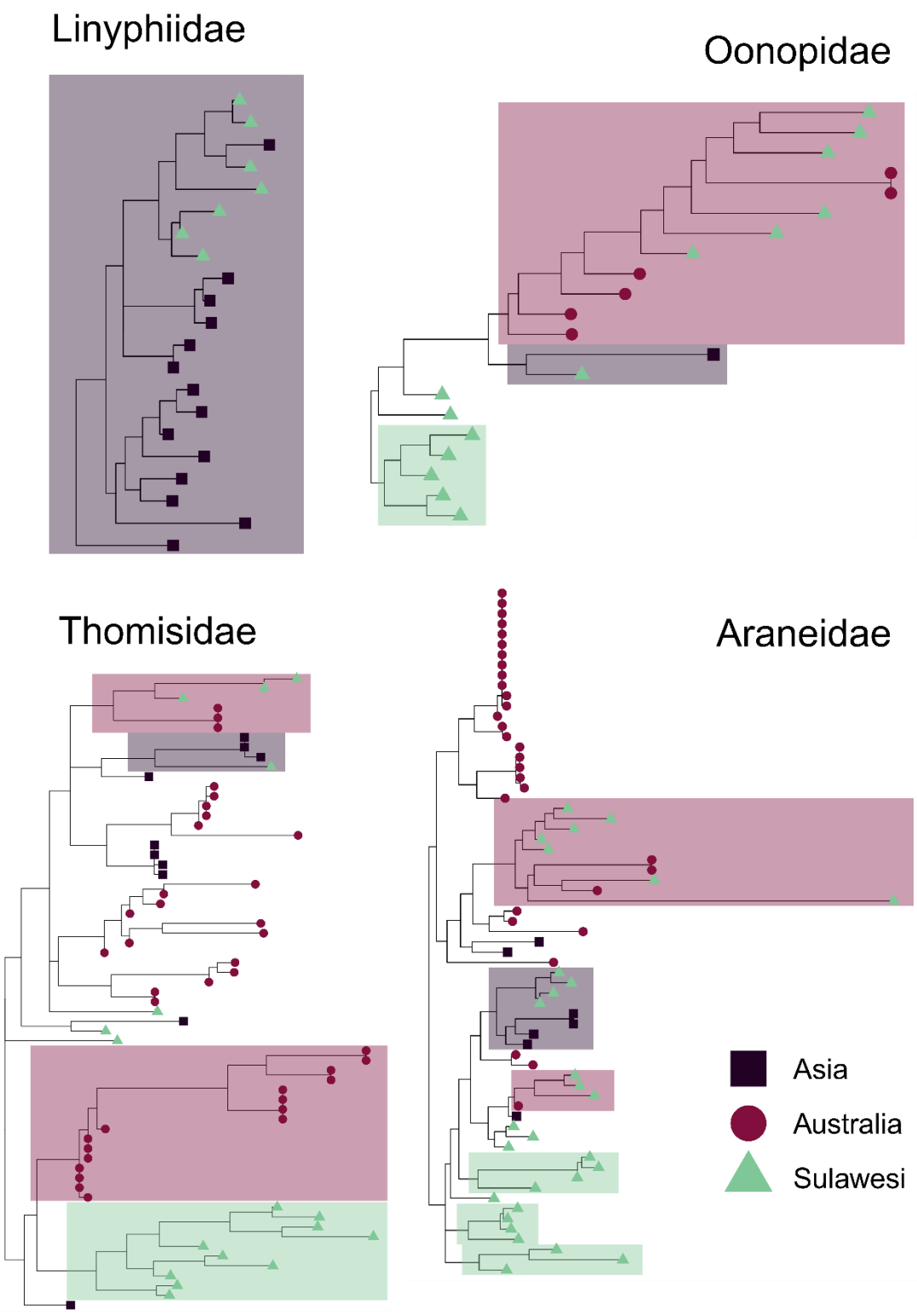


Figure 12. Clades in family Thomisidae, Araneidae, Linyphiidae and Oonopidae, showing relationships to taxa in Australia and Asia, as well as displaying distinct clades most distantly placed from known taxa.

1.4. Discussion

The use of OTUs generated from a mini-barcode region allowed us to test biodiversity patterns and community structure without the immediate need for morphological species identifications. Using these data, we documented novel biodiversity and found evidence for multiple well-established biodiversity patterns, including the elevational diversity gradient, areas of endemism, and the possibility of Sulawesi as a faunal transition zone for multiple spider lineages.

1.4.1. Documentation of biodiversity

The curated sequence data produced by the COI mini-barcode approach greatly expanded our knowledge of Sulawesi spider biodiversity. There are currently 278 species documented for Sulawesi, of which 197 species are endemic to the island[11]. 514 OTUs were defined using swarmv2 and, at a coarser 97% threshold, 337 clusters were defined. This is indicative of the incredible spider diversity that exists on Sulawesi. The utilization of juvenile samples increased the number of detected OTUs by 161, confirming their importance in exploring trends in biodiversity. The 514 OTUs were generated from 2,522 specimens, a relatively small sample size for arthropods; it is clear this is only the tip of the iceberg for the spider biodiversity on Sulawesi.

Similarly impressive is the high density of families we were able to detect in this relatively small sampling. We detected 26 families, three of which are currently not recorded on Sulawesi: families Theridiosomatidae, Scytodidae, and Hersiliidae (Figure 3). There are only three known species of theridiosomatids in Indonesia, all documented from Sumatra; we detected 16 OTUs on Sulawesi alone. There was a single species of hersiliid described from Sulawesi in 1877 but is now considered a *nomen dubium*[52]; however, this family is more widely documented across Indonesia and, therefore, was unsurprising to document on Sulawesi. Native species of scytodids in Indonesia have only been documented in Papua and the Aru Islands (Kabupaten Kepulauan Aru), possibly representing an Australasian lineage and an example of Sulawesi as a transitional faunal zone.

We could classify very few OTUs to species using BLAST; this is likely a result of both the extremely short fragment of COI screened here and the lack of closely related taxa in GenBank[53]. Six OTUs produced above 99% matches, including *Steatoda cingulata, Cyclosa bifida, Leucauge decorata, Leucauge celebesiana, Herennia* sp. and *Neoscona* sp. *Steatoda cingulata* is known from Java and Sumatra and other parts of South and Southeast Asia but not Sulawesi. *Cyclosa bifida* is documented in the Philippines and New Guinea but, again, not Sulawesi. In total, there are only two *Cyclosa* species noted to be in Indonesia, *Cyclosa caligata* and *Cyclosa seriata*, and these are endemic to Sumatra and Java, respectively. No members of *Cyclosa* nor *Steatoda* are currently known from Sulawesi. The other species detected (*Leucauge decorata, Leucauge celebesiana, Herennia* sp. and *Neoscona* sp.) are known from Sulawesi. The incredibly small number of BLAST matches reflects the paucity of information available not only for spiders of Sulawesi but for spiders in Indonesia and the Indomalayan region broadly. Because of the approaches adopted in our study, specimens can be further identified and used to generate a more robust reference collection for future research.

1.4.2. Patterns of biodiversity

Areas of endemism

Distinct species or haplotypes across Sulawesi have been documented for many vertebrate lineages, leading to the definition of areas of endemism across the island. This was first clearly defined by the macaques of Sulawesi[6] and more recent work has further supported these AOEs, such as in the *Draco* lizards[5] and endemic artiodactyls[9]. A similar pattern was detectable for spiders when using the mini-barcode region: we see each mountain acting as a distinct AOE, grouping distinctly in ordination space (Figure 10). We detected both highly distinct OTUs and ASVs on each mountain as well as distinct haplotypes across mountains (Figure 4). Diversity between mountains remained largely distinct until clustered at a 95% threshold, which may be considered a genus- or subfamily-like grouping level.

Elevational diversity gradient

Within-mountain diversity rivaled between-mountain diversity, with patterns of high elevational turnover. Mountains are known to be biodiversity hotspots themselves, associated with their complex topography that creates high habitat diversity and distinct climatic zones[54,55]. Sites at 500-1000m had the highest OTU richness regardless of clustering threshold. This aligns with the well-documented pattern of mid-elevational peaks in diversity, known as the elevational diversity gradient[55,56]. Sites below 500m were sampled only on Dako and Ilomata; despite their similar elevation, the sites sit far from one another in ordination space (Figure 10). This may be linked to the role of environmental filtering and disturbance in community composition. Lowland habitat on Dako has been converted to mixed agroforestry (Figure 2) while the lowland sites of Ilomata, located in a protected reserve, contain some of the only intact lowland forest on Sulawesi. The conversion of lowland forest on Dako could be related to low species diversity found on the mountain as well (Figure 7).

1.4.3. Patterns of diversification

The highest elevation sites were found to be lowest in OTU richness as well as share fewer ASVs with other elevational groupings. The highest number of shared clusters for > 2,000m sites were between 500-1000m sites at all clustering thresholds. This interesting pattern, in which the high elevation habitat does not share the most clusters with neighboring elevations, could be related to the age of the habitat; high elevation habitat on Sulawesi is young, emerging roughly 3Mya[7]. Rather than taxa from lower elevations diversifying into high-elevation habitat, communities at these elevations may be composed of more recent arrivals that are pre-adapted to high elevation conditions. This was found to be the case on Mount Kinabalu in Malaysia, another relatively young mountain; the biodiversity consisted mostly of pre-adapted lineages that colonized similar habitat, rather than descendants of lowland species that shifted to higher elevations [57].

1.4.4. Phylogenetic patterns

Sulawesi as a faunal transitional zone is well-documented in vertebrates but underexplored in invertebrates. While not robust, our preliminary analysis shows clades that appear to be closely related to both Asian and Australian taxa as well as clades that appear deeply diverged from known species and could represent relict taxa. We may then hypothesize that spiders show the

classic biogeographic pattern of this island, with lineages bridging the Australasian and Indomalayan regions. These results must be taken with a grain of salt – not only is the mini-barcode region too short to make strong phylogenetic hypotheses, but many spider species across this region are missing from the BOLD database and from our knowledge. This makes it difficult to say with certainty any one lineage is Australasian or Asian in origin. Larger genomic datasets and targeted collection of certain taxa across the Indomalay region will be necessary to test these patterns fully.

1.5. Conclusion

In the face of a biodiversity crisis, rapid assessment and monitoring of community-level diversity is essential in maximizing conservation efforts. The use of a mini-barcode to sequence whole communities of spiders allowed us to explore patterns of biodiversity and produce hypotheses based on our findings that can guide future research. We found high turnover both across mountains, supporting the presence of areas of endemism, as well as high turnover across elevations, including the classic pattern of an elevational diversity gradient. Clustering at different thresholds allowed us to detect differences across potential taxonomic levels and explore patterns of diversification. Paired with COI data extracted from BOLD, we were additionally able to identify certain clades that may represent Australasian lineages and support a hypothesis that Sulawesi acts as a faunal transitional zone for certain spiders. Adopting an approach such as ours can begin to build important biodiversity knowledge in a financially feasible manner that can be produced quickly, generate data for most specimens regardless of preservation method, and require no need for specialized taxonomic expertise. As studies grow and methods improve, the problems noted in this paper, such as low BLAST matches and multiple OTUs per individual, will continue to be reduced. While the short COI region used is not suitable for robust phylogenetic or biogeographic analyses, it was successful in identifying basic patterns of biodiversity across spiders and within lineages that together greatly increase our knowledge about the spiders of Sulawesi.

1.6. Supplementary Materials

Mountain	Site	Latitude	Longitude	Elevation
Dako	DKA1	1.08846	120.8848	338
Dako	DKA2	1.0755	120.889	495
Dako	DKB1	1.06264	120.8986	974
Dako	DKB2	1.06212	120.8995	992
Dako	DKC1	1.05335	120.9087	1380
Dako	DKC2	1.05325	120.9099	1435
Dako	DKD1	1.0497	120.9168	1751
Dako	DKD2	1.05043	120.918	1743
Dako	DKE1	1.06016	120.9349	2238
Dako	DKE2	1.05991	120.9344	2238
Torompupu	TPPA1	1.40231	119.9455	642
Torompupu	TPPA2	1.40133	119.946	646
Torompupu	TPPA3	1.40321	119.9442	656
Torompupu	TPPB1	1.41847	119.9303	999
Torompupu	TPPC1	1.42122	119.8989	1370
Torompupu	TPPC2	1.42184	119.8963	1335
Torompupu	TPPD1	1.4182	119.8804	1860
Torompupu	TPPE1	1.41179	119.8674	2227
Ilomata	ILOA1	0.70341	123.1907	324
Ilomata	ILOA2	0.70129	123.1883	320
Ilomata	ILOB1	0.71615	123.1854	873
Ilomata	ILOB2	0.71681	123.1874	902
Ilomata	ILOC1	0.72815	123.1943	1211

Supplementary Table 1. Coordinates and elevation for sites across elevation gradients on the three sampled mountains.

Taxonomic Summary	
Family ID	# of OTUs
Theridiidae	92
Araneidae	68
Tetragnathidae	43
Linyphiidae	40
Thomisidae	35
Salticidae	23
Uloboridae	17
Oonopidae	16
Theridiosomatidae	16
Sparassidae	15
Coriniidae	14
Zodariidae	12
Psechridae	9
Cheiracanthiidae	8
Pholcidae	7
Ctenidae	6
Oxyopidae	6
Clubionidae	5
Hahniidae	5
Hersiliidae	5
Lycosidae	5
Pisauridae	4
Mimetidae	3
Deinopidae	2
Scytodidae	2
Mygalomorph	1

Supplementary Table 2. Summary of the different families found using the mini-barcode. The right column shows the number of OTUs detected for each family.

Linear Regression

	q = 0	*q = 1*	*q = 2*
Observations	23	23	23
R^2	0.085	0.169	0.245
Adjusted R^2	0.041	0.130	0.209
Residual Std. Error	9.109	7.885	6.817
F-Statistic	1.939	4.278	6.810
Pr	0.178	0.051	0.016*
Note:	$^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$		

Supplementary Table 3. Results from linear regression, with elevation as predictor and Hill values as dependent variables.

q = 0

Statistic	*N*	*Mean*	*St. Dev.*	*Min*	*Max*
Df	2	11.000	12.728	2	20
Sum Sq	2	951.652	338.772	712.104	1,191.200
Mean Sq	2	207.806	209.652	59.560	356.052
F value	1	5.978		5.978	5.978
Pr(> F)	1	0.009**		0.009	0.0a09

q = 1

Statistic	*N*	*Mean*	*St. Dev.*	*Min*	*Max*
Df	2	11.000	12.728	2	20
Sum Sq	2	785.719	447.760	469.105	1,102.333
Mean Sq	2	144.835	126.880	55.117	234.553
F value	1	4.256		4.256	4.256
Pr(> F)	1	0.029*		0.029	0.029

Statistic	*N*	*Mean*	*St. Dev.*	*Min*	*Max*
Df	2	11.000	12.728	2	20
Sum Sq	2	646.248	531.326	270.544	1,021.952
Mean Sq	2	93.185	59.520	51.098	135.272
F value	1	2.647		2.647	2.647
Pr(> F)	1	0.096		0.096	0.096

Note: $^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$

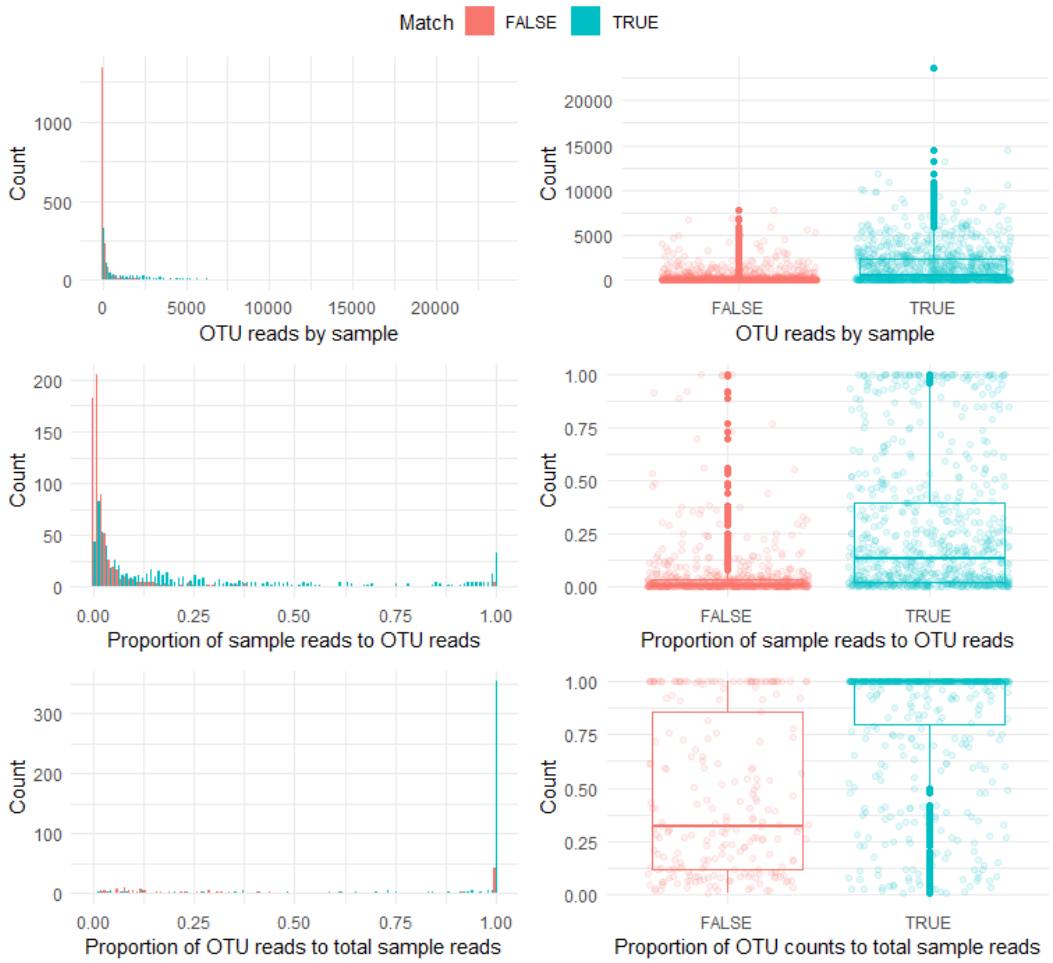
Supplementary Table 4. Results of ANOVA, testing association between elevation group and Hill numbers.

PERMANOVA results

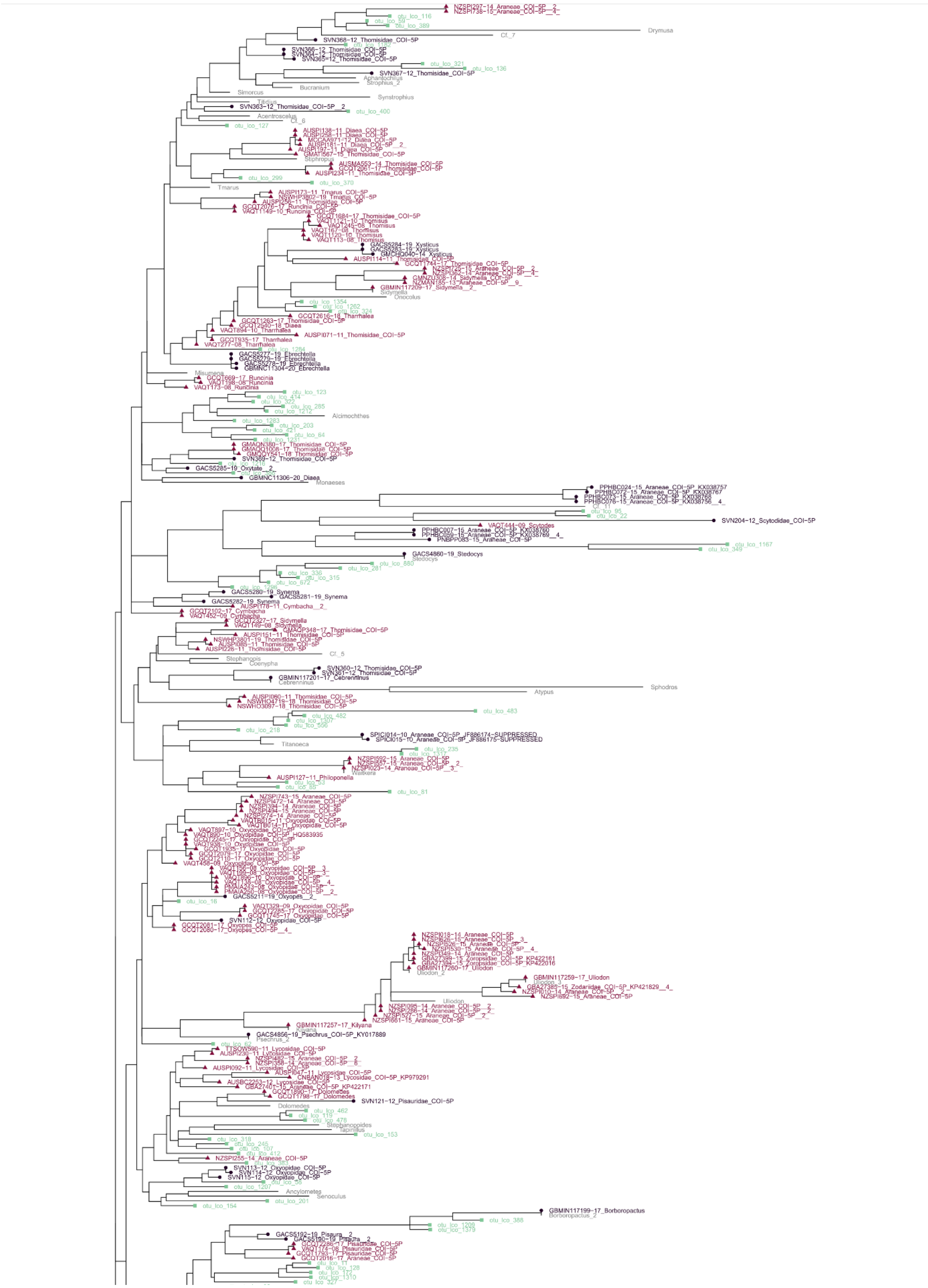
	:		
	Mountain	*Elevation group*	*Interaction*
Df	2	4	5
Sum of Sqs	0.884	1.244	1.440
R^2	0.153	0.215	0.249
F-statistic	2.193	1.544	1.423
Pr(>F)	0.001***	0.001***	0.001***
Note:		$^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$	

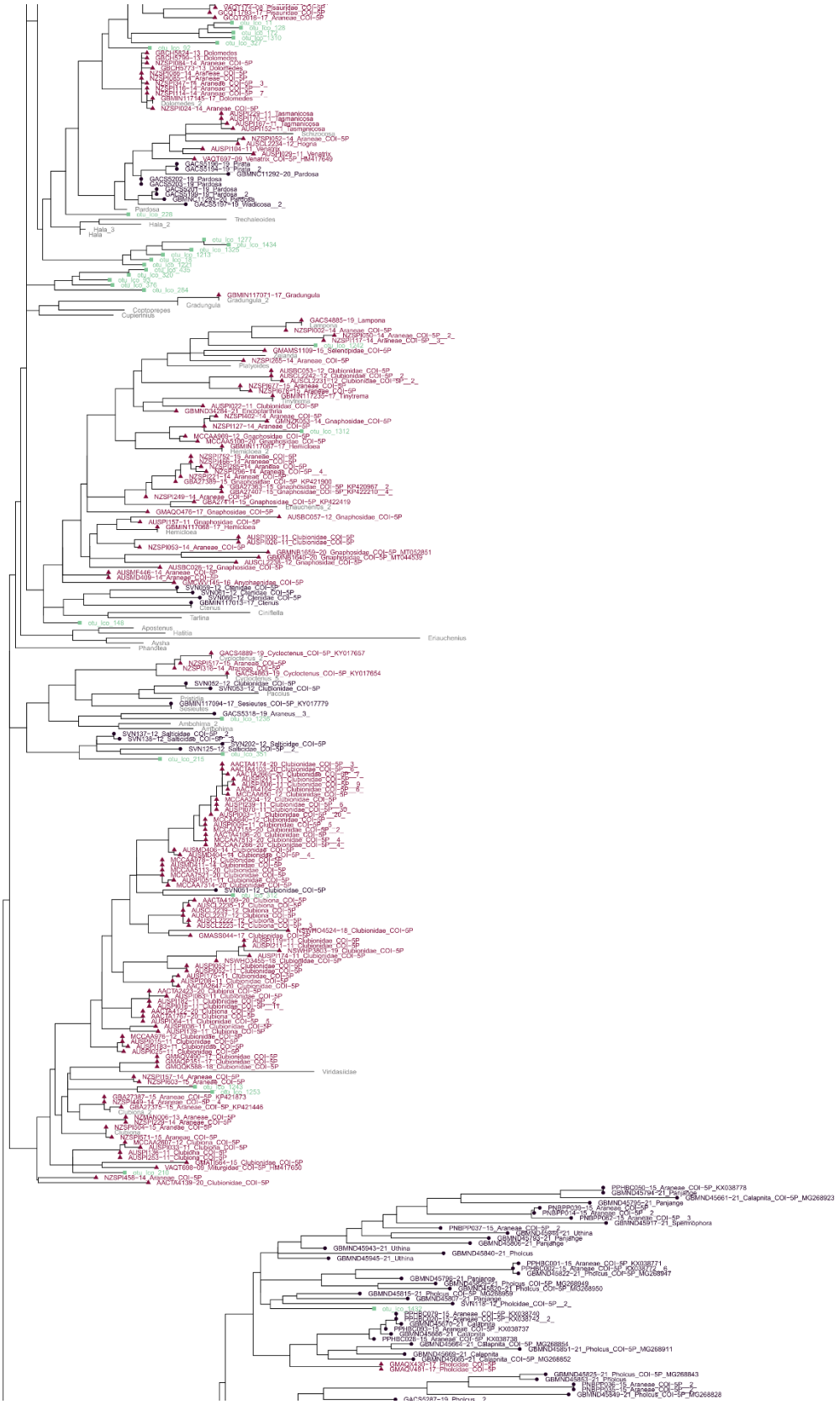
Supplementary Table 5. Results from PERMANOVA, using mountain, elevation group and an interaction between mountain and elevation group as the independent grouping variables.

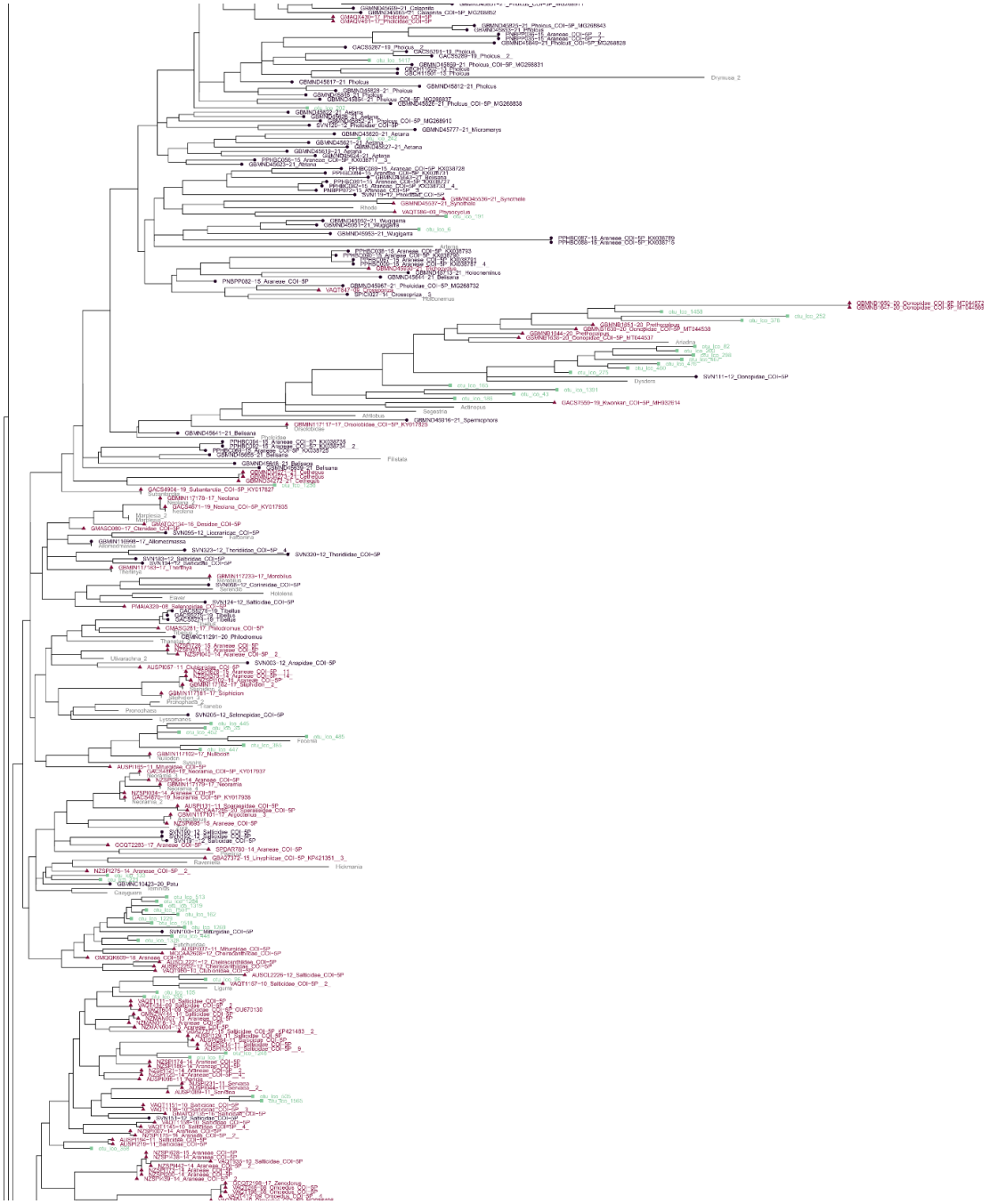
Supplementary Figure 1. Little to no sample degradation occurred following 3-hour lysing for adult specimens, leaving morphological characters intact for future morphological work. Level of degradation depended on degree of sclerotization as well as size.

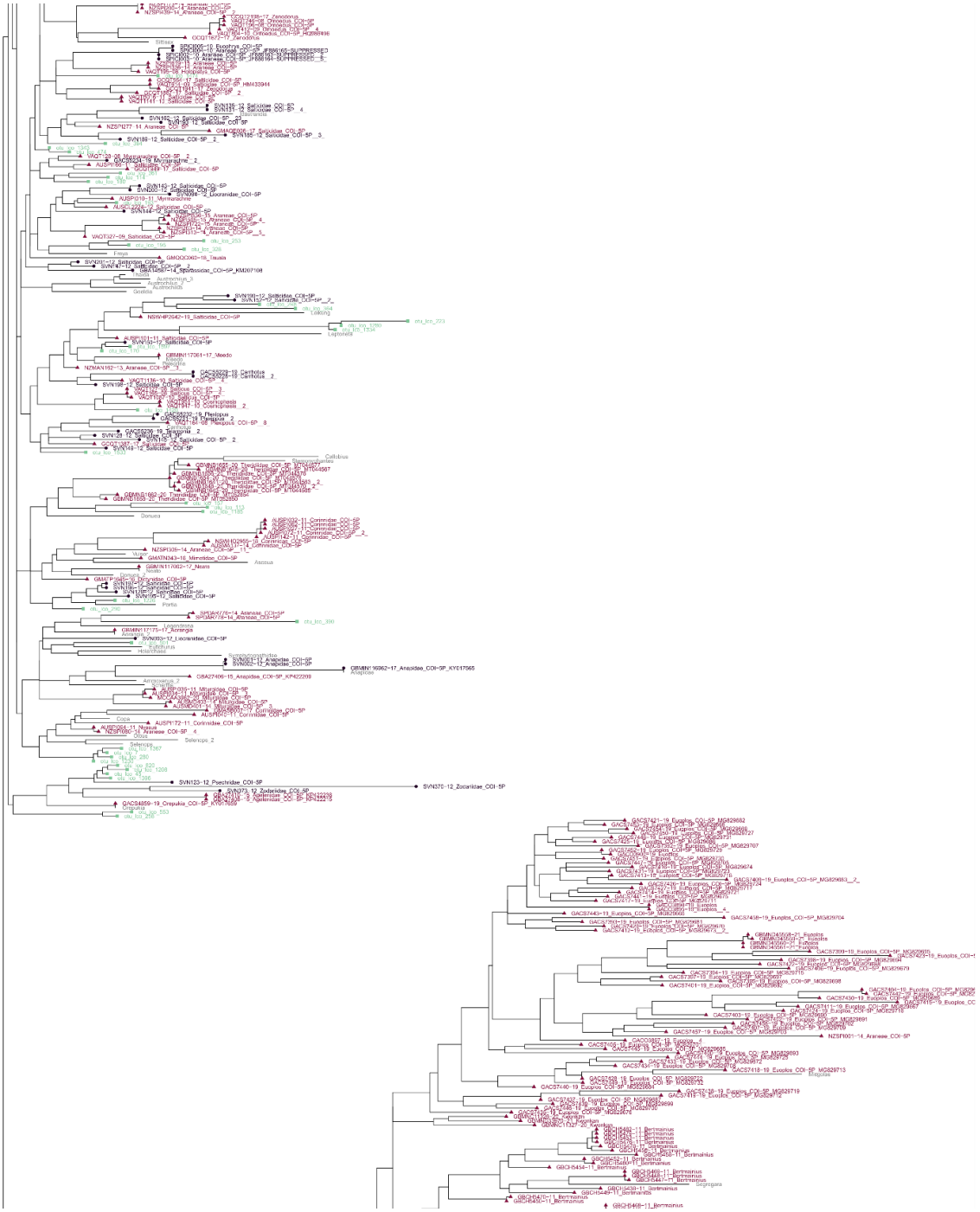


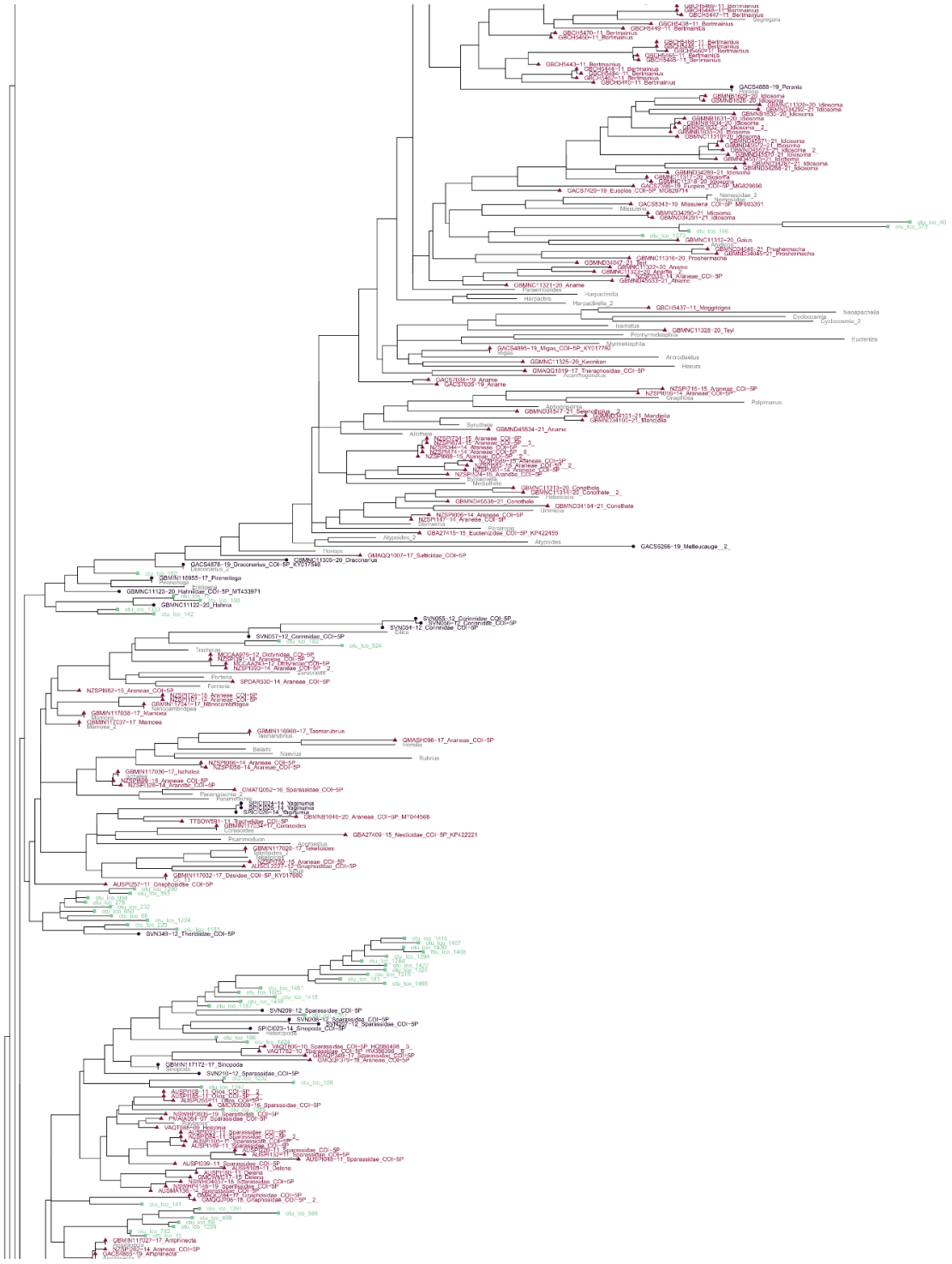
Supplementary Figure 2. In order to narrow the data to one OTU per sample, morphological identifications were compared to the OTU identifications for a given sample then different metrics were assessed to determine filtering thresholds, based on the 1) number of reads associated with an OTU in a sample, 2) the proportion of the OTU reads in a sample to the total OTU reads, and 3) the proportion of OTU reads in a sample to the total reads in a sample. False match indicates the morphological identification did not match the OTU identification of an OTU associated with a sample.

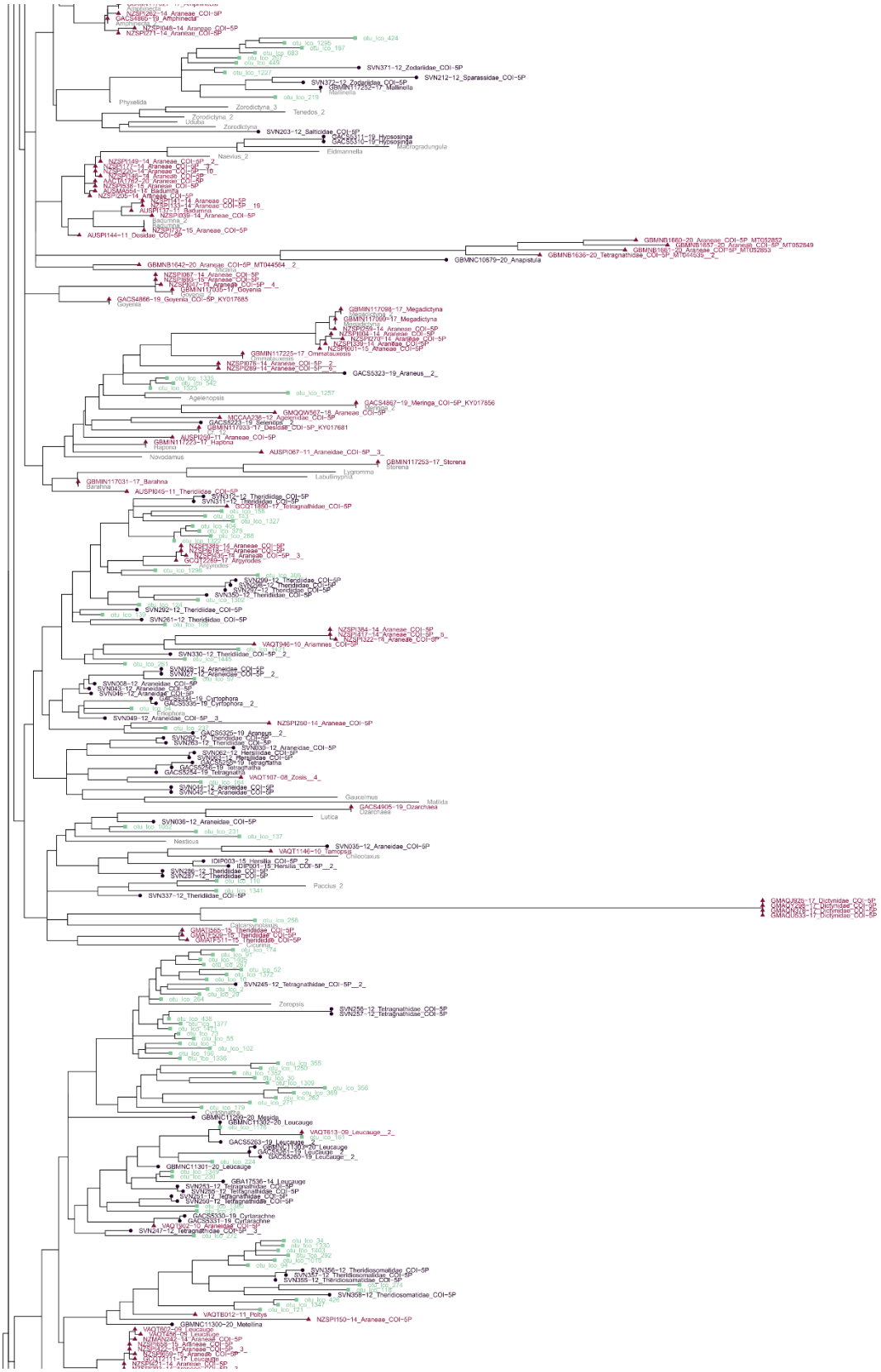


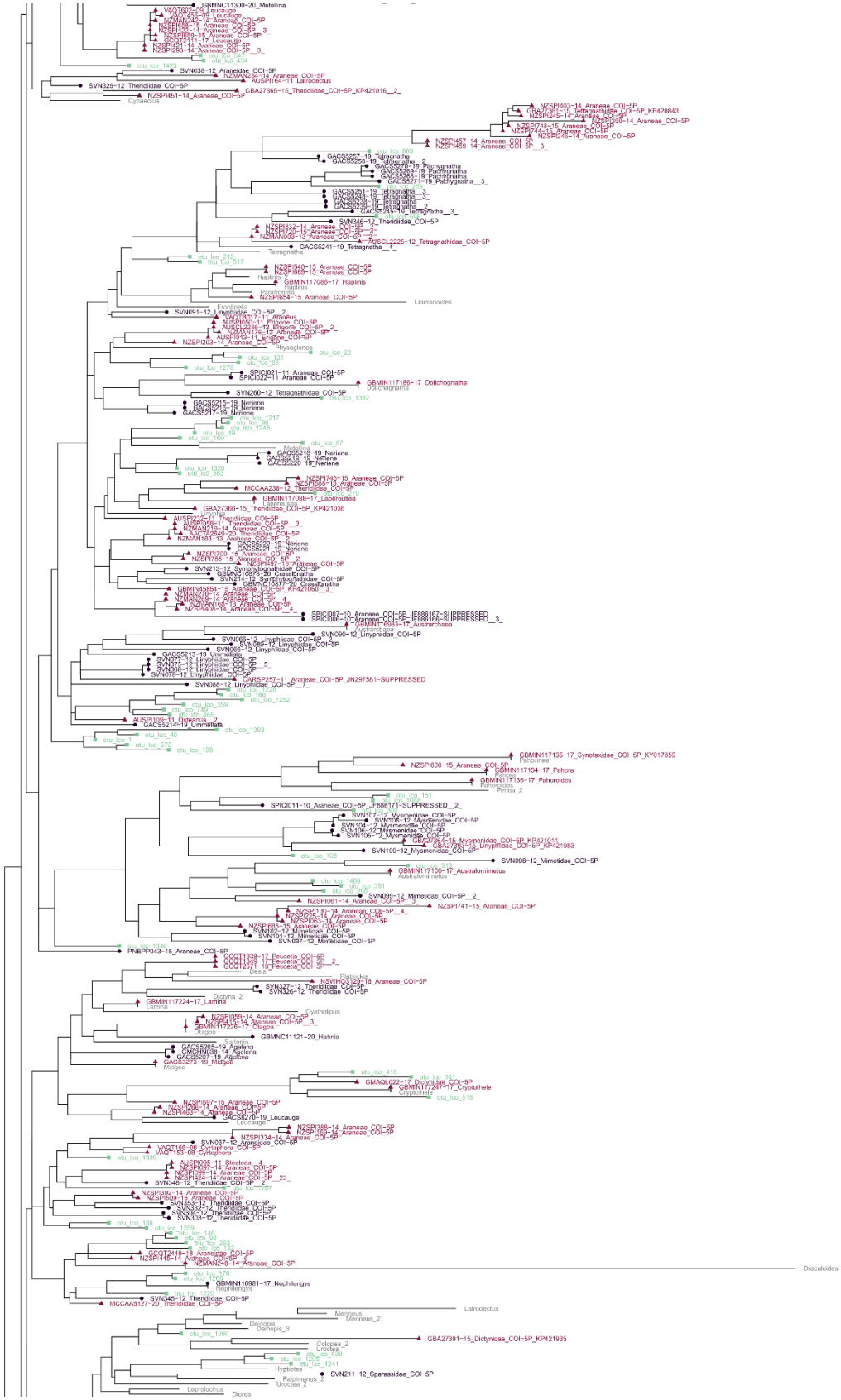


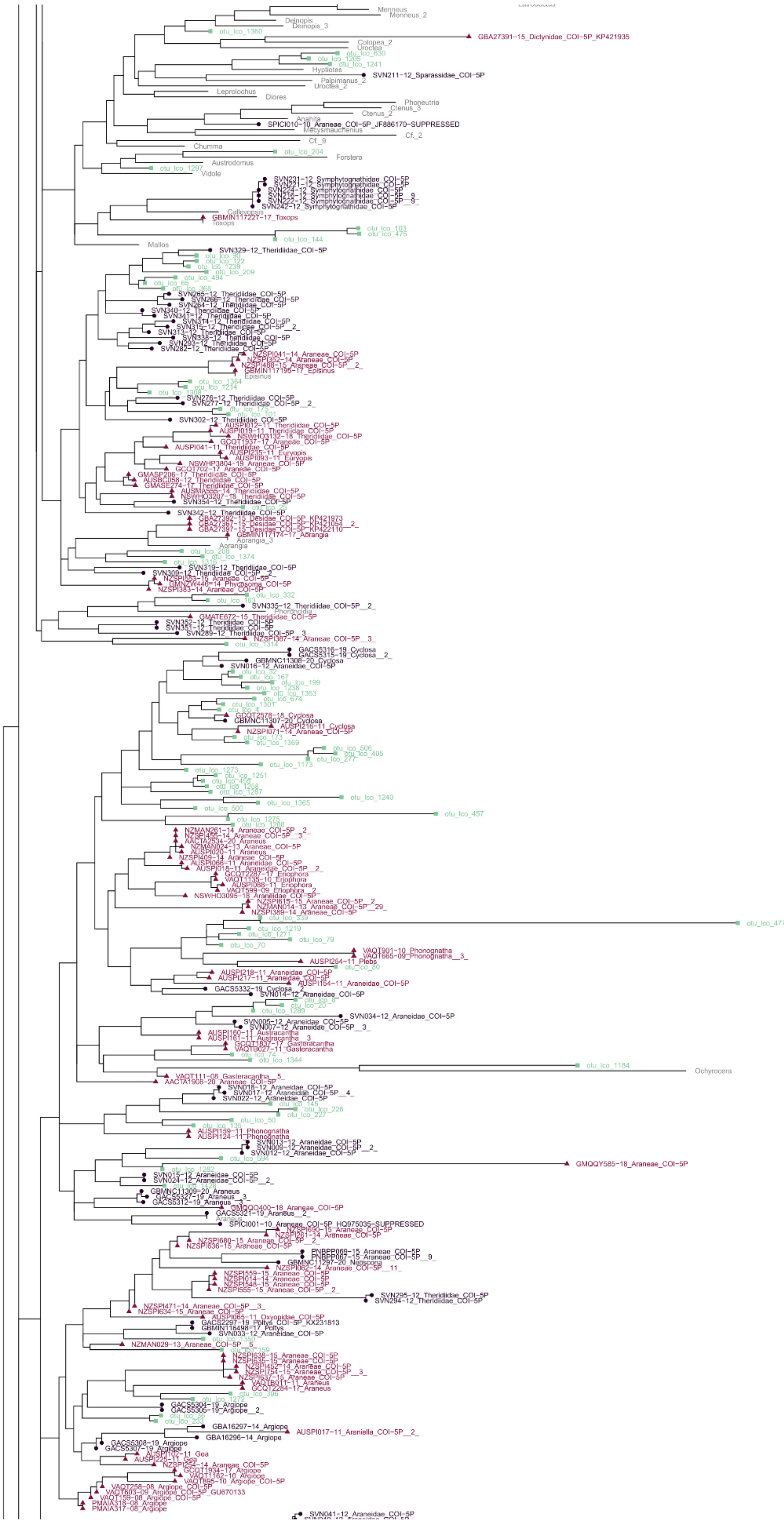












Supplementary Figure 3. Complete mini-barcode phylogeny constructed with OTU sequences and sequences downloaded from BOLD. Green labels with circle tips are Sulawesi OTUS, pink labels with triangle tips are BOLD sequences from Asian localities and purple labels with square tips are BOLD sequences from Australian localities.

Based on the results of **CHAPTER 1**, I identified the family Tetragnathidae to be a particularly intriguing group of spiders on Sulawesi, with what appeared to be extensive speciation across the three mountains sampled. In **CHAPTER 2**, I turn my focus to this family and studied patterns of colonization and *in-situ* diversification across multiple genera.

Chapter 2: Patterns of colonization and *in-situ* diversification of the tetragnathid spiders in Sulawesi, Indonesia

2.1. Introduction

The island of Sulawesi within the biodiversity hotspot of Wallacea has captivated biogeographers and biodiversity scientists alike for hundreds of years because of its distinct endemic diversity that has ties to fauna in both Australasia and Sundaland[58]. Alfred Russel Wallace, often called the father of biogeography, recognized Sulawesi as a fascinating location in the late 1800s[59] and wrote extensively on the fauna of the Indomalayan region[60–62]. In his paper on the geography of the region[61], he first proposed a demarcation between the Indomalayan region to the west of Sulawesi, and the islands to the east of this boundary, which he named the Austro-Malayan region; the line drawn became known as Wallace's line and led to the subsequent identification of alternative faunal divides in the region (e.g., Huxley's Line[63] and Lydekker's Line[64]) as well as the broader definition of zoological zones globally[65,66].

The zoological regions described in the 19th century remain supported by the more extensive zoological data we have today and are found to have been created through multiple forces, including plate tectonics and distinct climatic transitions[67]. The presence of four abutting biodiversity hotspots in the Indo-Australian region, each with its own pool of endemic diversity, is thought to be tied primarily to the paleogeology of the region, where multiple continental fragments moved together over the course of 150Ma to form the current extensive scattering of islands[68,69]. The particularly fascinating diversity of Sulawesi, consisting of both Australian and Indomalayan lineages and *in-situ* diversification structured into areas of endemism (AOEs)[5,6], is related to the island's complex geological history. Multiple previously submerged continental fragments collided ~23Ma[69], creating a primarily lowland archipelago that existed for ~15Ma[7]. This paleoarchipelago showed variability in available land area, number of areas above sea level, degree of island separation and topographical complexity through this time. Around 4Ma, the island underwent significant changes driven by tectonic and volcanic activity as well as changes in sea level. This resulted in the formerly disconnected paleo-islands becoming connected, and multiple mountainous regions forming with some peaks above 3000m[7].

This history has long been associated with the areas of endemism on Sulawesi[6]. In the modern era of genomic data, multiple studies have estimated colonization times in line with island formation, followed by rapid diversification in association with the connection of island fragments. A recent large phylogenomic study based an extensive island-wide sampling of *Draco* lizards was the first to explicitly incorporate the complex geological history of Sulawesi and provides clear evidence towards the role of the island's tectonic history in driving speciation[5]. Additional taxa showing patterns linked to Sulawesi's paleogeology include the tarsiers[8], artiodactyls[9], and squirrels[70]. The few studies that exist on invertebrate animals show astonishing amounts of *in-situ* diversification, such as in the terrestrial *Trigonopterus* weevils[22]

and the aquatic *Tylomelania* land snails[71,72]. These invertebrate studies display new stories not documented in vertebrates and can provide novel insight into both biogeography of the region as well as formation of biodiversity more broadly.

The formerly disconnected island fragments that existed for ~15Ma are thought to have been colonized through long-distance dispersal events. In some mammal lineages, species arrived early in the island's formation and underwent anagenesis but did not diversify. The emergence of more land area around 1-2Ma has been proposed as a driver of species expansions in some lineages[9]. Alternatively, taxa could have colonized Sulawesi and dispersed between island fragments within the paleo-archipelago. This pattern is clearly shown in *Draco* lizards[5] as well as in some insect taxa[73].

The formation of mid- to high-elevation habitat provided another opportunity for extensive speciation. While elevation above 1,000m was thought to have been formed upon fragment collision, this subsided and was not present in the 15Ma reconstruction; it was not until 10Ma that elevation above 1,000m formed stably, albeit with very limited extent[7]. More recent tectonic activity in the past 3Ma increased the extent and the elevation of mountain ranges. This opening of novel niche space was likely filled in one of two ways – either through ecological diversification, with lowland taxa expanding upwards and adapting to fill a new niche, or by pre-adapted lineages from outside of Sulawesi colonizing and maintaining their ancestral niche. Diversification patterns associated with mountain emergence are documented in some mammal groups[70,74].

To further unravel the complexity of Sulawesi fauna and understand formation of endemism associated with island formation and mountain building, we chose to study spiders in the family Tetragnathidae (Araneae). Tetragnathidae is a globally distributed, abundant and diverse family[11] that is often disproportionally common on islands. A notable example of their diversity is the radiation of *Tetragnatha* (Araneae, Tetragnathidae) in Hawaii[75]. In this example, spiders showed significant ecological divergence across Hawaii, with one lineage abandoning web building to form the "spiny-leg" clade of active hunters[41] with four distinct "ecomorphs" that co-occur[77]. A similar trend was noted in the web-building species, with distinct web "ethotypes" being identified that emerge convergently across islands[78]. In addition to their capacity for extensive species radiations, the group has been found to be capable of long-distance dispersal events, often found as one of the first groups to colonize *de-novo* habitats[79,80]. Tetragnathid spiders provide a compelling system to study species formation on islands, where evolution and ecology intertwine.

The main objectives of our study were to describe the diversity of family Tetragnathidae in Sulawesi more thoroughly and determine the phylogenetic relationships and evolutionary history of the prevalent genera, including the timing of colonization and patterns of *in-situ* diversification. We hypothesized that lowland taxa will be characterized by (1a) deep divergences, representing taxa that colonized early in Sulawesi's formation and diversified between paleo-islands. We may also expect to observe (1b) priority effects due to the open lowland habitat and dispersive capabilities of tetragnathids. In contrast to this, we hypothesize that mountain tops, being young and novel environments, will be characterized by (2a) more

recent divergences from species outside of Sulawesi that were pre-adapted to high elevation habitat. Ancestors of these taxa could have arrived via ballooning or other forms of long-distance sweepstakes dispersal, presenting contrasting patterns to those observed in less vagile terrestrial vertebrates. Because of the more specialized strategies needed for high elevation environments, we may also expect that (2b) sister species will be found between mountains rather than within mountains, as these more dispersive taxa could move to fill the niche to which they are best adapted.

2.2. Methods

2.2.1. Sample collection

Samples were collected in conjunction with field seasons described in Chapter 1, along elevation gradients on three mountains (Figure 1); these mountains will be referred to as TPP (Gunung Torompupu), DK (Gunung Dako) and ILO (Gunung Ilomata). Tetragnathids were collected using the standardized protocol described in Chapter 1. Field conditions prevented web documentation at some sites; this was especially true at the highest elevation on Gunung Torompupu, where multiple nights of high wind prevented web documentation. Specimens from webs were additionally collected during day collection, or during timed night collection, without associated web photographs. Samples were stored in 95% ethanol and at –20C until further processing.

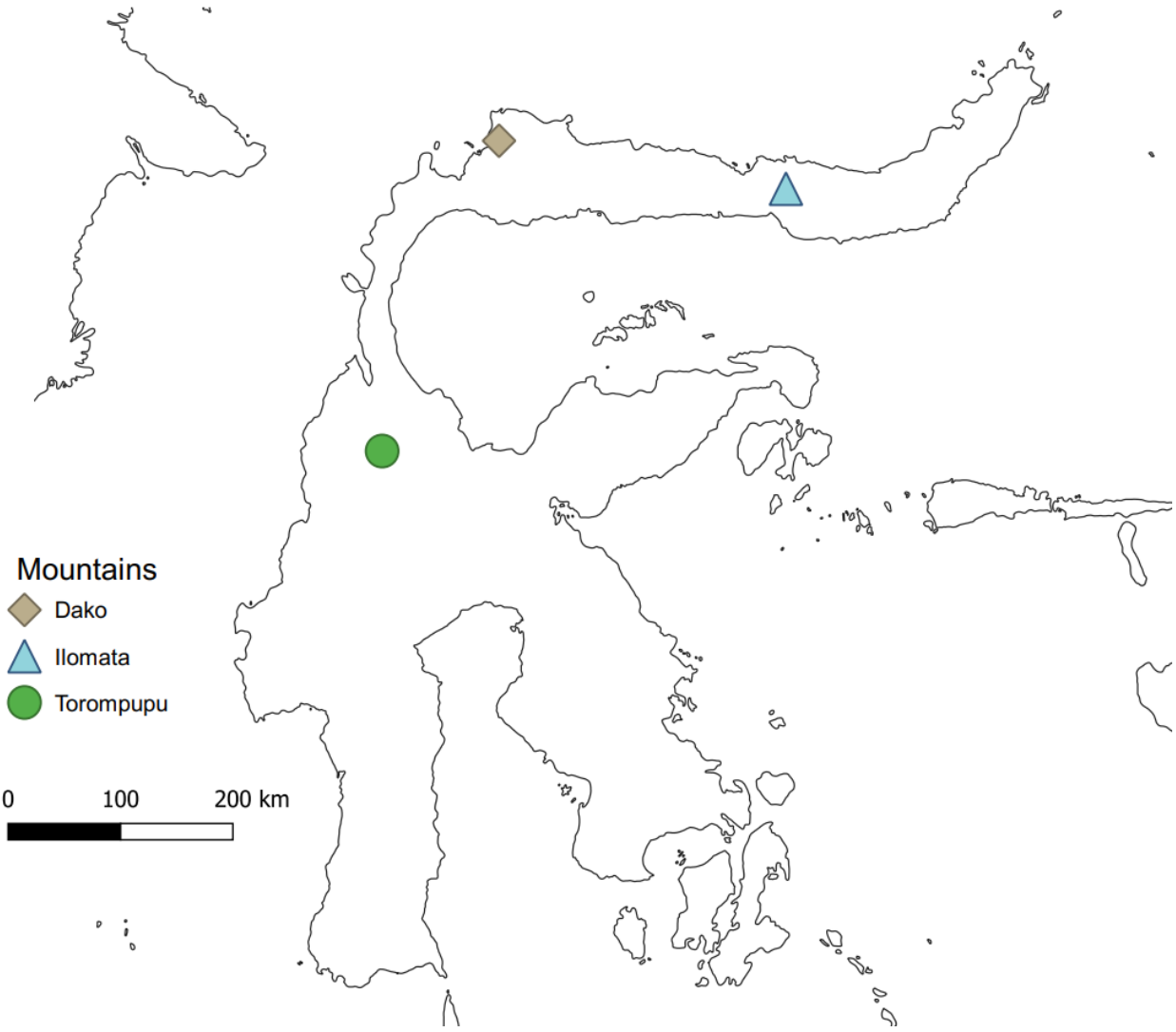


Figure 1. Map of mountains sampled on Sulawesi - Torompupu (TPP), in green, Dako (DK) in brown, and Ilomata (ILO).

2.2.2. Molecular procedures

DNA extractions were performed as described in Chapter 1. In addition to amplification of the short COI barcode, an additional four loci were amplified: mitochondrial 16s and nuclear H3, 18s and 28s. A longer fragment of COI was amplified to be combined with the COI "mini-barcode". The primers used are listed in Table 1; these primers were successfully used to construct a well-resolved phylogeny for other spiders[81]. Qiagen Multiplex PCR kit (Qiagen, Hilden, Germany) was used to amplify the mitochondrial and nuclear regions in two separate PCRs in the same volumes as noted in Chapter 1. Amplification of the nuclear loci was performed at an annealing temperature of 55°C for 35 cycles. Amplification of the mitochondrial loci was performed at an annealing temperature of 46°C for 35 cycles. The remaining steps were identical to the procedures described in Chapter 1; the generated amplicons were pooled with the

mini-barcode amplicons into a final library and sequenced in parallel; primers were used to demultiplex after sequencing.

Locus	bp	Forward	Sequence 5′-3′	Reverse	Sequence 5′-3′
COI	467	ARF1[82]	GCNCCWGAYATRGCNTTYCCNCG	Fol-degen-rev[83]	TANACYTCNGGRTGNCCRAARAAYCA
18s-1	421	SSU-F04[84]	GCTTGTCTCAAAGATTAAGCC	SSU-R22[84]	GCCTGCTGCCTTCCTTGGA
18s-2	351	18s_2F[85]	AACTTAAAGRAATTGACGGA	18s_4R[85]	CKRAGGGCATYACWGACCTGTTAT
16s	371	16S_F2[81]	AATYCAACATCGAGGTCGCAA	16s_R2[81]	TRACYGTRCWAAGGTAGCAT
28s	363	28s_F2[85]	TTTTGGTAAGCAGAACTGGYG	28s_4R[85]	ABTYGCTACTRCCACYRAGATC
H3	374	H2aF[86]	ATGGCTCGTACCAAGCAGACVGC	H3aR[86]	ATATCCTTRGGCATRATRGTGAC

Table 1. Primer combinations, all found successful in amplifying and constructing a well-resolved Bayesian phylogeny for family Tetragnathidae.[81]

2.2.3. Bioinformatics

Cutadapt was used both to trim primers, and to extract reads associated with each primer pair into separate FASTA files[38]. DADA2 was performed individually for each primer pair[39]. For any sample, an ASV must have made up at least 1% of total reads for that ASV as well as made up 1% of total reads for an individual sample.

Samples identified as Tetragnathidae based on sequencing results were confirmed morphologically. Next-generation sequencing approaches can generate erroneous reads that result in incorrect sequences being associated with a sample; index hopping is one major cause during sequencing[87,88]. In addition, even very minimal contamination or crossover can result in OTUs in low abundance in samples, due to the read depth of these platforms. These problems were largely addressed using the filtering approaches previously explained. However, to ensure the sequences were likely correct, phylogenies were constructed for each gene region. These included GenBank sequences representing every family of spider from the Wheeler et al. order level phylogeny[89],. Sequences were aligned using MAFFT[90]; the L-INS-I approach was used for protein coding regions and the Q-INS-I approach was used for the rDNA regions[91], which considers the secondary structure of RNA. A maximum likelihood phylogeny was estimated using IQTREE v1.6.12. Any sequence outside of the clearly defined Tetragnathidae clade was removed. Following the removal of non-Tetragnathidae sequences, the data frame of all samples and associated sequences was manually assessed to check for low read counts. Sequences that were found in low abundance in a sample were assessed using their topological placement on the associated phylogeny; if the sequence placement was similar to the placement of other gene regions associated with the sample, the sequence was retained. While our filtering approaches resulted in the loss of samples, we preferred to be conservative in our approaches to ensure confidence in the sequences we included in further analysis.

As outgroups, and to determine the number and approximate source of colonization events, we used taxa from previous studies, collected in the islands of Pohnpei and Kosrae of Micronesia in July 2006. We also queried sequences from GenBank representing every known genus of Tetragnathidae as well as attempted to include every sequenced species known from South-Southeast Asia to Australia. Additionally, sequences from *Mimetus* and *Austromimetus* (Araneae: Mimetidae) as well as *Nephila* (Araneae: Araneidae) were included as outgroups. Again, all sequences were aligned using MAFFT and the L-INS-I approach or the Q-INS-I approach. Accession numbers can be found in Supplementary Materials: Table 1.

Once we were confident in the representative sequences for each individual, regions with many gaps in rDNA alignments were removed using TrimAl[92] hosted on phylemon[93]. Gene regions were concatenated using SeqMat[94], which adds ambiguous characters for missing data while retaining gaps that may have been retained in the ribosomal loci. Samples sequenced from our project that had less than 800bp of data were removed. However, the specimens from Pohnpei and Kosrae had only COI data. Our final dataset consisted of 202 samples.

2.2.4. Divergence dating

To perform divergence dating, we used BEAST.v1.10.4[95]. BEAUti.v1.10.4[95] was used to construct the appropriate XML file for analysis. The mitochondrial clock rates were linked while all other clock rates were estimated independently. Based on the best partition scheme produced from ModelFinderPlus[48] as implemented in IQTREE v1.6.12, the GTR substitution model with gamma and invariant sites and estimated base frequencies was used 16s rDNA and the 1st and 2nd codon positions of COI. For 18s rDNA, 28s rDNA and the 1st and 2nd codon positions of Histone-3, the GTR substitution model with gamma and invariant sites was used as well; however, base frequencies were set to equal. For the 3rd codon position of COI and Histone-3, the HKY substitution model was used. Gamma and invariant site heterogeneity and estimated base frequencies were used for COI, while Histone-3 had equal base frequencies and no site heterogeneity model (Table 2).

Gene	ModelFinderPlus	BEAST
16s rDNA	TIM2+F+I+G4	GTR+I+G4; estimated base frequencies
18s rDNA	TIM2e+I+G4	GTR+I+G4; equal base frequencies
28s rDNA	TVMe+I+G4	GTR+I+G4; equal base frequencies
COI (1st and 2nd codon)	TN+F+G4	GTR+I+G4; estimated base frequencies
COI (3rd codon)	TN+F+I+G4	HKY+I+G4; estimated base frequencies
H3 (1st and 2nd codon)	TVMe+I+G4	GTR+I+G4; equal base frequencies
H3 (3rd codon)	K2P+R2	HKY; equal base frequencies

Table 2. Partition scheme proposed by ModelFinderPlus and the site model implemented in BEAST.

We used an uncorrelated relaxed clock with a lognormal distribution and a calibrated Yule tree prior as well as a random starting tree; however, monophyly of the ingroup was enforced, as well as monophyly between Mimetidae and Tetragnathidae. The fossil record for spiders is sparse, and the use of fossils is still widely debated[96]. Because of this, we used a mutation rate for the mitochondrial and nuclear loci for calibration. Additionally, we used secondary calibrations obtained from Garrison et al., 2016[97]; this study used a large phylogenomic data set at the order-

level along with well-supported spider fossils and produced estimated dates of divergence for major nodes.

A normal prior with the ucld.mean equal to 0.0112 site^{-1} My^{-1} and ucld.stdv set to 0.01 was used for the mitochondrial genes[98]. A diffuse uniform prior with a lower bound of 0.0001 and upper bound of 0.0115 was used for the nuclear genes[99]. A normal prior was set on the root height with a mean of 148My and stdev of 56.12My. An additional normal prior was set on the MRCA of Mimetidae and Tetragnathidae, with a mean of 100My and stdev of 41.11My. Three independent chains were run for 100 million generations with sampling every 10,000 trees and combined using LogCombiner. Tracer.v1.7.2 was used to assess convergence of each chain individually as well as the combined chains (Supplementary Materials, Figures 1 & 2). A 10% burn-in was chosen based on the results in Tracer. TreeAnnotator was run to generate a consensus tree. Trees were first visualized using FigTree.v1.4.4 then figures wer created using ggtree[50] and treeio[51] in R. Species delimitation was performed using the generalized mixed Yule coalescence model with the *gymc* function from the package splits[100]; the input phylogeny was the ultrametric consensus tree produced from BEAST.

No formal phylogeographic analyses were performed because of the limited range of sampling. Patterns of diversification and colonization dates were compared to paleogeological reconstructions of Sulawesi adapted from Nugraha & Hall, 2018[7] (Figure 2) to make hypotheses about the patterns found.

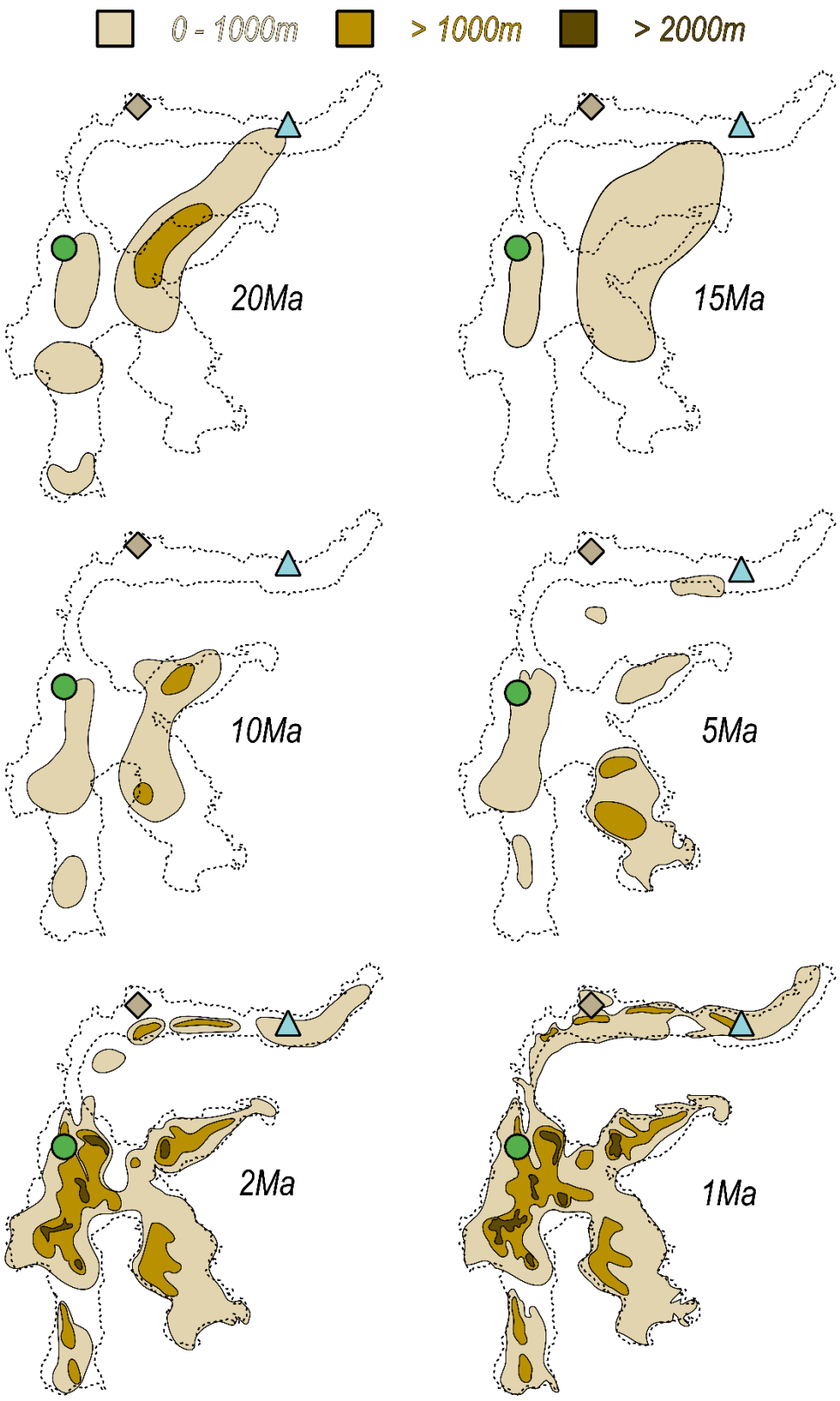


Figure 2. Hypothesized paleogeography of Sulawesi from 20Ma to present. Figures based on Nugraha & Hall, 2018 reconstructions[7]

2.3. Results

2.3.1. Summary of results

A total of 141 Sulawesi spiders were included in the final phylogeny, with an additional 61 samples representing different genera of Tetragnathidae taken from GenBank, and COI data from tetragnathids of Micronesia (Gillespie, R; unpublished). Specimens from Sulawesi were classified to five genera – *Mesida* (Kulczyński, 1911), *Tylorida* (Simon, 1894), *Leucauge* (White, 1841), *Dolichognatha* (O. Pickard-Cambridge, 1869) and *Tetragnatha* (Latreille, 1804) (Table 3). Using GMYC, 42 putative species were defined (Table 3, Figure 3). *Mesida* was the most diverse, with 20 putative species. This was followed by *Tylorida* with 11 species. Both groups were more abundant as well, with 118 specimens collected across both genera during the standardized sampling approaches. *Leucauge* was found less commonly with 18 specimens collected representing 7 putative species. This was true of *Dolichognatha* as well, with only two species and five individuals. Lastly, a single specimen was collected that was placed in *Tetragnatha* (Table 3; Figure 3). Species were distributed by mountain and by elevation. Gunung Torompupu had the highest number of genera and had 17 species. Ilomata had four genera with 10 species, followed by Gunung Dako with only three genera but 17 total species (Figure 7). The number of genera present in a community decreased with elevation; four genera were detected in sites below 1000m, three genera between 1000-1500m, two genera at 1500-2000m and a single genus above 2000m (Figure 8).

Genera	# of putative species	# of samples	# of species in Indonesia	# of species globally
Mesida	20	84	2	14
Tylorida	11	33	2	8
Leucauge	7	18	28	167
Dolichognatha	2	5	2	32
Tetragnatha	1	1	11	324
Total	42	141	45	545

Table 3. Five genera were found in Sulawesi. This table shows the number of putative species estimated using GMYC and the total number of samples. This is compared to the known species in Indonesia and the known species globally for each genus, taken from WSC[11]

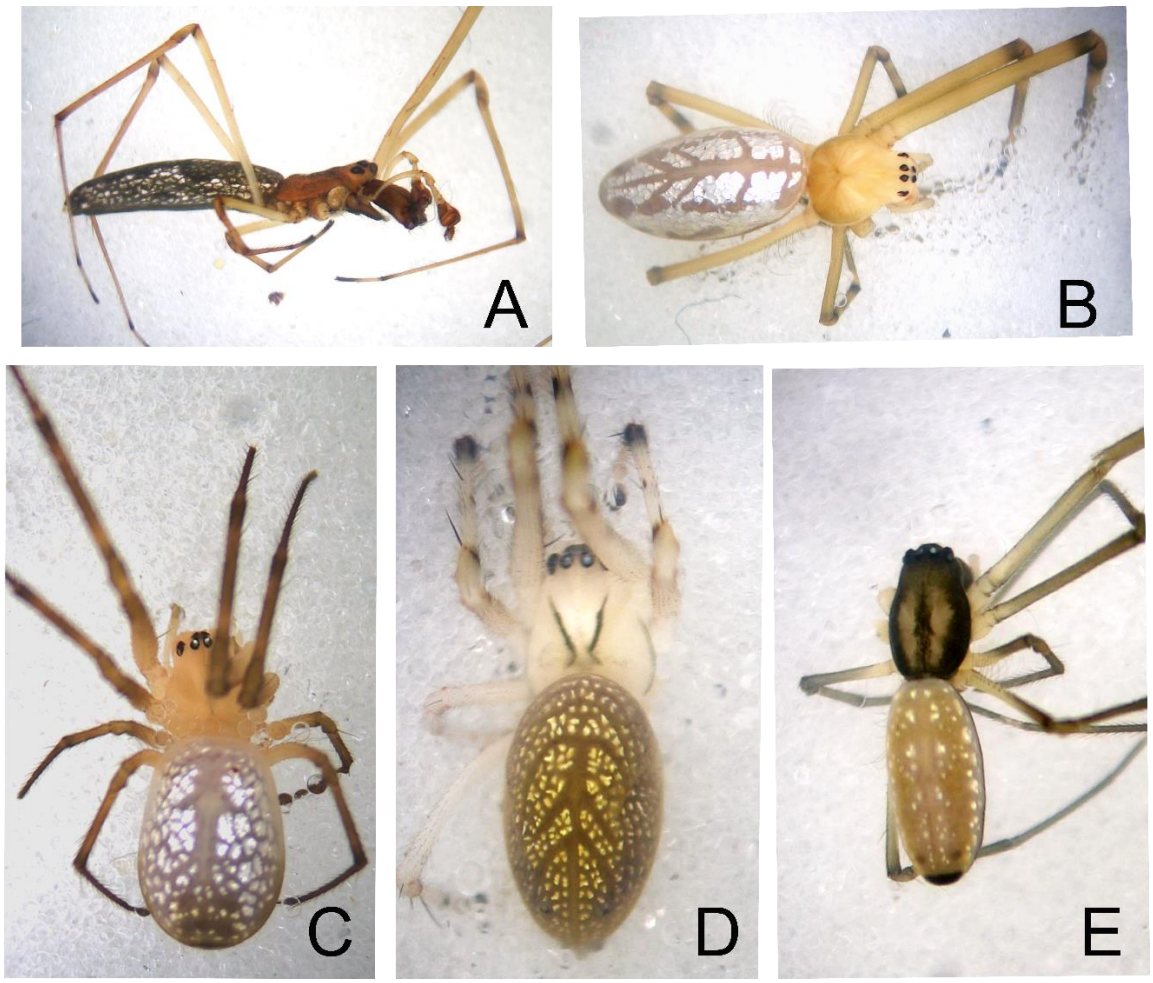


Figure 3. Examples of different specimens from Sulawesi, including A) *Tetragnatha* B) *Leucauge* and C) *Mesida.* D) and E) are two distinct *Tylorida* spp. from Gunung Torompupu.

Figure 4. Images of A) *Dolichognatha* sp. from Gunung Torompupu as well as B) the web from which it was captured.

Figure 5. Second *Dolichognatha* sp. collected from Ilomata, demonstrating many characters similar to family Arachaeidae – narrow, long chelicerae with recursive fang orientation, and elongated neck-like modification to the cephalothorax. The abdomen is much smaller than those of many other *Dolichognatha* spp., such as the Torompupu specimen shown in Figure 5A.

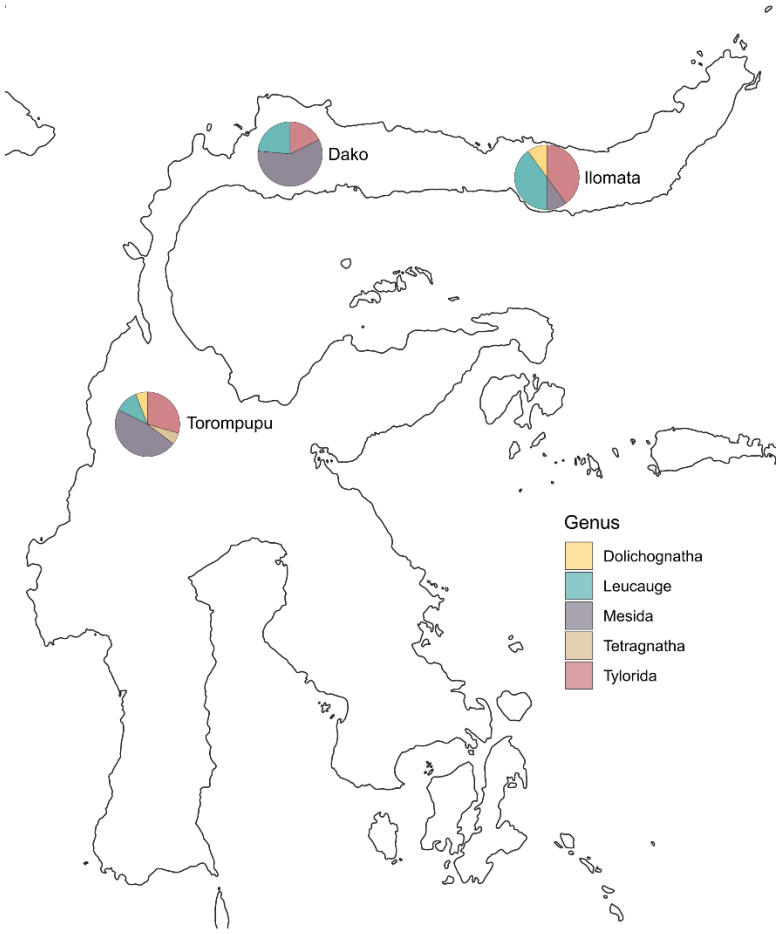


Figure 6. Map of Sulawesi, with pie charts indicating number of species found in each genus on each mountain.

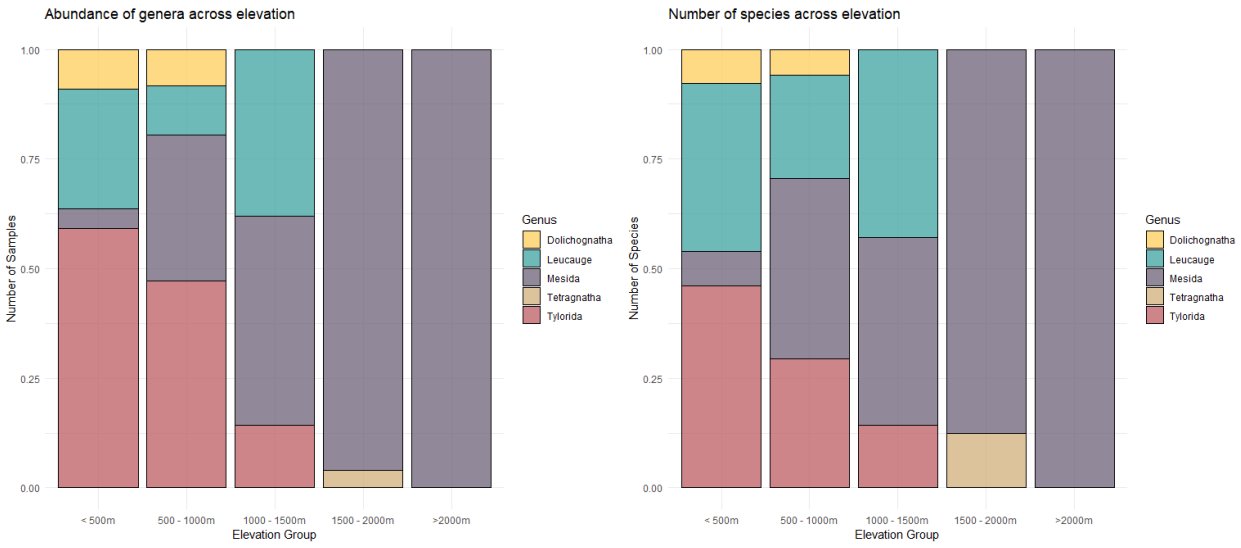


Figure 7. Proportion of (a) individual samples and (b) species belonging to each genus across elevation groups.

2.3.2. Phylogenetic relationships and evolutionary history

For ease of interpretation, a phylogeny with tips collapsed into putative species is included in the main text (Figure 8). The tree with all terminals can be found in the supplemental materials.

Tetragnatha (Figure 9)

The single species of *Tetragnatha* documented at 1860m on TPP was estimated to have split from a clade containing a *Tetragnatha* sample from Singapore and a *Tetragnatha* sample from Reunion Island around 22.13Ma [13.32, 32.32]. This split is well supported, with a posterior value of 0.97.

Dolichognatha (Figure 10)

Two putative species were found in *Dolichognatha*. A single species found on TPP is sister to a species found in India, Myanmar and Thailand (*Dolichognatha longiceps)*(Thorell, 1895*),* as well as an unidentified *Dolichognatha* sample collected from Thailand. The TPP clade is estimated to have diverged 7.14Ma [3.03, 12.80] although with a relatively low posterior value of 0.766. The second Sulawesi *Dolichognatha* is found only on ILO and has a distinct morphology from other known *Dolichognatha;* in fact, it resembles family Archaeidae with the distinct cephalic region and elongated "neck" that inspired Arachaeidae to be referred to as the pelican spider. It is sister to *Dolichognatha incanescens* (Simon, 1895) and an unidentified *Dolichognatha sp.*; however, the posterior value is very low (0.26) making the date of divergence and relationships to other *Dolichognatha* unclear.

Leucauge (Figure 11)

In the genus *Leucauge,* we found 7 putative species on Sulawesi. The closest affinities were to 1) *Leucauge argentina* (van Hasselt, 1882), 2) *Leucauge decorata* (Blackwall, 1864) and 3) *Leucauge celbesiana* (Walckenaer, 1841). *L. decorata* and *L. celebesiana* have both been previously found on Sulawesi. One Sulawesi species complex appears to have diverged from *L. argentina* (known from the Philippines, Taiwan, Singapore, and Sumatra) 16.06Ma and then diverged within Sulawesi 3.57Ma, with one species found on TPP and another on ILO. Additionally, there is a single specimen on ILO belonging to *L. Argentina.* A species found on DK is sister to *L. decorata,* which has a widespread distribution from Africa to South-Southeast Asia and Indonesia. The age of divergence between *L. decorata* and the DK species is estimated at 4.14Ma [1.03, 8.63]. Two species on DK were sister to *L. celebesiana,* known throughout South and Southeast Asia and documented on Sulawesi; the posterior values, however, are very low (< 0.75). A final Sulawesi *Leucauge* clade remains, with species found on all three mountains and not found closely related to any other sequenced *Leucauge* sp. This clade has an estimated divergence of 9.46Ma [5.83, 13.46] but with a low posterior value of 0.38. The splits between mountains were well-supported, with a TPP-DK split estimated at 1.62Ma [0.55, 3.10], and then a TPP+DK – ILO split 2.98Ma [1.54, 4.82]. A third species from ILO represented by one specimen had an estimated divergence of 5.89Ma [3.34, 9.08] from the other species.

Tylorida *(Figure 12)*

11 putative species belonging to *Tylorida* were identified. Only two species of *Tylorida* are known from Indonesia, with *T. striata* (Thorell 1877) known from Sulawesi. One putative species from ILO was sister to *T. striata* and had an estimated divergence of 4.17Ma [1.22, 7.91] with high support (> 0.99). Two additional species, one from ILO and one from DK, were sister to this clade (*T. striata* and ILO sp.) with a split estimated at 8.55Ma [4.98, 13.08]. Relationships between the remaining *Tylorida* clades had low posterior support values. There appears to be a TPP-GOR clade, sister to another clade found only on TPP; this TPP species complex shows a divergence between elevation bands at 4.62Ma [2.18, 7.90] with high posterior support (> 0.99). There is a second clade found on DK and TPP, with a split between mountains estimated at 2.76Ma [0.46, 7.19] and well-supported (> 0.99). Finally, there was an entirely distinct *Tylorida* clade consisting of a clade of specimens collected from Micronesia that were sister to a clade composed of species from TPP, DK and GOR. There was a well-supported divergence (> 0.99) between the species on ILO and the species on DK-TPP dated at 4.88Ma [2.52, 8.35]. The Sulawesi clade was estimated to have diverged from the Micronesia clade 17.90Ma and from the clade containing *T. striata* and all other Sulawesi species 21.93Ma. However, both nodes had extremely low posterior values (< 0.75).

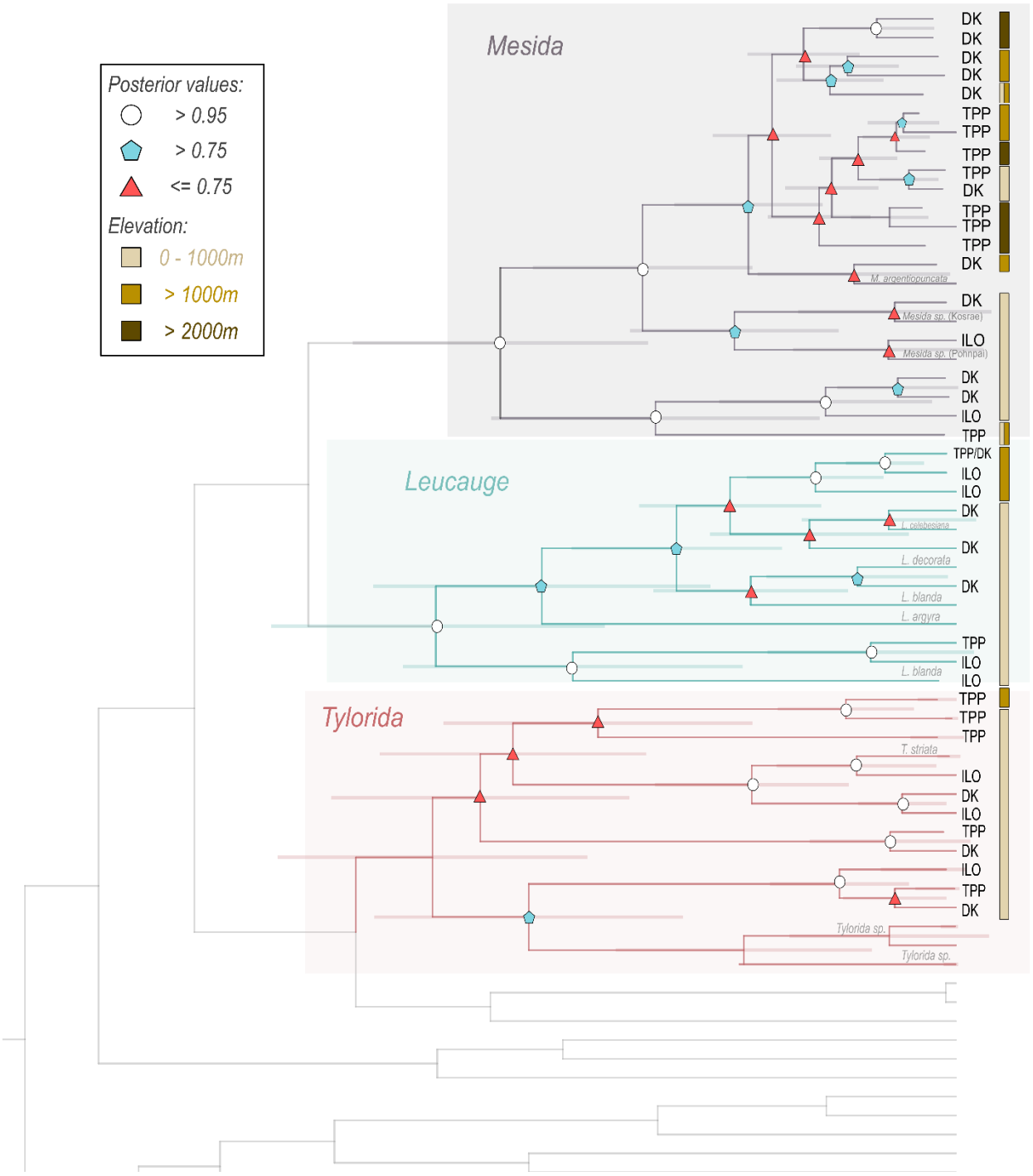
Mesida *(Figure 13)*

Mesida was the most diverse genus with 20 putative species. Monophyly of *Mesida* did not emerge. The species known from South Asia, *Mesida yini* (Zhu, Song & Zhang, 2003), was placed in a clade with *Tylorida*, while the Sulawesi *Mesida* spp. as well as *Mesida argentiopunctata* (Rainbow, 1916) from Australia and two *Mesida* spp. from Micronesia were placed in a separate clade that was sister to *Leucauge*. This is the relationship resolved in other Leucauginae phylogenies (*Mesida* as sister to *Leucauge*) which is then sister to *Tylorida*.

On Sulawesi, there was a single clade with high support that had an estimated divergence 19.09Ma [13.58, 25.95] from the rest of the Sulawesi *Mesida*. Within this clade, there were species on each mountain found between 800-1,300m. TPP specimens split from the DK-ILO clade 12.59Ma [6.18, 19.93] followed by an additional split between DK and ILO at 5.47Ma [1.86, 10.77]. The next *Mesida* clade contains species from Micronesia and *M. argentiopunctata* from Australia. Within the Micronesian complex, there was a species on DK that was sister to a species from Kosrae as well as a species on ILO sister to a species from Pohnpei. This split was relatively well-supported, with a posterior value of 0.88. This Sulawesi-Micronesian complex was estimated to have split from a clade that included *M. argentiopunctata* and the mid-high elevation Sulawesi *Mesida* 13.13Ma [8.945, 18.14] and was well supported, with a posterior value of 0.97.

The final Sulawesi *Mesida* spp. were estimated to have diverged from a clade including *M. argentiopunctata* and one DK species 8.70Ma [5.93, 12.14] with a posterior value of 0.89. The remaining taxa were largely split between DK and TPP and diverged from one another 7.71Ma [5.48, 10.43] although posterior values were very low for the higher relationships. The DK clade had one clade above 2,000m and another clade between 900-1700m that diverged 6.39Ma but

with low support. There were estimated to be two species at the highest elevation sites that were estimated to have diverged 3.32Ma [1.33, 6.14]. In the sister clade, there were distinct groups between elevations, with taxa at 900-1400m diverging from taxa at 1700m elevations about 5.28Ma [3.29, 7.78]. The TPP clade shows similar trends, with clades largely structured by elevation. There was one mid-elevation DK species sister to a mid-elevation TPP species with a posterior value of 0.84 and a divergence estimate at 1.98Ma [0.96, 3.43]. Most other relationships within the TPP clade had posterior values below 0.5. However, we did find strong support for the presence of three clades above 2000m (posterior value >= 0.99).


Posterior values:
> 0.95
> 0.75
<= 0.75
Elevation:
0 - 1000m
> 1000m
> 2000m
Mesida
Leucauge
Tylorida
DK
TPP
ILO
TPP/DK
M. argentiopuncata
Mesida sp. (Kosrae)
Mesida sp. (Pohnpai)
L. celebesiana
L. decorata
L. blanda
L. argyra
T. striata
Tylorida sp.

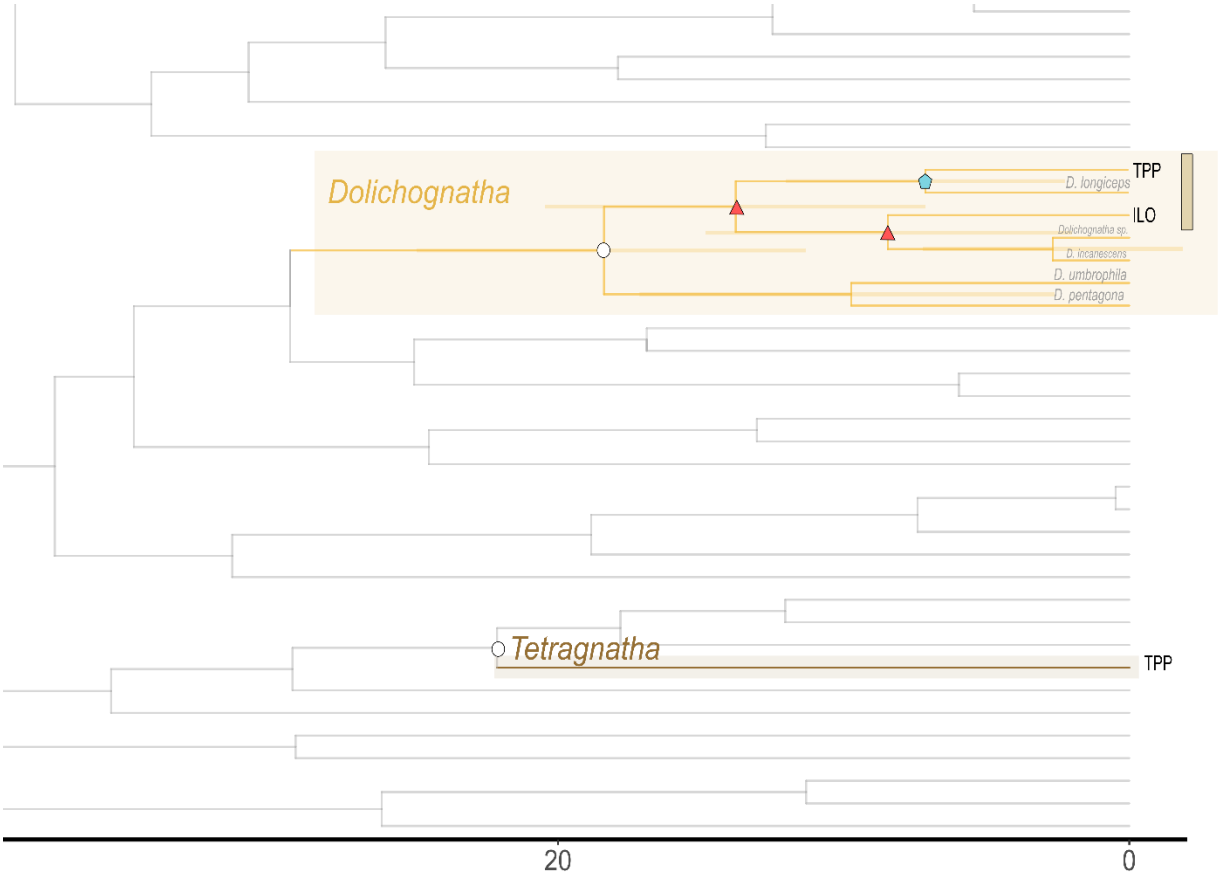


Figure 8. Simplified dated phylogeny. Branches represent putative species, estimated using GMYC. Grey branches were genera not found on Sulawesi, taken from GenBank. Node shapes represent posterior values. Color blocks represent genera. Tips are labeled with mountain abbreviation (TPP: Torompupu, DK: Dako, ILO: Ilomata). A tree containing the root node is in the supplementary materials (Supplementary Materials, Figure 2).

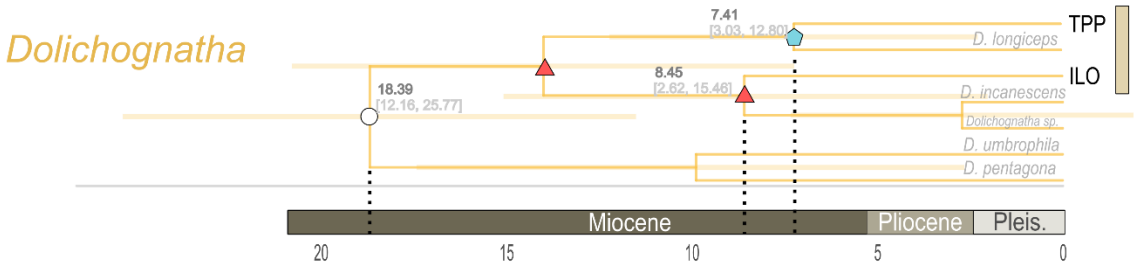


Figure 9. Genus-level phylogeny of *Dolichognatha* spp.

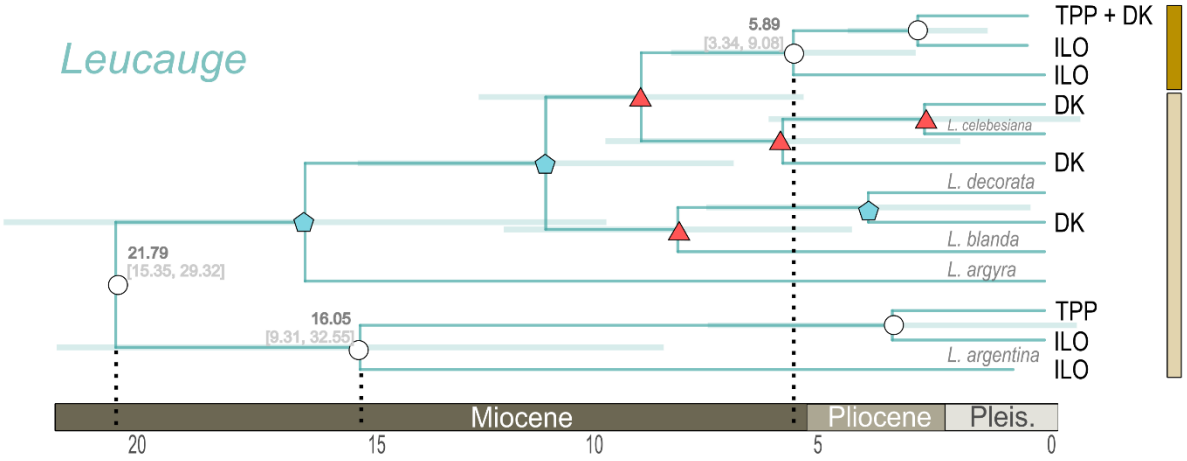


Figure 10. Genus-level phylogeny of *Leucauge* spp.

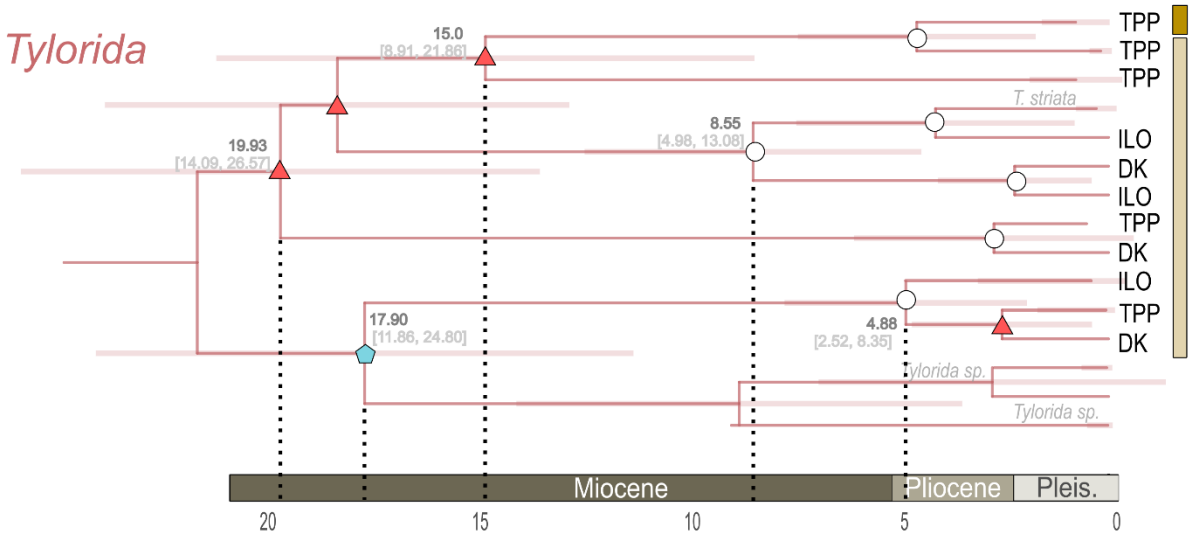


Figure 11. Genus-level phylogeny of *Tylorida* spp.

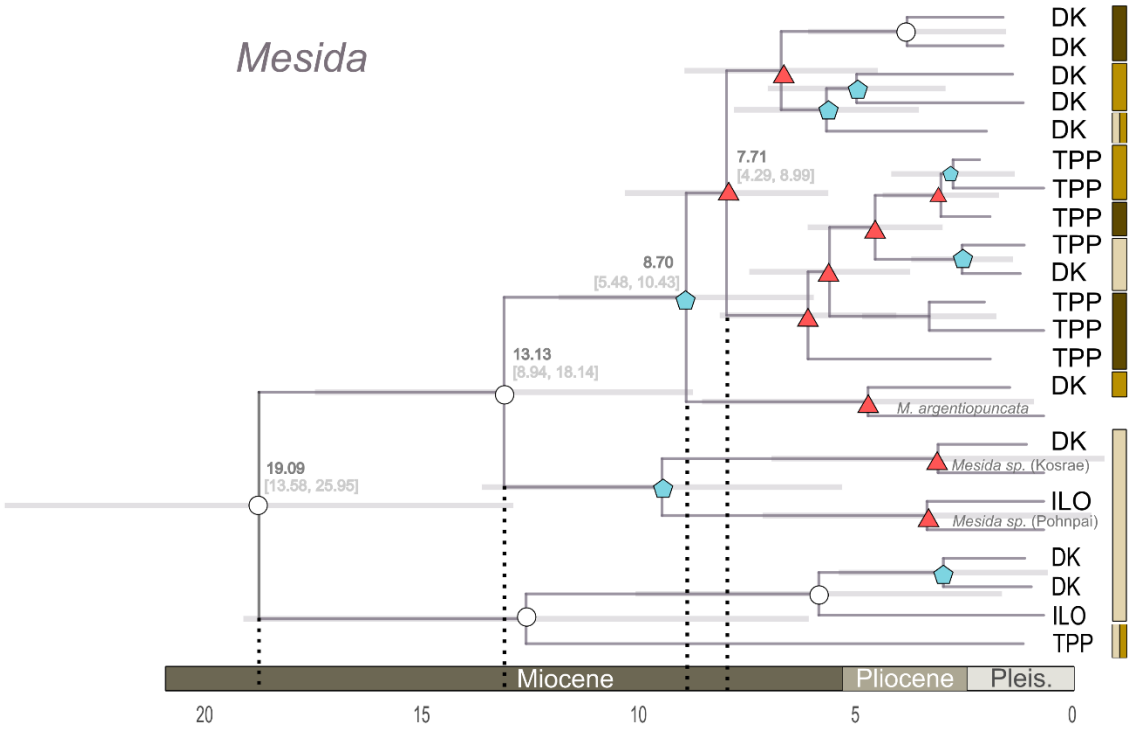


Figure 12. Genus-level phylogeny of *Mesida* spp.

2.4. Discussion

Spiders in the family Tetragnathidae on Sulawesi were found to be incredibly diverse, including 42 putative species belonging to five genera. This included: *Mesida, Tylorida, Leucauge, Dolichognatha* and *Tetragnatha. Mesida* and *Tylorida* were the most diverse genera, with more species estimated from the three mountains of Sulawesi than are currently known globally. Despite their more widespread occurrence and diversity, *Leucauge*, *Dolichognatha* and *Tetragnatha* were much less common.

2.4.1. Lowland tetragnathid diversity

Deep divergences

In lowland *Tylorida*, *Mesida* and *Leucauge,* we see deep divergences aligning with the formation of Sulawesi. This supports our hypothesis that the lowland taxa would represent older lineages that colonized Sulawesi when land emerged and maintained their lowland niche. In *Tylorida* and *Mesida*, we found estimated divergences aligning closely with the origins of Sulawesi, with *Tylorida* colonization estimates at 19.93Ma and *Mesida* colonization estimates at 19.09Ma (Figure 9 & 10). Two species of *Leucauge* were lowland taxa and appear to have colonized early in the history of Sulawesi as well (16.06Ma), each found on different mountains (TPP and ILO). Three lowland clades were found in *Tylorida*, with divergence dates at 15.0Ma, 19.93Ma, and 21.93Ma respectively. The split at 21.93Ma divides the clade which contains the widespread *T. striata* from a clade that contains unknown Micronesian species. This could indicate two subgenera that colonized Sulawesi independently. Alternatively, Sulawesi could have served as a

species pump. This has been noted in the incredibly speciose *Trigonopterus* weevil genus, in which Sulawesi served as a "colonization hub" for the rest of the Indomalayan region to the west[73]. While we see one older clade of *Leucauge*, the other species belonging to this genus seem to have arrived more recently, potentially establishing as more land area opened 4Ma.

Taxa diversifying between mountains

In-situ diversification appears to have occurred between paleo-islands in *Tylorida* and *Mesida* as well as in the single lowland *Leucauge* clade. In *Tylorida*, we documented extensive lowland diversity. Most clades show a species on each mountain, with divergence dates aligning roughly with the creation of new islands throughout the region. This aligns with our hypothesis that lowland taxa would disperse between paleo-islands and diversify. There is one clade on Gunung Torompupu, however, with three species - one lowland species that diverged from the two low and mid-elevation species 15.0Ma. Here, we could be detecting niche partitioning; the two lowland species may have adopted different ecological strategies and diverged early on. When high elevation habitat opened, the species with the more suitable strategy moved up the mountain and diverged; the date of divergences align with the emergence of higher elevation habitat and may provide support for this hypothesis.

In *Mesida*, there are splits between all mountains in the lowland clades, but the dates are less clear. The TPP and DK-ILO clade diverged around 12.6Ma, well before the formation of islands in the northern region. The species on TPP were found to occupy some higher elevation habitat than the species on DK or ILO. The emergence of mid-elevation habitat above 1,000m could have begun a process of ecological diversification similar to what we noted in *Tylorida*. Alternatively, the other land fragments that existed during the archipelago period could have served as the source population for DK and ILO. Regardless, we find clear evidence of lowland taxa dispersing and diversifying between land fragments as they emerged throughout the area.

Priority effects

Despite their prevalence across Indonesia, we found low diversity of *Tetragnatha* and *Leucauge* species. The early colonization by *Tylorida* and *Mesida* may have prevented other species from establishing without specialized adaptations. Most of the lowland *Leucauge* spp. we found were closely related to well-known and widely distributed species in Southeast Asia (*L. celebesiana*, *L. decorata*, and *L. argentina*) with much more recent divergences. In this case, the increasing land area on Sulawesi within the past 4 million years could have allowed establishment of more species in the lowlands than during the island's archipelagic period.

Dolichognatha could present an example of a lowland species that was able to establish alongside *Mesida* and *Tylorida* due to a different ecological strategy. *Dolichognatha* webs are easily distinguishable from other orb weavers because they build webs parallel to the ground, usually in the nook of a tree. This is a more specialized strategy with associated unique behaviors, such as hanging prey items from silken lines at the bottom of the webs. In addition to finding *Dolichognatha* spiders that employ the classic web ecological strategy (Figure 5), we found a *Dolichognatha* species that has potentially convergently evolved a similar trap-jaw predation mechanism to family Arachaeidae, although this is hypothesized solely on morphological

similarities between the two. Regardless, both species of *Dolichognatha* from Sulawesi appear to be adopting functional strategies that differ from the strategies of *Mesida* and *Tylorida*, genera that rely on more typical orb webs; this may have allowed establishment of *Dolichognatha* within lowland communities, albeit in lower abundances, while other taxa that depend on the more classic orb web, such as those in *Leucauge* and *Tetragnatha*, were not able to establish.

2.4.2. Mid- to high-elevation Tetragnathid diversity

More recent colonization associated with mountain building

The opening of higher elevation seemed to prompt a second colonization of *Mesida* with a divergence of 8.7Ma between Sulawesi species and *M. argentiopunctata.* This species is endemic to Australia and found in coastal habitat. High elevation habitat, with more exposure to wind, rain and sunlight, can require more specialized web architectures as well as physiological adaptations. The ancestor of these taxa may have employed web architectural strategies suitable for more wind-swept, extreme environments, as one might find in both coastal regions and in more exposed mountain tops. This supports our hypothesis that high elevation taxa will show more recent divergences from species outside of Sulawesi with some level of pre-adaptaion, although not seemingly to high-elevation specifically. However, the date of divergence in Sulawesi *Mesida* spp. was estimated to be before the emergence of elevations about 2,000m; instead, this alternate ecological strategy could have been more suitable for mid-elevations and provided more lability or adequate functionality in higher elevations that allowed *Mesida* in this clade to move up the mountain.

The creation of mid-elevation habitat may have additionally enabled the colonization and establishment of *Leucauge* spp. There is a *Leucauge* clade with an estimated divergence of 9.5Ma that we collected only in mid-elevation habitat (1,000-1,400m). Similarly, the single species of *Tetragnatha* we found was collected at ~1800m. As noted in the previous section, priority effects may have limited both genera's capabilities of establishing in the lowlands. The emergence of stable mid-elevation habitat above 1,000m may have provided opportunity for establishment due to open niche space. We found only a single species of *Tylorida* and one species of *Mesida* from the basal clade in this mid-elevation habitat, which may also support this idea – the taxa that colonized the lowlands largely stayed low, and the mid-high elevation niche space were filled by outside colonizations.

Connectivity across mountains

While we did see evidence for a secondary colonization associated with the emergence of mid-elevation habitat, we did not find taxa diverging around the time of high elevation mountain emergence, in the last 3-4Ma. Additionally, we do not see sister taxa across mountain tops. Instead we identified extensive within-mountain diversification in *Mesida*. There was a well-defined clade of mid-high elevation on DK and another clade on TPP, with an estimated divergence of 7.71Ma. This date is later than the emergence of land that would become DK, and could be a result of undersampling. However, the current pattern could indicate that a mid-elevation species of *Mesida* established, dispersed between other mid-elevation habitats, and then diversified upwards on each mountain as elevation increased. This causes us to reject our hypothesis that the most well-adapted

taxa in each elevation will disperse to fill the same niche, resulting in sister species across mountains.

A notable finding that may support the idea of an environmental filter preventing lowland taxa from entering high-elevation habitat was the presence of sympatric sister species at higher elevations. Unlike in low-land *Tylorida,* where multiple species were found at similar elevations but were seemingly from different subgenera or clades, co-occurring species of *Mesida* above 1,500m were closely related. At each elevation, we found a species that was collected primarily from webs in addition to a sister species that was not found in webs, and was less abundant. While we have no ecological evidence for this sister species beyond their absence in webs, we may be detecting different functional strategies being adopted by co-occurring *Mesida*; similar to the spiny-leg clade of Hawaiian *Tetragnatha* that lost web-building, these sister species may have undergone ecological diversification and filled unoccupied niche space that was not filled by other taxa because of the strong environmental filter imposed by high elevation climates. This again goes against our hypothesis that taxa would colonize from outside of Sulawesi to fill high-elevation niche space, and may indicate the degree of isolation on mountain tops combined with the strength of the environmental filter drove diversification within habitats.

2.5. Conclusion

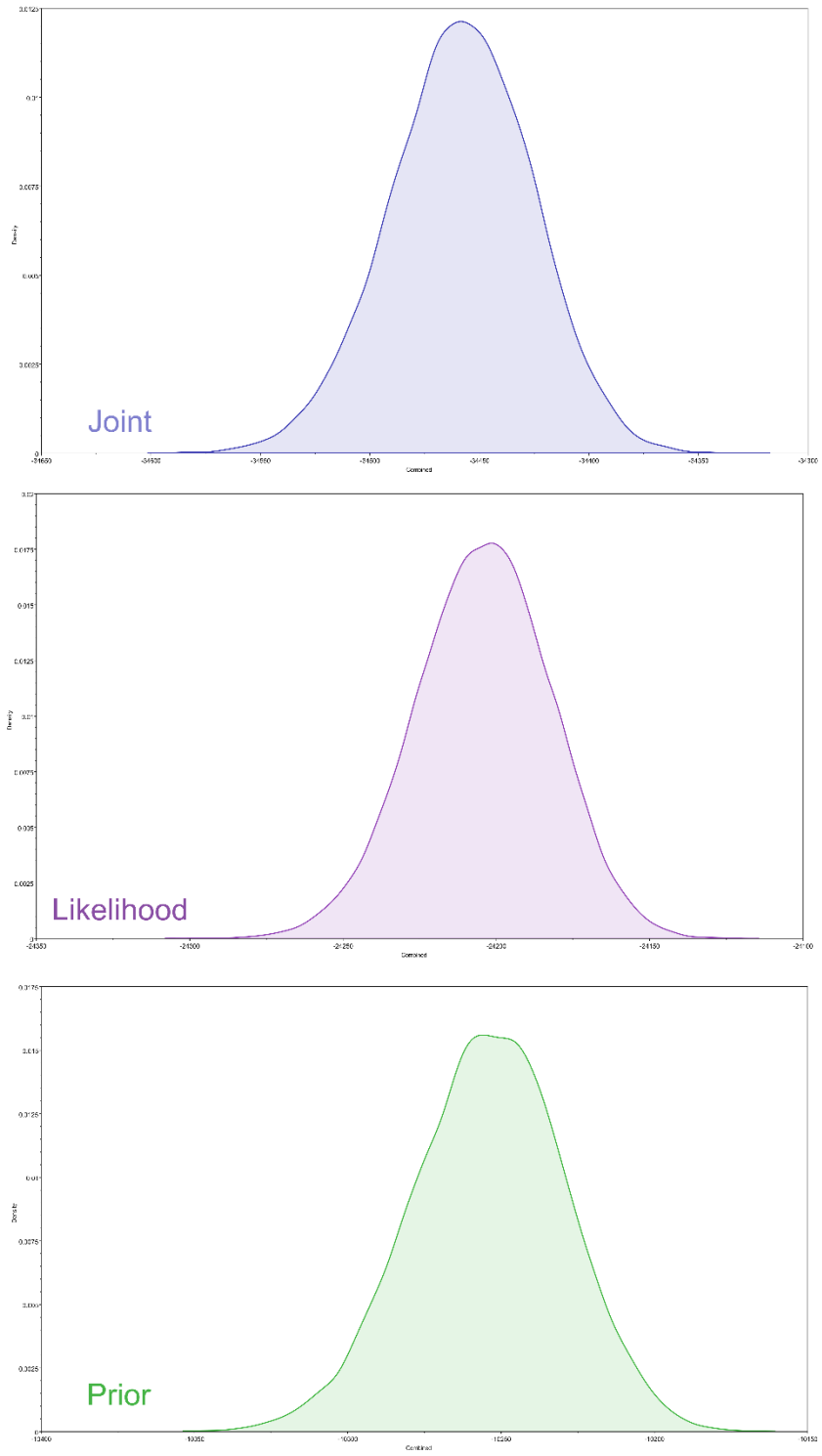
This research exemplifies the astonishing spider diversity left to be discovered on Sulawesi and the novel patterns of speciation that we can study using the complex paleogeology of the island paired with such an ecologically diverse taxonomic group. We found support for three of our hypotheses – 1a) lowland taxa showed deep divergences and diversification across former paleoislands; 1b) priority effects may have excluded widely occurring species from the low elevation communities; 2a) taxa at higher elevations are younger taxa that arrived as higher elevation emerged. The isolated nature of Sulawesi likely meant that taxa would arrive infrequently and the species that established would be determined by dispersal abilities and random chance, resulting in priority effects. More distantly related taxa with different ecological strategies, such as *Mesida*, *Tylorida*, and *Dolichognatha,* could fill available niche space and diversify, then prevent establishment of later arrivals with similar ecological strategies. High elevation habitat, on the other hand, seemed to create a strong environmental filter that led *Mesida* spp. to dominate elevation above 1,500m. Rather than see connectivity across mountains, we see within-mountain diversification. The lack of other taxa being able to establish due to environmental constraints and isolation could have allowed niche partitioning in this high elevation clade of *Mesida*, resulting in co-occurring sister species. These results are exciting and important both for biodiversity discovery and moving towards untangling the processes involved in shaping the endemic species of Sulawesi.

2.6. Supplementary materials

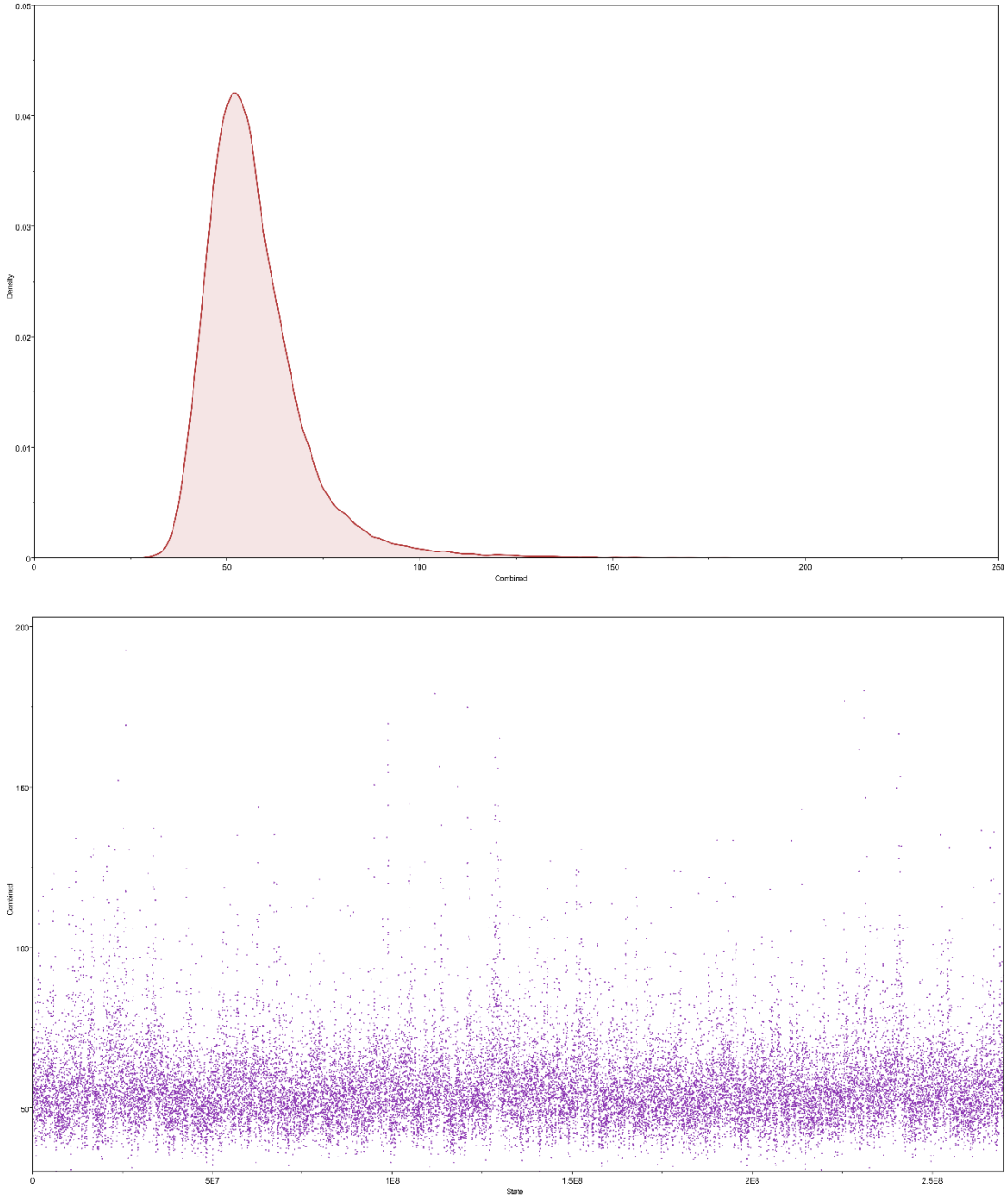
Tip Label	*16S*	*18S*	*28S*	*COI*	*H3*
Allende	KY016143.1	KY016734.1	KY017393.1	KY017950.1	KY018452.1
Antillognatha_lucida		GU129576.1	GU129603.1	GU129631.1	GU129647.1
Australomimetus	KY015963.1	KY016547.1	KY017192.1	KY017794.1	KY018304.1
Azilia_guatemalensis	EU003262.1	EU003371.1	EU003399.1	EU003280.1	EU003313.1
Azilia_histrio		MZ604144.1	MZ604206.1	MZ562628.1	MZ562842.1
Azilia_sp_834	GU129570.1	GU129581.1	GU129606.1	GU129624.1	GU129641.1
Azilia_sp_838		GU129582.1	GU129607.1	GU129625.1	GU129642.1
Chrysometa_alboguttata	KY016145.1	EU003389.1	EU003400.1		EU003314.1
Chrysometa_poas	MF663663.1	MF668716.1	MF668708.1	MF674516.1	MF681788.1
Chrysometa_zelotypa	MF663664.1	MF668717.1	MF668709.1	MF674517.1	MF681789.1
Cyrtognatha	KY016146.1	KY016735.1	KY017394.1	KY017951.1	KY018453.1
Diphya_sp_LT-2015	KT003115.1	KT002924.1	KT003022.1	KT002726.1	KT002825.1
Dolichognatha	KY016147.1	KY016736.1	KY017395.1	KY017952.1	
Dolichognatha_incanescens	MF663667.1	MF668718.1	MF668710.1	MF674518.1	MF681790.1
Dolichognatha_longiceps	MF663666.1	MF668719.1	MF668711.1	GU129632.1	GU129648.1
Dolichognatha_pentagona	MF663669.1	MF668720.1	MF668712.1	MF674519.1	MF681792.1
Dolichognatha_sp_FAPDNA061		EU003346.1	EU003405.1	EU003285.1	EU003317.1
Dolichognatha_umbrophila		MF668721.1	MF668713.1		MF681793.1
Harlanethis_lipscombae		MK841587.1		MK840828.1	
Leucauge_argentina		MK071501.1		MK057518.1	MK071608.1
Leucauge_argyra		KY016737.1	KY017396.1		KY018454.1
Leucauge_blanda	JN816495.1	JN816717.1	JN816926.1	JN817129.1	
Leucauge_celebesiana	JN816497.1	JN816719.1	JN816928.1	JN817131.1	
Leucauge_decorata	MK071493.1	MK071498.1	MK071540.1	MK057512.1	MK071606.1
MGrTetPohnSal27F				-	
MesKosFink31F				-	
MesiKosFin610m30F				-	
Mesida_argentiopunctata		MZ604197.1	MZ604256.1		MZ562833.1
Mesida_yini		KU836912.1		KU836880.1	
Mesida_yini_2		KU836913.1		KU836881.1	
Meta	KY016148.1	KY016739.1	KY017398.1	KY017954.1	
Metabus_ebanoverde	EU003265.1	EU003354.1	EU003414.1	EU003296.1	EU003326.1
Metellina	KY016149.1	KY016740.1	KY017399.1	KY017955.1	KY018456.1
Metleucauge_sp_897		GU129599.1	GU129621.1	GU129636.1	
Metleucauge_yunohamensis	JN816500.1	JN816722.1	JN816931.1	JN817134.1	
Mimetus	KY015966.1	KY016549.1	KY017195.1		
Mollemeta_edwardsi	EU003269.1	EU003374.1	EU003419.1	GU129634.1	EU003330.1
Nanometa_sp_LB-2014a	KM486317.1	KM486175.1	KM486387.1	KM486458.1	

Nanometa_sp_LB-2014b	KM486318.1	KM486176.1	KM486388.1	KM486459.1	
Nanometa_sp_LB-2014c	KM486319.1	KM486177.1	KM486389.1	KM486460.1	KM486505.1
NephilaAntipod				-	
NephilaClavata				-	
NephilaMaculata				-	
Okileucauge_hainan		MZ604199.1	MZ604259.1	MZ562630.1	MZ562835.1
OrsNa660m12F				-	
OrsNa750m2F				-	
OrsNan660m11F				-	
OrsNan730m6F				-	
OrsNana750m2F				-	
OrsiHNan750m3F				-	
Orsinome_sp_FAPDNA052	EU003267.1	EU003362.1	EU003424.1	EU003305.1	EU003337.1
Pinkfloydia_sp_LB-2014a	KM486328.1	KM486190.1	KM486401.1		KM486515.1
Taraire_oculta	MK841586.1	MK841588.1	MK841584.1		MK840830.1
TetNitensBrasSec18HReun				-	
Tetragnatha	KY016151.1	KY016743.1		KY017958.1	KY018459.1
Tylorida_sp		KY016744.1	KY017402.1		KY018460.1
Tylorida_striata		KY016745.1	KY017403.1		KY018461.1
Zhinu_manmiaoyangi	MF663662.1	MF668723.1	MF668715.1	MF674521.1	MF681794.1

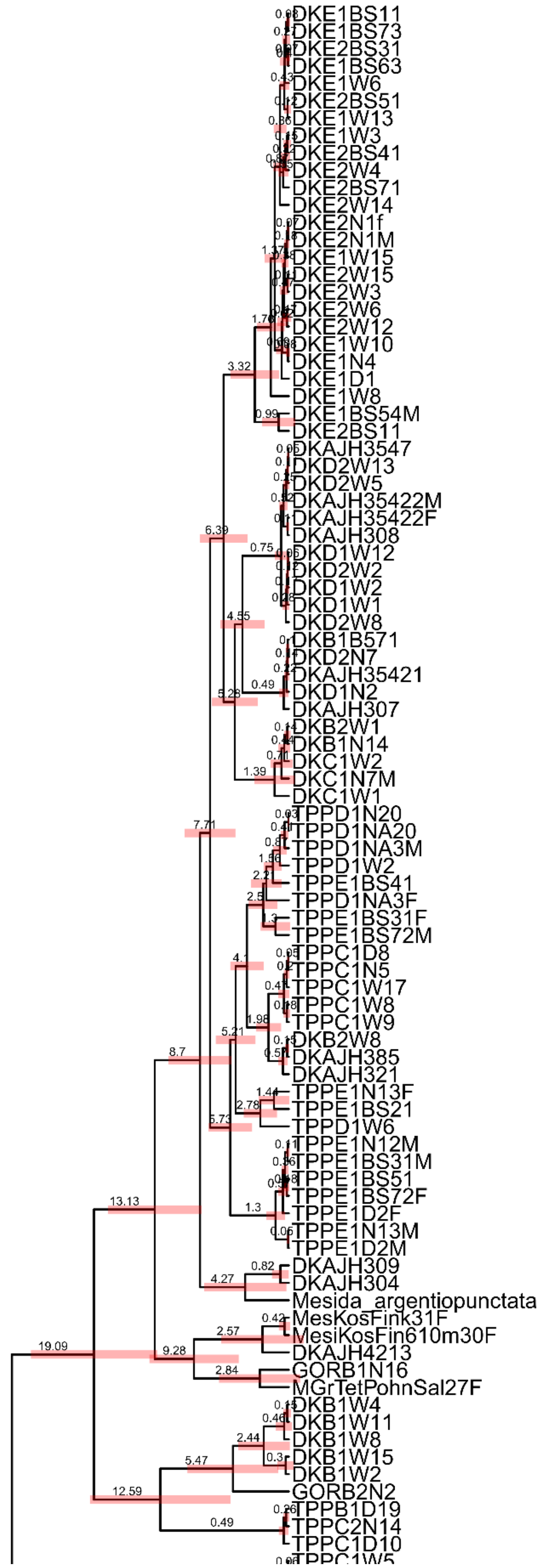
Supplementary Table 1. Accession numbers for the tetragnathids queried from GenBank. Blank cells represent no data. Dashed represent unpublished data (Gillespie, unpublished).

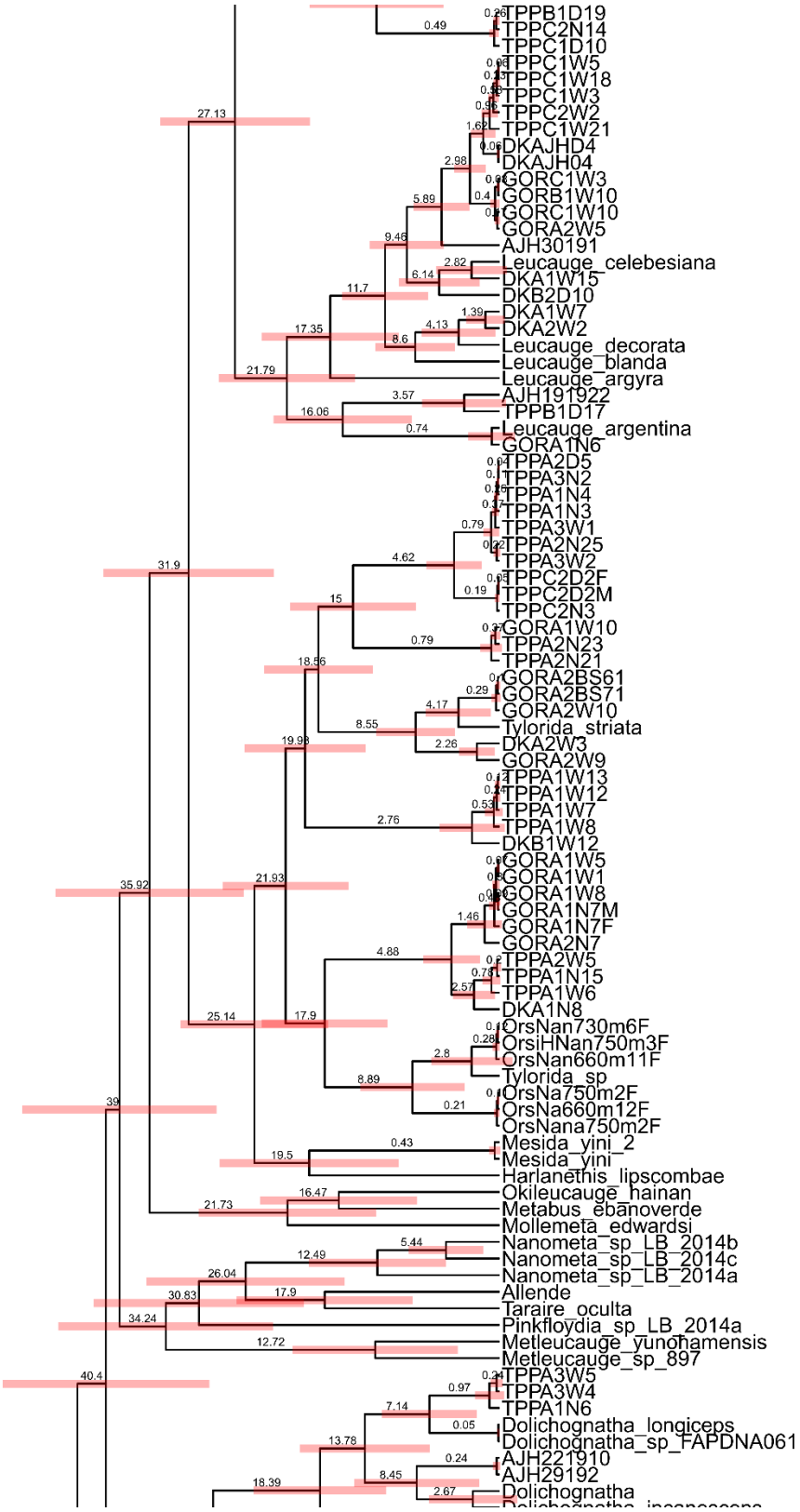


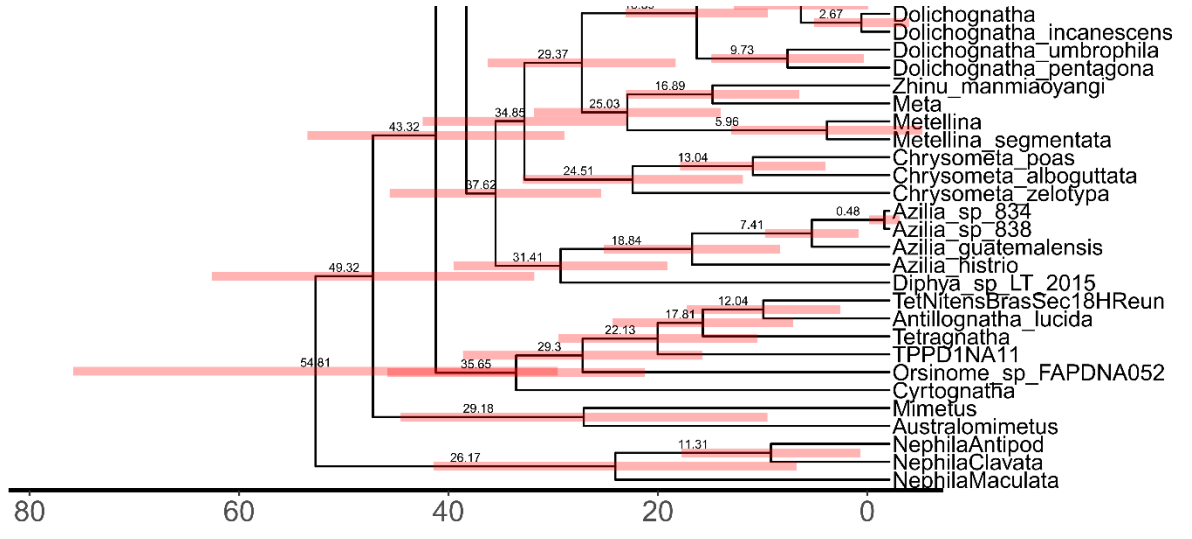
Supplementary Figure 1. Combined plots of the likelihood, joint, and prior distributions produced from running three independent chains using BEASTv1.10.4[95]. Plots produced using Tracer v1.7.2.



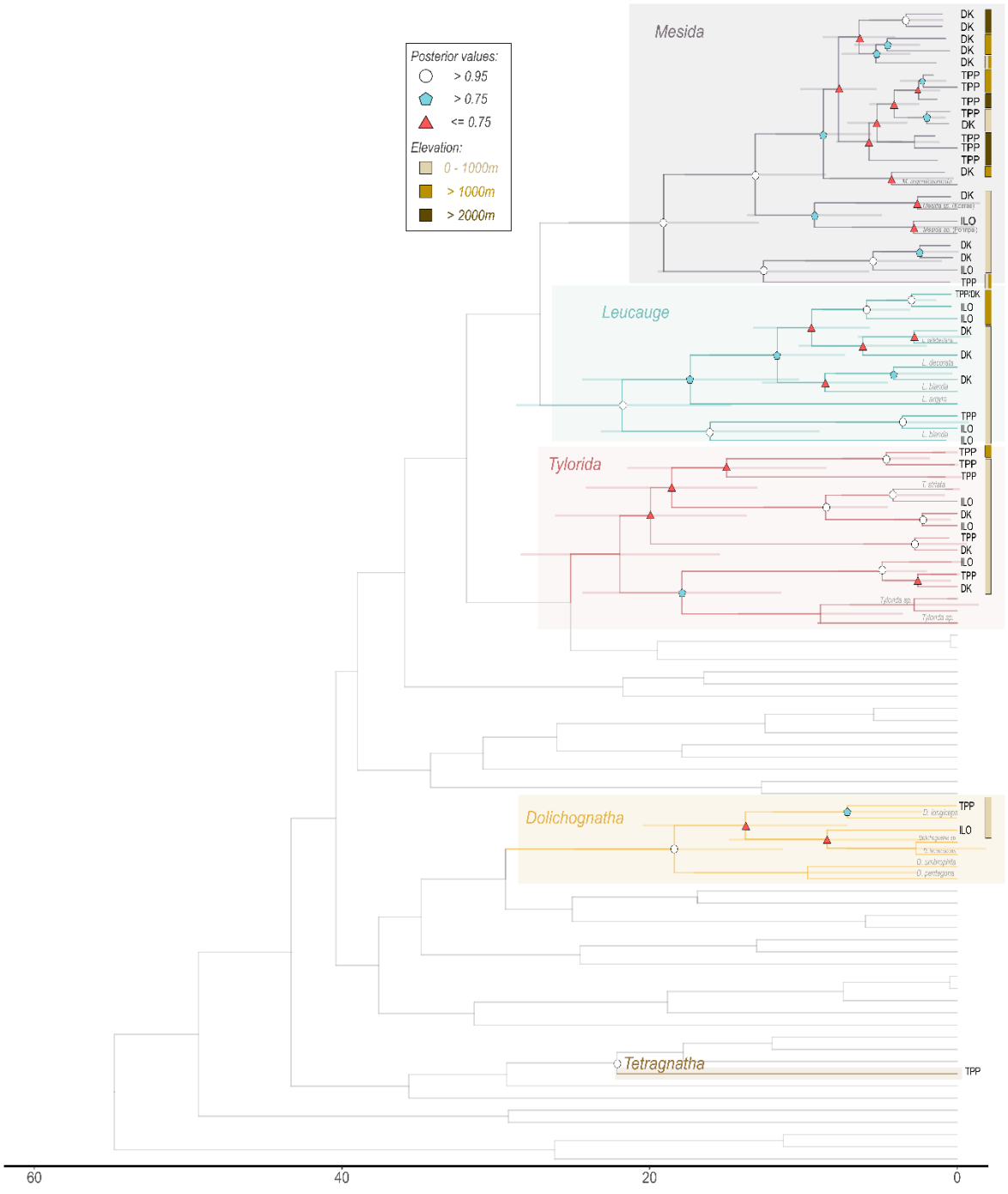
Supplementary Figure 2. Combined plots of the tree root height estimation, produced from three independent chains using BEASTv1.10.4[95], ESS = 959. Plots produced using Tracer v1.7.2.


DKE1BS11
DKE1BS73
DKE2BS31
DKE1BS63
DKE1W6
DKE2BS51
DKE1W13
DKE1W3
DKE2BS41
DKE2W4
DKE2BS71
DKE2W14
DKE2N1f
DKE2N1M
DKE1W15
DKE2W15
DKE2W3
DKE2W6
DKE2W12
DKE1W10
DKE1N4
DKE1D1
DKE1W8
DKE1BS54M
DKE2BS11
DKAJH3547
DKD2W13
DKD2W5
DKAJH35422M
DKAJH35422F
DKAJH308
DKD1W12
DKD2W2
DKD1W2
DKD1W1
DKD2W8
DKB1B571
DKD2N7
DKAJH35421
DKD1N2
DKAJH307
DKB2W1
DKB1N14
DKC1W2
DKC1N7M
DKC1W1
TPPD1N20
TPPD1NA20
TPPD1NA3M
TPPD1W2
TPPE1BS41
TPPD1NA3F
TPPE1BS31F
TPPE1BS72M
TPPC1D8
TPPC1N5
TPPC1W17
TPPC1W8
TPPC1W9
DKB2W8
DKAJH385
DKAJH321
TPPE1N13F
TPPE1BS21
TPPD1W6
TPPE1N12M
TPPE1BS31M
TPPE1BS51
TPPE1BS72F
TPPE1D2F
TPPE1N13M
TPPE1D2M
DKAJH309
DKAJH304
Mesida_argentiopunctata
MesKosFink31F
MesiKosFin610m30F
DKAJH4213
GORB1N16
MGrTetPohnSal27F
DKB1W4
DKB1W11
DKB1W8
DKB1W15
DKB1W2
GORB2N2
TPPB1D19
TPPC2N14
TPPC1D10
TPPC1W5
3.32
0.99
6.39
0.75
4.55
0.49
1.39
7.71
8.7
5.21
5.73
1.98
2.78
1.3
4.27
0.82
13.13
19.09
9.28
2.57
2.84
12.59
5.47
2.44
0.49


TPPB1D19
TPPC2N14
TPPC1D10
TPPC1W5
TPPC1W18
TPPC1W3
TPPC2W2
TPPC1W21
DKAJHD4
DKAJH04
GORC1W3
GORB1W10
GORC1W10
GORA2W5
AJH30191
Leucauge_celebesiana
DKA1W15
DKB2D10
DKA1W7
DKA2W2
Leucauge_decorata
Leucauge_blanda
Leucauge_argyra
AJH191922
TPPB1D17
Leucauge_argentina
GORA1N6
TPPA2D5
TPPA3N2
TPPA1N4
TPPA1N3
TPPA3W1
TPPA2N25
TPPA3W2
TPPC2D2F
TPPC2D2M
TPPC2N3
GORA1W10
TPPA2N23
TPPA2N21
GORA2BS61
GORA2BS71
GORA2W10
Tylorida_striata
DKA2W3
GORA2W9
TPPA1W13
TPPA1W12
TPPA1W7
TPPA1W8
DKB1W12
GORA1W5
GORA1W1
GORA1W8
GORA1N7M
GORA1N7F
GORA2N7
TPPA2W5
TPPA1N15
TPPA1W6
DKA1N8
OrsNan730m6F
OrsiHNan750m3F
OrsNan660m11F
Tylorida_sp
OrsNa750m2F
OrsNa660m12F
OrsNana750m2F
Mesida_yini_2
Mesida_yini
Harlanethis_lipscombae
Okileucauge_hainan
Metabus_ebanoverde
Mollemeta_edwardsi
Nanometa_sp_LB_2014b
Nanometa_sp_LB_2014c
Nanometa_sp_LB_2014a
Allende
Taraire_oculta
Pinkfloydia_sp_LB_2014a
Metleucauge_yunohamensis
Metleucauge_sp_897
TPPA3W5
TPPA3W4
TPPA1N6
Dolichognatha_longiceps
Dolichognatha_sp_FAPDNA061
AJH221910
AJH29192
Dolichognatha



Supplementary Figure 3: Complete dated phylogeny with all terminals, 95% HPD intervals, and median heights



Supplementary Figure 4: Annotated phylogeny with root visible.

.

Understanding the assemblage and structure of biotic communities is crucial to predict ecosystem response and develop scientifically informed management strategies. In **CHAPTER 3**, I study the processes involved in structuring leaf litter arthropod communities across an elevation gradient on Gunung Dako. This included studying communities in native forest as well as studying communities in cacao agroforests

Chapter 3: The relative importance of stochastic and deterministic processes in the assemblage of leaf litter arthropod communities across an elevation gradient

3.1. Introduction

Many questions remain about the interplay of deterministic and neutral processes in community assembly. Deterministic or niche-based processes, in which abiotic and biotic variables shape species membership, have been a primary focus of community ecology[13,101]. The concept of an environmental filter or competitive exclusion through biotic interactions represents deterministic processes that have been historically considered the primary mechanisms in assembly[102,103]. Neutral or stochastic processes, such as randomness in the order of species arrival and ecological drift, are additionally important in shaping communities. Computational advances have allowed the creation of more complex models that estimate the importance of deterministic and neutral approaches in shaping communities[104–106]. Further exploration using empirical data is crucial to assess how the processes involved in assemblage relate to community stability and resilience, especially in the face of rapid global change[107]. For example, communities structured more by dispersal and stochasticity versus those structured strongly by interactions between abiotic and biotic factors may respond differently to disturbances and global change. Identifying the importance of neutral and niche-based processes in shaping community biodiversity is therefore necessary to support ecosystems in the future[108].

Degree of isolation and dispersal capabilities are key factors involved in early stages of community assembly. Regional source populations of potential colonizers dictate what species are likely to arrive in a location. Community assembly in less isolated regions with nearby species sources is thought to initially be driven by a combination of environmental and stochastic processes – whoever arrives first and is suitable for an open niche will establish[79,109]. Given enough time, stochastic assembly processes may give way to deterministic processes. This is an assemblage trend noted in other settings, such as in microbial succession[110] Open niche space will continue to be filled by lineages most well-adapted and other taxa will be unable to establish due to competitive exclusion[111]. This is often predicted to result in a pattern of phylogenetic overdispersion, where niches are first filled by distantly related lineages with distinct functional strategies.[112] The assembly process in more isolated locations that experience low rates of immigrating taxa due to dispersal limitation may be shaped more quickly by deterministic processes; the number and quality of empty niches may promote speciation in taxa that successfully established, with increasing diversity driving more diversification[113]. This results in species within a community being closely related, resulting in a pattern of phylogenetic under-dispersion[114,115]. Once all niche space is filled, communities may reach a state of equilibrium, where differences emerge through more stochastic processes because resource use has been optimized and all niche space is filled[116].

Islands have served as fundamental natural laboratories to study this process as they provide simplified systems ideal for studying the interplay of neutral and deterministic processes[4]. Classic ecological theories, such as MacArthur and Wilson's theory of island biogeography[117], emerged through the study of island systems. Island systems often have a definable time history, which allows studying community assembly temporally. For example, the archipelago of Hawaii provides a fascinating system to study how communities formed on isolated *de novo* islands, each with different ages[4]. The isolation of many islands is an additional benefit, where infrequent immigration is thought to make deterministic processes more important at the early stages of assembly and, therefore, we can better study evolutionary processes through the radiations that occur[118]. Lastly, many island systems additionally contain mountains which provide an opportunity to disentangle the processes of assembly related to the environment because of their high turnover in habitat within a small area[54,119]. Especially in tropical locations where climates remain largely stable, this turnover allows assessing the trajectory of assembly in the context of distinct environmental conditions.

The biogeographic regions of Sundaland, the Philippines and Wallacea contain an extraordinary number of topographically complex islands with different endemic diversity and paleogeographic histories[2]. Sulawesi, the largest island in Wallacea, is one island that is ideal for studying community assembly. The island formed 23Ma following collision of land fragments with biotic communities formed through long distance dispersal[69]. It existed as an isolated archipelago with lowland habitat for an extended amount of time until roughly 1-3Ma ago, rapid tectonic activity connected the archipelago, extended the land area and created high elevation mountains[7]. This system provides an exciting opportunity to assess the interplay of deterministic and neutral processes through time, with lowland habitat existing for 25Ma and highland habitat being relatively young. In addition to the environmental and temporal gradient provided by Sulawesi mountains, the lowest elevations are highly disturbed and can therefore also provide insight into the assemblage mechanisms in degraded habitat. We may expect a) deterministic processes, specifically environmental filtering, to be most important in the highest elevation communities given their young age and more extreme climates and b) neutral processes to be most important at low-mid elevation communities, given habitat stability through time that has allowed communities to reach near-equilibrium. Lastly, in the highly disturbed lowland sites, we may predict that c) communities have assembled through a combination of neutral and niche processes, with recent generalist colonizers establishing stochastically based on order of arrival, and native or endemic taxa being strongly shaped by the environmental filter of the degraded forest.

To test our hypotheses, we use metabarcoding of arthropods collected from soil across an elevation gradient ranging from 338 meters to 2,238 meters. We first examined the overall biodiversity structure using phylogenetic alpha and beta diversity metrics. Then, to test for neutral versus deterministic processes, we assessed phylogenetic over- and under-dispersion; phylogenetic under-dispersion is predicted in communities shaped by an environmental filter, phylogenetic over-dispersion is predicted in communities shaped by competitive exclusion, and a lack of phylogenetic signal is predicted in communities shaped stochastically. Lastly, to assess the importance of specific processes in assemblage, we used the iCAMP model[106] which

partitions phylogenetic signal into deterministic (homogeneous and heterogeneous selection) and neutral (dispersal limitation, dispersal homogenization and drift) processes. We found neutral processes were important in structuring all communities, but deterministic processes were only identified as important in the low and high elevation sites. Rather than finding evidence of underdispersion, we instead found phylogenetic overdispersion at two of the higher elevation sites while the most disturbed site showed under-dispersion, as anticipated given a strong environmental filter. The mid-elevation sites show low phylogenetic signal and support our hypothesis of neutral-based processes currently structuring these communities.

3.2. Methods

3.2.1. Sample collection

Leaf litter arthropod communities were collected in conjunction with the Sulawesi biotic inventory described in Chapter 1. We focused on the sampling of Gunung Dako for this project. At each site across the elevation gradient, a 2-gallon bag of pre-sifted leaf litter was collected. This was then sorted by hand to find all leaf litter invertebrates. Spiders were stored separately. All other invertebrates (Arthropoda, Annelida, Mollusca) were stored in 95% EtOH.

3.2.2. Molecular procedures

Samples were first sorted by body size then extracted using the approach in Henrik et al. 2017[120] Tissue was disrupted using 5mm steal beads and a Genogrinder (Spex SamplePrep, Matuchen, NJ, USA) for 2 min at 1,200 hz. Extractions were performed using the same protocol as in Chapter 1, based on the Qiagen Puregene kit. A longer fragment of COI was amplified using the primer pair mlCOIintF[121] and Fol-degen-rev[83], commonly used in arthropod metabarcoding[120,122]. We additionally used a newer primer set for arthropod metabarcoding (BF3, BR2) which was proposed as a remedy for primer slippage and increased taxonomic resolution[123] (Table 1). Three PCR replicates were used for each sample and indexed individually. Library construction otherwise was identical to that of Chapter 1.

Forward	Sequence 5′-3′	Reverse	Sequence 5′-3′
mlCOIintF[121]	GGWACWGGWTGAACWGTWTAYCCYCC	Fol-degen-rev[83]	TANACYTCNGGRTGNCCRAARAAYCA
BF3[123]	CCHGAYATRGCHTTYCCHCG	BR2[123]	GCHCCHGAYATRGCHTTYCC

Table 1. Primer pairs used to amplify COI.

3.2.3. Bioinformatics

All sequence processing described in Chapter 1 was used on the produced reads. Amplicon sequence variants (ASVs) are used throughout this chapter. While amplification and sequencing produced a high number of reads for the BF3/BR2 primer set, a large portion were low quality and removed during processing with DADA2. This led us to question the quality of the retained reads and we correspondingly chose to exclusively use the results from the mlCOIintF/Fol-degen-rev primer pair.

ASVs were aligned and visualized in Geneious Prime v.2022.0.2 to identify any reading frame interruptions or stop codons, which can indicate NuMts[124]. These sequences were removed. Sequences were then assigned taxonomic identities using megablast, ran in Geneious. The top 10 hits were extracted and consensus identities were generated based on percent identity matches above 80%. We used conservative percent identity thresholds for taxonomic identities to ensure no incorrect assignments were made. Species-level IDs were kept if a sequence was above 99%, genus above 97%, family above 92% and order above 85%. Matches below 80% were removed. Sequences belonging to Arthropoda were retained. Results were then filtered by PCR replicate; ASVs were required to be detected in two of the three replicates to be retained in the dataset. Lastly, ASVs represented with less than 10 reads in a sample were removed and the number of reads for an ASV within a sample needed to be at least 0.1% of total reads for the ASV.

Figure 1. Lowland agroforests of Gunung Dako. Sites consisted primarily of cacoa trees but there was other plants and groundcover throughout.

To assess for any remaining problematic sequences, we constructed an exploratory phylogeny in Geneious using a neighbor-joining approach. ASVs with abnormal branch lengths were removed. A maximum-likelihood phylogeny was then estimated with 1,000 bootstrap replicates and 1,000 alrt tests using IQTREE v1.6.12. ModelFinderPlus[48] was used to find the best substitution model. The phylogeny was then made ultrametric using function *chronos* in R

package ape[44], with lambda = 1 and 10 million iterations. Both trees were used in analyses of phylogenetic diversity (Figure 2).

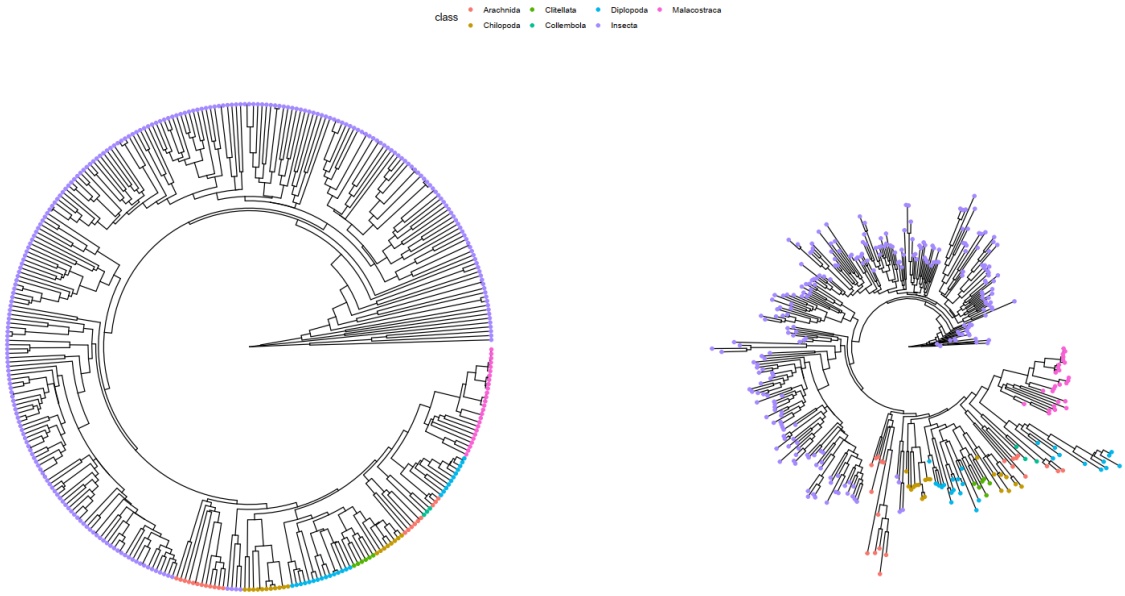


Figure 2. Phylogeny constructed using maximum likelihood. The a) ultrametric tree was made by applying function chronos to b) the non-ultrametric tree.

3.2.4. Statistical analysis

Density plots were constructed to visualize BLAST percent identity matches across elevation as a proxy for amount of unknown taxa. Differences between percent identity matches across elevation were tested using an ANOVA, along with a Tukey multiple comparisons of means test to identify the significant mean differences between elevation groups.

The ASV matrix was converted to an incidence matrix to be used in analyses, as sequencing reads are not directly correlated with abundances. Faith's phylogenetic diversity (PD) was calculated using the R package BAT; values were calculated using both the ultrametric and non-ultrametric trees. We tested for differences in PD values when using either tree with the nonparametric paired Wilcoxon signed-rank test. We used multiple linear regression to test for association been PD and environmental variables. Environmental variables were downloaded from WorldClim 2.1[125] along with fine-scale forest cover data from 2010[126]. We constructed a correlation matrix and tested for significance between environmental variables (Supplementary Materials, Figure 1). Based on the results, we selected elevation, annual precipitation, monthly diurnal range (mean of the temperature range each month), and forest cover for our environmental variables.

Phylogenetic beta diversity was calculated using the function *beta* from the package BAT for both trees with the Jaccard dissimilarity index and incidence matrix. This function produces total phylogenetic dissimilarity, as well as the components of dissimilarity, specifically richness and replacement differences. We tested for differences in phylogenetic beta diversity values when using either tree again with the nonparametric paired Wilcoxon signed-rank test. We used linear

regression to test the relationship between spatial distance and phylogenetic distance between communities. We performed a partial Mantel test using the function *mantel.partial* in vegan to assess phylogenetic dissimilarities and environmental distance while controlling for spatial distance. The input environmental distance matrix was calculated using the *vegdist* function in the R package vegan and Euclidean distance using the uncorrelated environmental variables discussed above; elevation was removed because it is directly tied to spatial distance.

Because we were particularly interested in the processes of assembly related to the formation of Sulawesi, and therefore processes temporally grounded, we used the ultrametric tree for tests of phylogenetic over- and under-dispersion. We calculated the mean pairwise distance between communities and the standardized effect size using *ses.mpd* in picante. This is equivalent to the inverse of the nearest relative index (NRI). We used the "independent swap" null model, which uses a swapping algorithm to randomize the community[127,128] . We used ANOVA to test differences in observed effect size by elevation group and the Kruskal-Wallis test to evaluate differences by disturbance (disturbed or undisturbed). We then used multiple linear regression to test the relationship between the effect size and all environmental variables.

To determine the importance of deterministic versus neutral processes, we used the iCAMP model (community assembly mechanisms by phylogenetic-bin-based null model analysis[106]). This model uses a phylogenetic binning method to classify individual taxa and determines the processes structuring each bin using the net relatedness index (NRI) and the Raup-Crick metric (RC). If the absolute value of NRI is above 1.96, this is labeled as deterministic processes shaping the phylogenetic bin; positive versus negative values are used to estimate if selection is homogenous or heterogeneous. Comparisons below 1.96 are assessed using both RC and NRI to define the importance of homogenizing dispersal, dispersal limitation and drift. Then, the fraction of processes shaping the phylogenetic bins found within a community are calculated and weighted by relative abundance. The resulting values represent the importance of various factors shaping the whole community, while taking into account lineage-specific differences in the processes dictating their community membership.

Because relative abundance is a key component in the iCAMP model, we used a matrix with raw reads as the input, applied the Hellinger-transformation, and then used Euclidean distance. For the phylogenetic and the taxonomic null models, we used both randomization within phylogenetic bins and between phylogenetic bins using the "across.all" argument. For the phylogenetic distance threshold used to cluster bins, we used 0.1 and a minimum bin size of 12 ASVs. Using bootstrapping analysis as implemented in package iCAMP, we calculated the variation within groups and compared assemblage processes driving differences between groups based on elevation (low, mid and high) using 10,000 bootstrap replicates.

3.3. Results

Following all filtering, 685,159 reads were retained, grouped into 319 OTUs and 349 ASVs. Most sequences had extremely low BLAST matches. Percent identity match varied significantly by elevation group (F-value = 4.512, Pr = 0.00139) with the sites under 500m having the highest mean percent identity match of 88.9% (Figure 3). Only 12 ASVs had above a 97% match, ten of

which were found in low elevation sites. The other two ASVs were found at the highest elevations but were identified only as “Hymenoptera sp”. There were five confident species identifications (> 99%), all from the sites below 500 meters: *Nesogaster aculeatus, Pheidole umbonate, Lobopterella dimidiatipes, Monomorium bifidoclypeatum* and *Monomorium floricola*. *Monomorium floricola* is a widespread invasive ant species. The other species, *M. bifidoclypeatum,* is only known from Madagascar. *P. umbonate* is another ant species, known from Papua and other Pacific islands. *N. aculeatus* is an earwig species known from Papua. *L. dimidiatipes* is a cockroach species known from Japan and Christmas island. The sum of all branch edges was 62.333 for the non-ultrametric tree and 63.483 for the ultrametric tree.

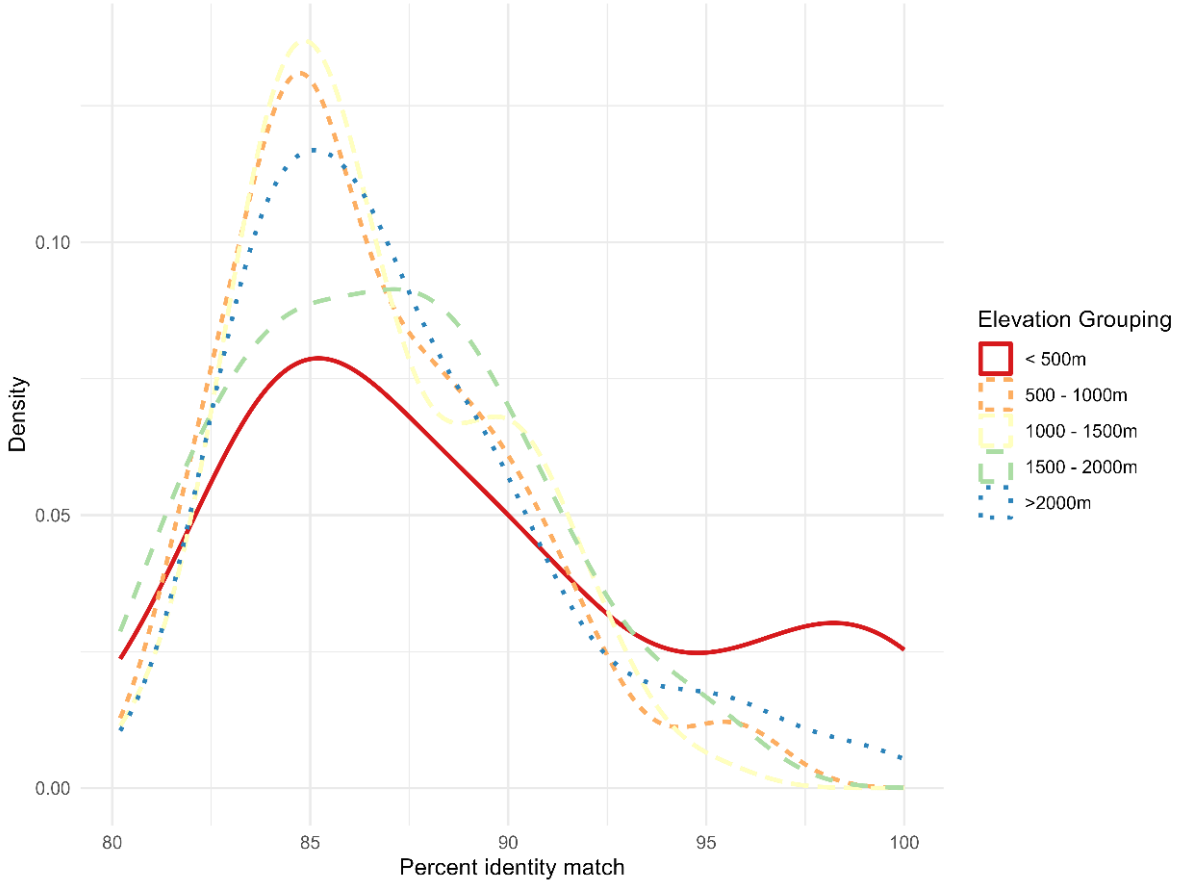


Figure 3. Density curves showing the number of ASVs with a certain percent identity match. The density peaks around 85%. The highest identity matches were found in the low elevation sites.

3.3.1. Phylogenetic diversity

Phylogenetic alpha diversity was highest at mid elevation sites (1000-1500m) and lowest below 500m. There was no significant relationship between PD and any of the environmental variables (Figure 4). Phylogenetic diversity calculated using a non-ultrametric tree versus the ultrametric tree varied slightly; this was most pronounced against elevation and annual precipitation as predictors, with values becoming more similar towards the highest elevation site, which were also the sites with the highest precipitation. However, PD values between calculated from either tree were not significant (V = 8, p-value = 0.195).

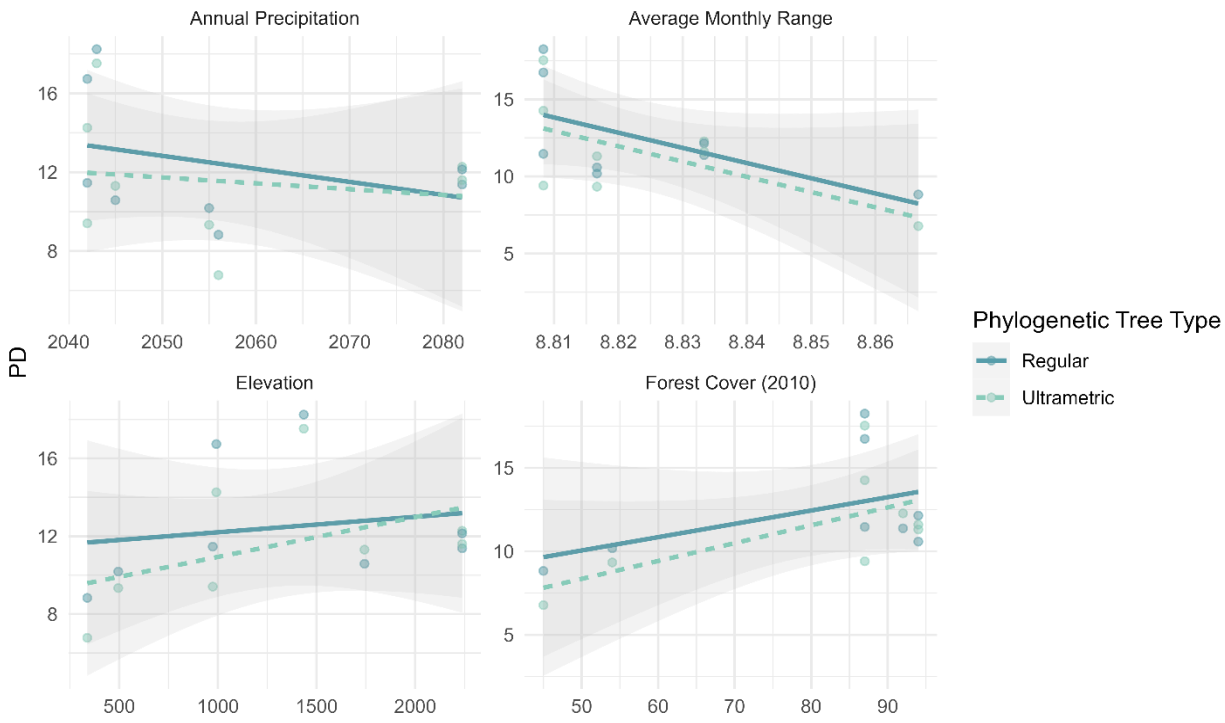


Figure 4. Variation in Faith's PD by four environmental predictors: a) annual precipitation, b) elevation, c) forest cover in the year 2010 and d) average monthly diurnal range. There were no significant linear relationships between PD and any environmental variables.

Communities in the same elevation band were generally most similar to one another, although the middle elevation sites showed higher variability (Figure 5). There was a significant relationship between spatial distance and phylogenetic dissimilarity; the farther apart the sites, the more dissimilar they became. This was strongest using the non-ultrametric tree (R-squared = 0.384, p-value = 3.45e-06) but was also significant using the ultrametric tree (R-squared = 0.272, p-value = 0.0001)(Figure 6). Replacement was the primary component of beta diversity values. When using the non-ultrametric tree, replacement driven dissimilarity increased as sites became more distant from one another (R-squared = 0.146, p-value = 0.0254). This trend was not observed when using the ultrametric tree; there was instead an increase in richness-related changes with distance, although a very weak relationship (R-squared = 0.046, p-value = 0.142)(Figure 7). However, for the total beta diversity, there were not strong differences when calculated using either tree (V = 3813, p-value = 0.088). Using a partial Mantel test to control for spatial distance, there was no significant relationship between the environmental variables and phylogenetic distances (p-value = 0.294, 0.144).

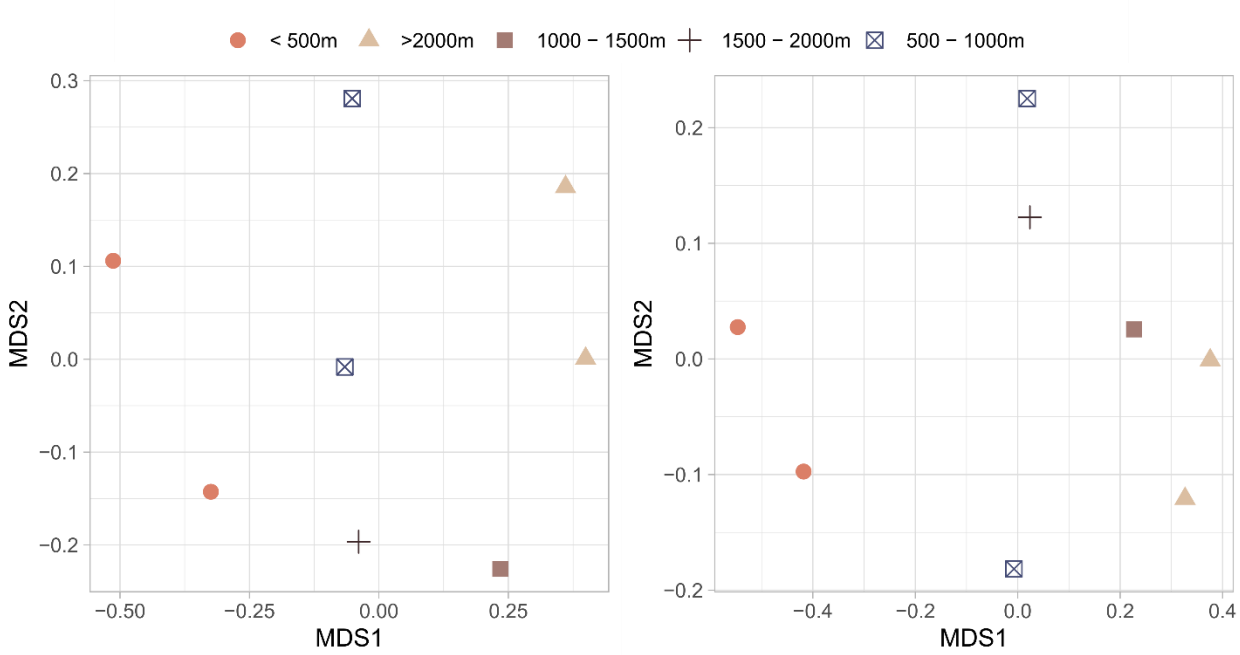


Figure 5. Ordination plot from non-metric multidimensional scaling. Each point represents a site, with shape and color associated with elevation group. Plot a) is calculated using the ultrametric tree while plot b) is calculated using the non-ultrametric tree. In both cases, the highest and lowest elevation sites cluster together. The mid-elevation sites show more spread.

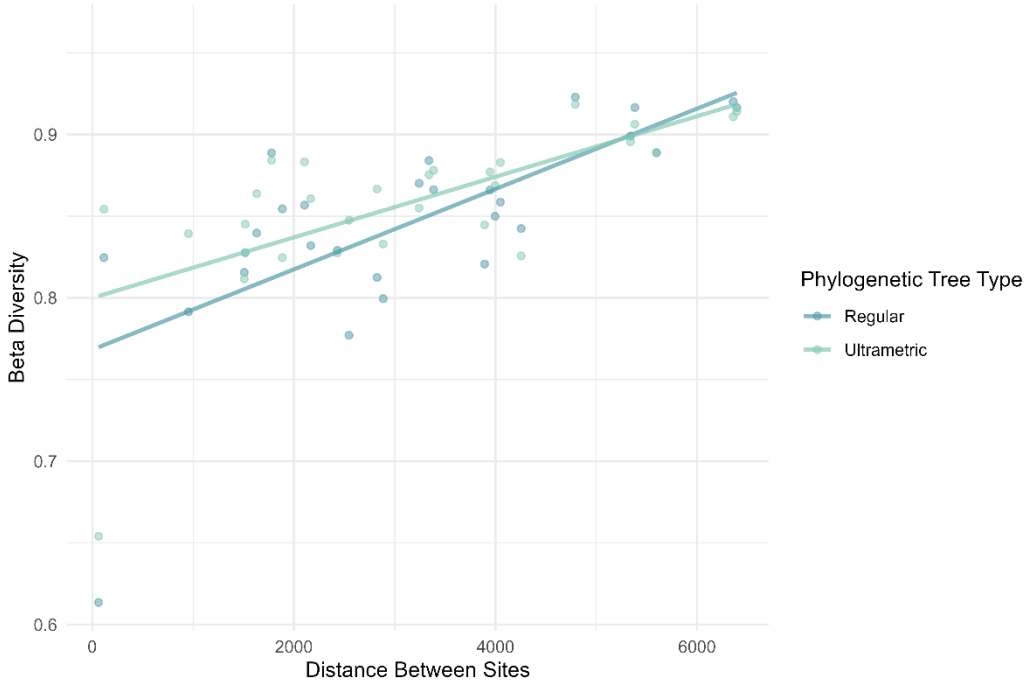


Figure 6. Phylogenetic beta diversity against spatial distance between sites. There was a significant positive relationship between spatial and phylogenetic distance.

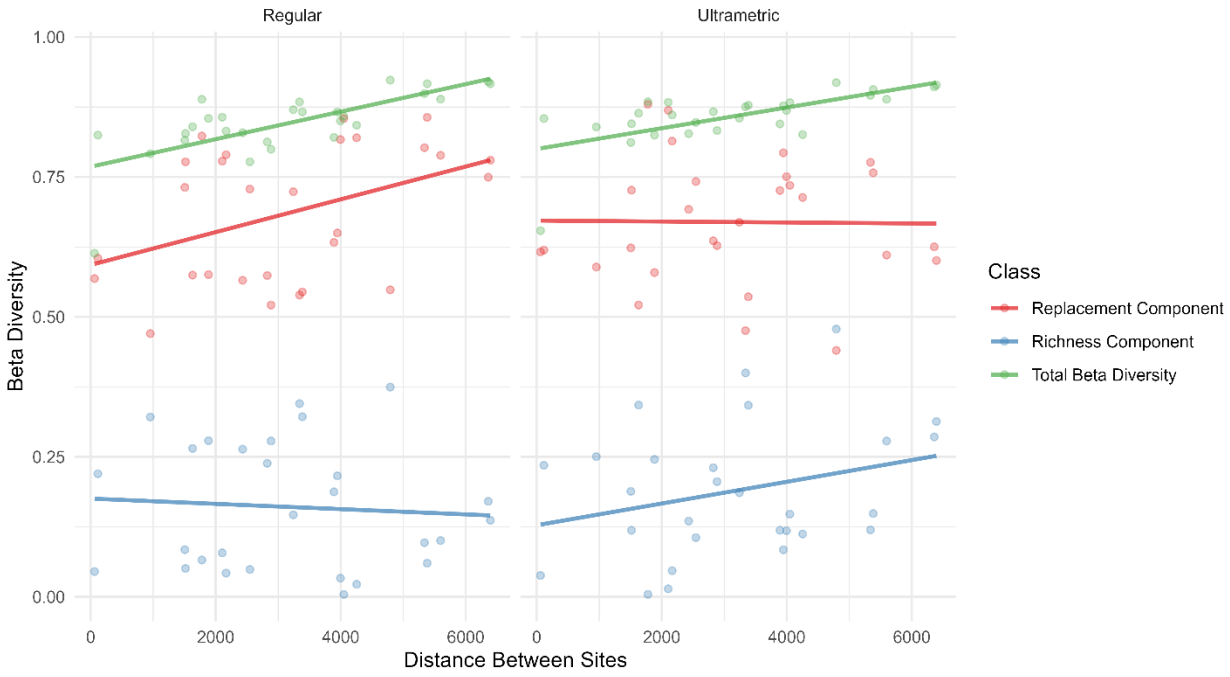


Figure 7. Phylogenetic beta diversity broken into individual components (richness and replacement) and plotted against spatial distance. The blue line represents total beta diversity. The red line shows the replacement component, and the green line shows the richness component. Using the non-ultrametric tree, the importance of replacement increases across distance. This is not found when using the ultrametric tree.

3.3.2. Over- and under-dispersion

There was a near significant difference in the effect sizes between elevation groups (F = 4.83, p-value = 0.068). This was driven by a large difference between the lower elevations (< 700m) and the higher elevations (> 1400m)(Figure 8). There was not a significant difference between comparisons with mid-elevation plots. Elevation was the only slightly significant predictor in the multiple linear regression. When modeled with elevation as the single predictor, the relationship was stronger (R-squared = 0.571, p-value = 0.018)(Figure 9). Additionally, when modeled with forest cover as the single predictor, the relationship was significant (R-squared = 0.547, p-value = 0.023). There is some collinearity in elevation and forest cover, as the lowland sites are the disturbed sites (Supp). Using Kruskal-Wallis test, there was no significant difference when grouped as disturbed and undisturbed.

The lowest elevation site had the lowest observed MPD of 0.634 and was significantly different than the null community (z = -3.44, p-value = 0.001; Table 2). This site additionally had the lowest forest cover (45%). The other lowland site, at 495m, did not have a significant difference from the null community. Sites at higher elevations showed the opposite pattern, with observed MPD higher than the null expectation. This was strongest in the site found at 1743m, with MPD equal to 0.922 (z = 1.78, p-value = 0.944). The site at the highest elevation, 2238m, additionally showed overdispersion, although not significant, with a MPD score of 0.890 (z = 1.01, p-value = 0.857). The second highest elevation site did not show significant deviation from the null model

(MPD = 0.844 versus MPD = 0.834). The observed values for mid elevation site (~1400m) were near identical to the null MPD estimate.

Site	Observed MPD	Z-score	P-value	Null mean	Null S.D.
DKA1	0.6341590	-3.43680981	0.001	0.8218851	0.05462219
DKA2	0.7836728	-0.70603754	0.276	0.8247902	0.05823676
DKB1	0.7480144	-1.35012458	0.063	0.8249545	0.05698742
DKB2	0.8025744	-0.51342715	0.320	0.8243845	0.04247932
DKC2	0.8270518	0.05772203	0.549	0.8248466	0.03820405
DKD2	0.9222167	1.79747487	0.944	0.8239007	0.05469670
DKE1	0.8790246	1.10192573	0.855	0.8235206	0.05037003
DKE2	0.8309210	0.18654961	0.596	0.8215094	0.05045094

Table 2. Results from function *ses.mpd* in package picante, used to calculate the standardized effect size of mean pairwise distance based on a null model. DKA1 was significantly under-dispersed (p = 0.001) while DKD2 was nearly significantly over-dispersed (p-value = 0.944).

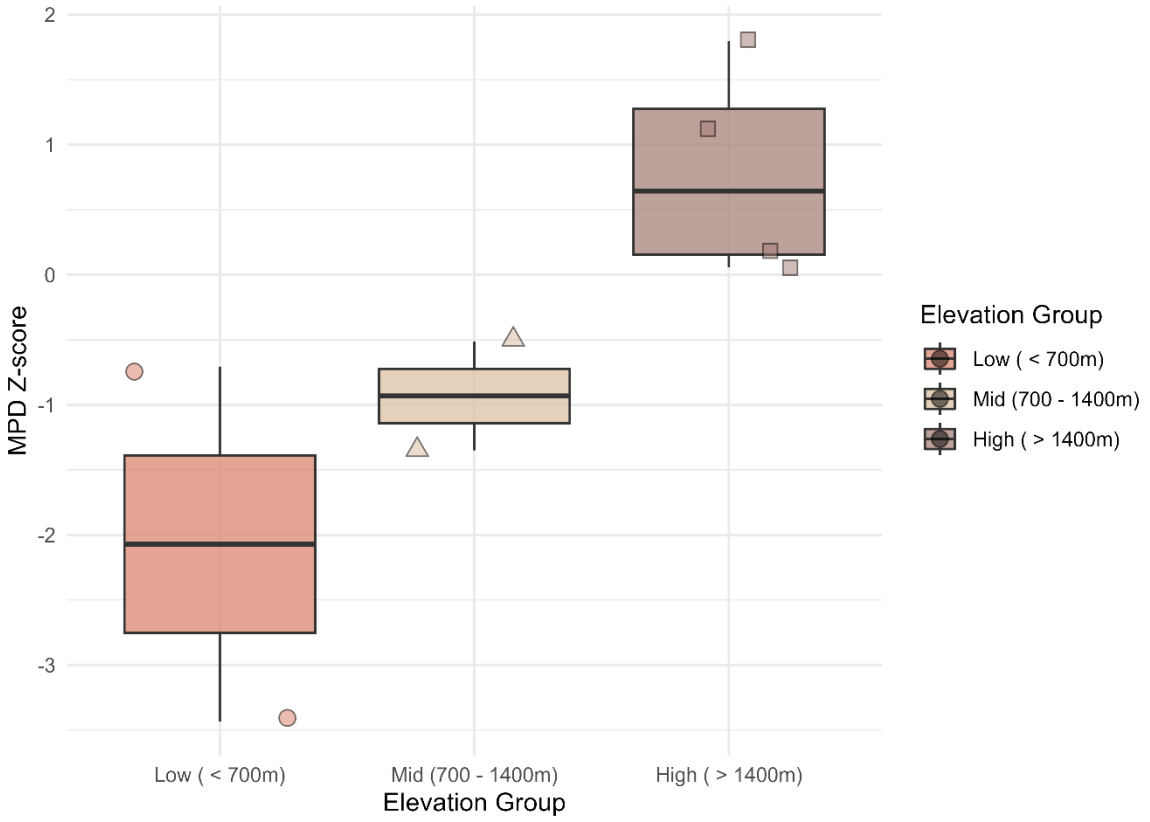


Figure 8. The z-score for mean pairwise distances of sites, grouped by elevation. There was a nearly significant difference in scores by elevation. This was due to the difference in values between the low and the high elevation sites.

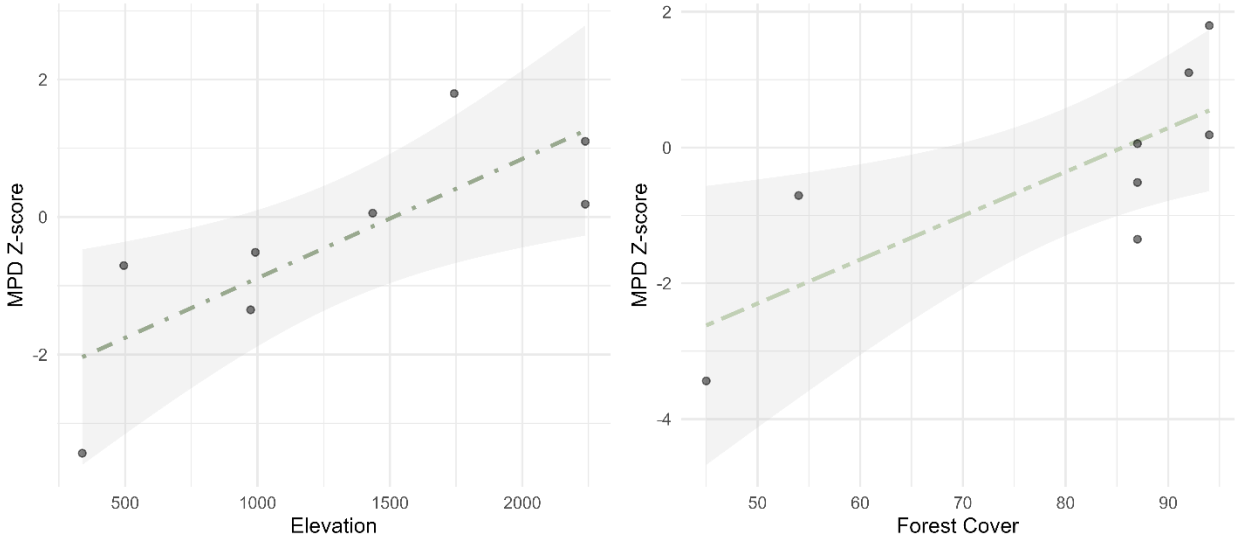


Figure 9. Linear regression of MPD z-scores against a) elevation and against b) forest cover. There was a significant relationship between z-scores and both predictors when modeled individually.

3.3.3. Major processes involved in community assembly

Drift was identified as most important in assembly (0.675) followed by dispersal limitation (0.152) and heterogeneous selection (0.136) (Figure 10). Stochastic processes were the only important processes identified for the mid-elevation sites (900-1400m), consisting of drift (mu = 0.808) and dispersal limitation (mu = 0.192)(Figure 11). Stochastic processes were also most important in high-elevation (mu = 0.799) and low-elevation (mu = 0.624) communities. In the high-elevation sites both drift (mu = 0.638) and dispersal limitation (0.152) were predicted as important, with a small amount of homogenizing dispersal as well (0.008). The stochastic processes at low-elevation were primarily drift-related (0.615) but a small amount of homogenizing dispersal was also detected (0.009). Both low and high elevation sites additionally had deterministic processes that were identified as important, specifically heterogeneous selection. Low-elevation habitats had a large amount of variation associated with heterogeneous selection (0.376) while high-elevation habitats had less (0.201).

When comparing elevation bands, both dispersal limitation (Cohen's d = -2.740, p-value = 0.012) and heterogenous selection (Cohen's d = 5.292, p-value < 0.001) were identified as significant drivers of the differences between communities in the low elevation agroforest sites and the mid-high elevation sites. There were other processes shaping differences between communities that produced large effect sizes; however, only those between the low and mid-high elevation sites were significant.

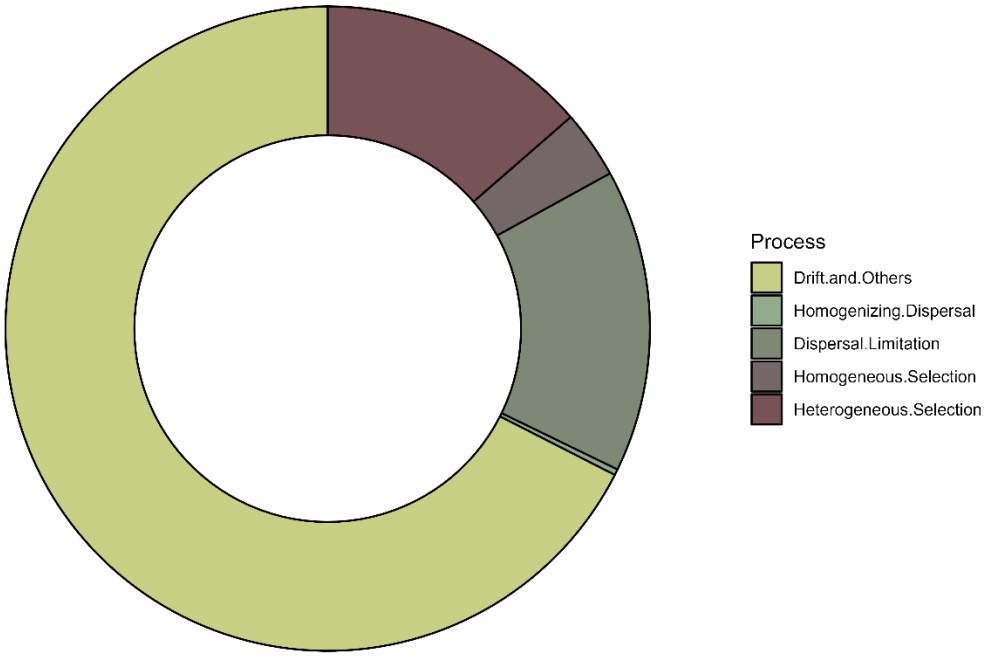


Figure 10. The relative importance of different processes in community assembly. Stochastic drift was identified as most important, followed by dispersal limitation and heterogenous selection.

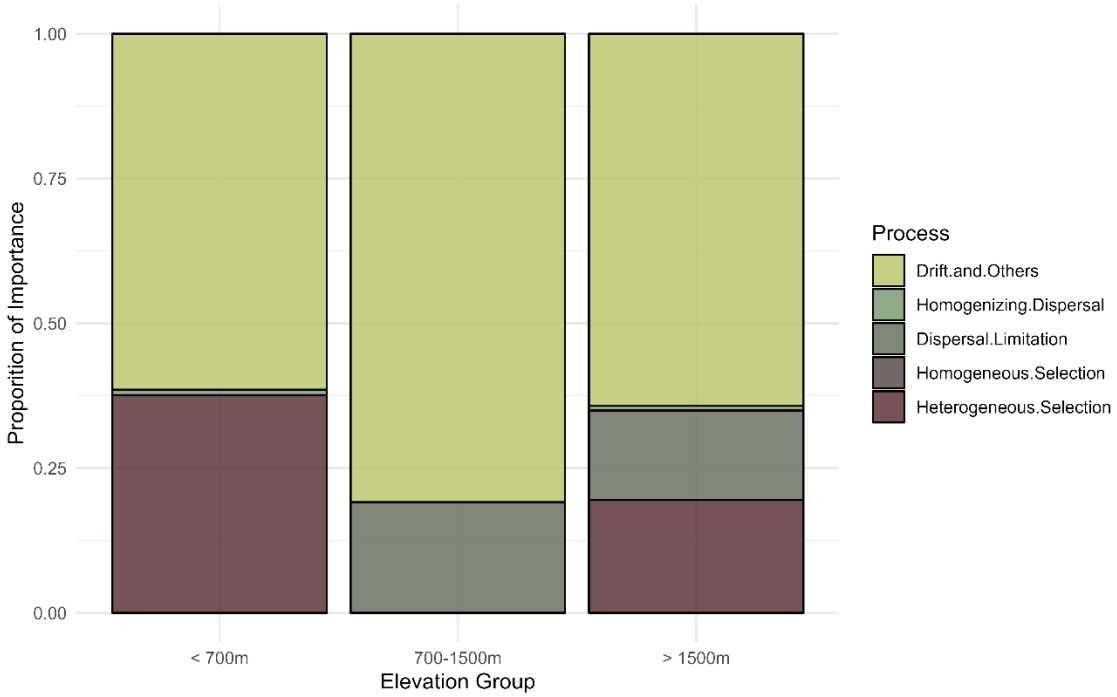


Figure 11. Importance of different processes by elevation group. Stochastic processes alone were identified as important for the mid-elevation sites. The deterministic process of heterogeneous selection was identified as important for both low and high elevation sites. Dispersal limitation was also an important stochastic process identified in both the mid and high elevation communities

3.4. Discussion

Strong differences in community patterns were documented in Sulawesi leaf litter communities across an elevation gradient. Differences were most extreme at the lowest elevations and the highest elevations, with mid-elevation sites showing little phylogenetic structure as well as having stochastic processes, primarily drift, as the most important processes in assembly. The distinctness of the low elevation sites and the high elevation sites are likely related to different mechanisms, with the low elevation sites being shaped most by recent disturbance and deforestation while the signatures at the highest elevation sites reflect the young nature of the habitat as well as the unique environmental pressures imposed by high-elevation habitat. Using this system, we were able to assess the interplay of determinism and stochasticity in structuring leaf litter communities and provide evidence in support of our three hypotheses: a) high-elevation communities are filled by pre-adapted lineages, enforced by niche requirements, b) the low-mid elevation communities had highest diversity and showed no phylogenetic signal, indicative of a near-equilibrium state and c) the communities in the converted agroforestry site caused underdispersion, likely due to the novel environment imparting a strong filter.

3.4.1. Soil diversity

Leaf litter communities are known to be incredibly diverse[129,130], especially in tropical locations[131]. The low percent identity match of almost all sequences in our study is indicative of the degree of unknown diversity. Of our 349 ASVs, only 12 had a 97% or higher identity match (Figure 2). The highest matches were found in lowland habitat. We may expect this to be due to a prevalence of non-native taxa that perform well in altered habitats. In deforested tropical forest sites on Barro Colorado Island, richness differences overall did not change but there were significant differences in soil composition related to removal of leaf litter[132]. We similarly find strong compositional differences (Figure 5) but did not detect a high number of invasives; in fact, only a single ant species was confidently identified as a widespread invasive.

We saw an increase in taxa matching above 95% in the lowland sites (Figure 3); this may symbolize young divergences from closely related taxa in other parts of SE Asia. Tropical sites undergoing restoration methods that increase habitat heterogeneity have been shown to support native soil arthropod fauna[133]. While the sites we sampled were not undergoing restoration, they were agroforestry plots and the maintenance of mixed habitat and ground cover may support some native species (Figure 1). In addition to a general increase in identity match, we found five species that matched above 99% in the lowlands. The species identifications were intriguing. Two species (one ant and one earwig) are known from Papua, a cockroach known from Japan and Christmas Island and finally an ant species only known from Madagascar. For many years, the nature of Sulawesi as a faunal transitional zone was linked to vicariance – species rafted on island fragments that eventually collided to form Sulawesi. This theory has been replaced by one of dispersal because the Sula Spur fragments are now thought to have been submerged throughout their movement and land area only emerged upon fragment collision[69,134]. However, a handful of invertebrate studies do find evidence for an out-of-Australia vicariance hypothesis[71,73], but the evidence is weak and not supported by many other studies on taxa on Sulawesi.

3.4.2. Neutral processes shaping mid-elevation sites

At mid-elevation sites (900-1400m) only stochastic processes were identified as important in structuring communities, with drift being most important followed by dispersal limitation. There was little to no phylogenetic signal detected using MPD. Mid-elevation sites showed the highest phylogenetic diversity as well as a spread in ordination space, indicating high turnover that exists in these communities regardless of environmental similarity or proximity. Given low-mid elevation habitat has existed for 15-25Ma, deterministic processes leading to fine-scale resource partitioning could have created high diversity and filled all available niche space, resulting in communities today being at near-equilibrium states. Short range endemism is common in arthropods and other invertebrates that are soil dwelling[135,136]. The importance of dispersal limitation in assemblages at these sites could support the idea that these communities consist of short-range endemics that have been established in these areas for an extended period of time, and now at a stable state where differences exist solely due to stochasticity.

3.4.3. Niche-based processes in high elevation communities

In both high and low elevation communities, we identified important deterministic processes involved in the assemblage process, specifically heterogenous selection. Upper montane or cloud rainforests in the tropics differ significantly from low elevation rainforest and require more specialized ecological strategies. A large analysis of different taxonomic groups on Mount Kinabalu, Malaysia found taxa rarely made shifts from lowland tropical rainforest to higher elevations[57]. This is likely due to several physiological, morphological and behavioral adaptations that may be required to shift upwards on a mountain. For example, arthropods in tropical alpine habitats display changes in body size, diapause, thermoregulatory behaviors, temperature sensitivity, and other attributes when compared to lower elevation taxa[137,138].

Heterogeneous selection, then, aligns with a strong environmental filter preventing many taxa from establishing without specialized adaptations. However, rather than find phylogenetic under dispersion in line with an environmental filter, we found that two higher elevation sites (1750m and 2200m) had pattern of phylogenetic overdispersion, usually indicative of competitive exclusion. In the higher elevations, niche space may be more limited due to the reduced area. While an environmental filter may drive the initial stages of assemblage in these communities, the most well-adapted taxa may now occupy the limited niche space and prevent other closely related taxa with similar functional roles from establishing. The communities are additionally isolated, demonstrated by the importance of dispersal limitation. Priority effects may have then played an important role in how these communities formed, with pre-adapted lineages entering and establishing then outcompeting other closely related colonizers.

3.4.4. Niche-based processes in low elevation communities

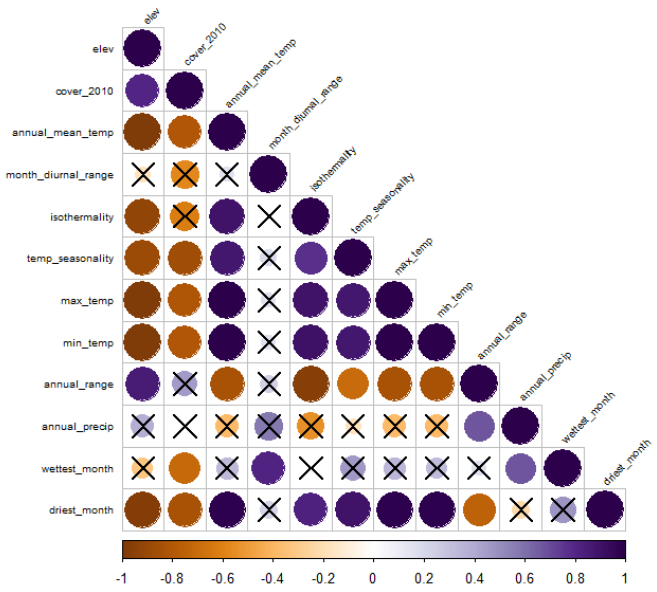
Both low elevation sites clustered together and away from other sites in the ordination plot, indicative of their distinct community composition. We found the differences between the low elevation sites and the mid-high elevation sites to be driven strongly by both dispersal limitation and by heterogenous selection. Similar to the high elevation sites, the exposed and degraded habitat found at the low elevation sites likely requires different ecological specializations.

However, this degraded agroforestry zone represents young habitat, and the conversion has been sudden. This contrasts with the mountain tops; while geologically young, community assembly has been ongoing for 1-3Ma in an environment that has been relatively stable. The conversion to agroforesty in more recent history likely applied a strong filter that quickly altered the native community. The lowlands are additionally less isolated, enabling more widespread species with suitable ecological strategies to establish. One example would be the large *Nephila* orb-weavers; *Nephila pilipes* and *Nephila argentiopunctata* are widespread spider species in Southeast Asia that are frequently found in close proximity to one another in lowland forest and in more disturbed habitat. *Nephila* are well-documented to be aerial dispersers, capable of ballooning long distances[139]. While our leaf litter samples did not include spiders, this gives one example of a lineage-specific adaptation that may lead to closely related species colonizing a given area and capitalizing in an environment that is hostile to other taxa, which could lead to the significant underdispersion we found in the lowest elevation sites.

3.5. Conclusions

Our study demonstrates how different processes shape communities across an environmental gradient. As we hypothesized, neutral processes were most important in shaping mid-elevation communities. Deterministic processes, specifically heterogeneous selection, were identified as important in the low and high-elevation habitats. The differences we see in phylogenetic signal between these sites could be related to the history of each community – the high elevation communities are relatively young and still undergoing deterministic selection processes, but the isolation and age of mountain building may have led to competitive exclusion becoming a prevalent process. The recently converted and less isolated lowland agroforests may instead be shaped by successful generalist species that are closely related and can capitalize on a habitat that is not suitable for other native species. Additionally, because this is a recently disturbed habitat, we may not yet see the full effect on endemic lowland zones. Consideration of these different mechanisms shaping arthropod communities is important when we think about community succession and efforts to restore habitat or promote higher diversity in degraded forest.

3.6. Supplementary Materials



Supplementary Figure 1: Correlation matrix of environmental variables from WorldClim[125]. Variables without significant correlations are marked in X shapes.

PART 2

The impact of global change on biotic communities

In the second part of my dissertation, I focus on two major types of global change – the introduction of non-native taxa and changes in disturbance regimes. I assess how biotic interactions change following invasion by studying the diet of an endemic generalist predator in Hawaii as well as how arthropod community structure in California is altered following novel wildfire regimes in different habitat types

Chapter 4: Invasion by an ecosystem engineer changes biotic interactions between native and non-native taxa

Anna J. Holmquist, Seira A. Adams, Rosemary G. Gillespie

In press: Ecology and Evolution
DOI: 10.1002/ece3.9820

4.1. Introduction

Over recent decades, global human transportation networks have led to the establishment of once geographically restricted species into new ecosystems[140–142]. While many introduced taxa go unnoticed, some may act as "ecosystem engineers" by altering abiotic and biotic factors, leading to changes in the structure of the original ecosystem[143]; such non-native taxa may then be called invasive because of their negative impact on ecosystem services and native species[144]. Invasive plants are particularly pervasive and act as ecosystem engineers by changing soil chemistry[145], nutrient cycling[146], microclimates[147] and the presence and abundance of native taxa[148]. Because of their role as primary producers, the introduction of new plants and the resulting displacement of native flora can produce strong bottom-up effects with far reaching consequences[149].

Native arthropods, and in particular herbivores, are often strongly affected by plant invasion due to their direct interaction with native plants. Non-native flora can induce evolutionary traps for native insects by producing attractive habitat that in fact has fitness costs for species with particular physiological adaptations[150–152]. Alternatively, native taxa with higher behavioral plasticity may be capable of rapid host shifts and benefit from the increased access to resources provided by new plant species[152,153]. The contrasting response of different species to plant invasion results in varied changes to the overall arthropod community. Most studies document decreases in richness and abundance levels in invaded patches but this result is inconsistent[148,149,154].

A major driver of abundance and/or richness changes are shifts in biotic interactions, such as increased predation, lack of the most nutritionally beneficial prey items or a higher interaction with parasites[155,156]. The displacement of native taxa and associated changes in biotic interactions can result in invasional meltdown, in which modification driven by one invasive provides opportunity for the establishment of other non-native taxa[157,158]. The introduction of multiple non-native taxa can result in the significant disruption to the native food web, causing broad ecosystem effects. Because of the widespread consequences that can occur following invasion, it is important to go beyond common diversity metrics such as abundance or species richness and study community network structure across multiple trophic levels[159].

Kāhili ginger (*Hedychium gardnerianum*), a plant native to the Himalayas, has expanded its range globally to the Azores, Madeira, Jamaica, Réunion, New Zealand and Hawaii, as well as across South and Central America, Australia and Southern Africa[160]. *H. gardnerianum* was brought to Hawaii in 1940 and has become an aggressive invasive, capable of establishing in

intact native rain forest, displacing understory vegetation, altering composition of soil microbial decomposers and promoting the establishment of other plant invasives[161,162].

Here, we use a recent invasion by kāhili in Hawaii to measure how the establishment of an invasive ecosystem engineer changes biotic interactions. By assessing the major shifts that occur in biotic interactions following the introduction of ginger using the diet of a native generalist predator, we can ask if major shifts in biotic interactions occur following invasion. Towards understanding the effect of ginger invasion on the interactions between native and non-native taxa, it is essential to identify a highly simplified system where ginger invasion is the primary differentiation between sites. Our sampling sites were located in the mesic forest of Waikamoi on East Maui, where a sharp boundary exists between ginger invasion and the native forest due to the efforts of the Nature Conservancy of Hawaii in protecting their lands[163]. The adjacency of native forest and ginger-invaded sites allows studying invasion in discrete units at a small spatial scale not usually possible in invasion biology.

To further reduce variability, we chose a single endemic genus of spider predators (Pagiopalus, Philodromidae) to assess biotic interactions. While little is known about the ecology of this endemic genus, species in the family Philodromidae are generalist active hunters found on foliage[164]. Like most spiders, they are expected to consume prey at rates roughly proportional to their availability in the environment. Therefore, we use these spiders as a vehicle to study relative proportions of native and non-native species, as well as shifts in biotic interactions. We utilized metabarcoding to compare diet composition and parasite loads in spiders across adjacent sites in ginger-invaded habitat and native-forest. By assessing biotic interactions using the diet of a native generalist predator in a simplified system, we can ask if major shifts in the community network occur following invasion.

We have three hypotheses related to the effect of an invasive ecosystem engineer on relative proportions of native and non-native taxa and the associated shifts in biotic interactions. First, the altered environmental conditions in ginger will result in arthropod communities differing from native forest sites; we expect to see this reflected in compositional differences in the diets between spiders collected in ginger-invaded habitat and native forest. Second, ginger sites will host more non-native prey items, reflected again in the diet of Pagiopalus specimens. Lastly, a higher interaction between arthropods and non-native parasites will exist in invaded forest, due either to parasitism of the spider or by secondary consumption.

4.2. Methods

4.2.1. Study system

The study site sits across two adjacent reserves on East Maui - The Nature Conservancy (TNC) of Hawaii's Waikamoi Preserve and the Makawao Forest Reserve. This area was invaded by ginger in the early 1980s. The ginger has spread across the reserves, significantly increasing in density and coverage over the last decades[163]. The Waikamoi Preserve is actively managed by TNC, who regularly remove ginger seedlings throughout the preserve to maintain a largely native landscape within the fenced off area. In comparison, the Makawao Forest Reserve is

maintained less frequently and thus a thick stand of ginger covers much of the reserve, meeting the fence line that separates the adjacent Waikamoi Preserve.

4.2.2. Sampling

To investigate the effects of invasive ginger, we laid five transects in the native mesic forest habitat of the Waikamoi Preserve and five transects in ginger-invaded habitat of the Makawao Forest Reserve (Supplementary Materials Figure 1; see Data Accessibility for coordinates). Each transect ran 30 meters long and was 3 meters in width. Spiders were collected along transects between June 8 to June 21 in 2017 using vegetation beat sampling. Each transect was broken into 3 blocks, each 10 meters in length. Combined transect number and block are used as "sampling units" throughout the paper for a total of 30 unique units. Vegetation along transects was sampled using a beat sheet. Fifteen areas of vegetation were beaten in each block of each transect, totaling 75 seconds of beating per block. Spiders were collected from the beat sheet using an insect aspirator and preserved in 100% EtOH in individual 2mL vials. Samples were stored in a -20°C freezer until further use.

4.2.3. Molecular procedures

Spiders were identified morphologically and Pagiopalus specimens were retained. To extract DNA from the gut of each spider, the opisthosoma (abdomen) was first cleaned using 70% ethanol and water to remove external DNA contamination. The opisthosoma was removed using a sterile scalpel blade and placed in Qiagen cell lysis solution. The opisthosoma was then ground using two 3 mm steel beads on a Genogrinder (Spex SamplePrep, Matuchen, NJ, USA) for 2 min at 1,200 hz. DNA from ground samples was extracted using the Qiagen Puregene kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. DNA was then amplified using primer sets for 28s, 18s, and 16s, optimized for spider gut content amplification[165] (Supplementary Materials: Table 1). COI was unsuccessful at amplifying prey in this genus and was not used[165]. 28s has been used for fungal identification[166] and was therefore appropriate for detecting both arthropods in spider gut content and fungi. Amplification was performed using the Qiagen multiplex PCR kit (Qiagen, Hilden, Germany) in 10 µl reactions with 1 µl of template DNA, and 1 µl 10uM primer dilutions of each primer for 35 PCR cycles. Nuclear markers were multiplexed and amplification was performed at an annealing temperature of 55°C. Amplification of the mitochondrial 16s was performed independently at an annealing temperature of 46°C. Negative PCR controls were included in each round of amplification. Primers included a 5` tail that allowed binding of 8bp indexing primers and TruSeq adapters (Illumina San Diego, CA, USA) performed in a second indexing PCR of five cycles[165]. PCR products were visualized using a 1.5% agarose gel. Products were pooled in approximately equal amounts based on band strength. 1X AmpureBeads (Beckman Coulter Life Sciences, Indianapolis, IN USA) were used to clean the pooled products. Negative controls were included in the final library. The cleaned library was then sequenced on an Illumina Miseq (Illumina San Diego, CA, USA) using V3 chemistry alongside other libraries not associated with the project; 1.71 million reads were expected.

4.2.4. Bioinformatics

Sequences were demultiplexed using Illumina Basespace (Illumina San Diego, CA, USA). Demultiplexed sequences were batch processed using CutAdapt to remove primer sequences and perform preliminary quality filtering[38]. The denoising algorithm DADA2 was run in R to produce amplicon sequencing variants (ASVs) and remove chimeras[39]. Parameterization and length trimming was dependent on locus. Widespread contaminants were identified using package *decontam* in R using sequences identified in controls[40]; the threshold argument was set to 0.5, which removes sequences more prevalent in controls than in true samples. Once contaminants were removed, ASVs were further curated using the LULU algorithm in R, which identifies NuMTs and any remaining artefactual sequences[41]. Finally, sequences were filtered at the ASV and sample level by read counts: ASVs represented by less than 10 reads were removed. and the number of reads found for an ASV within a sample that represented less than 0.01% of the total ASV reads or that represented less than 1% of the total reads for the sample were removed. Additionally, one sample had missing metadata and was removed.

ASVs were then written to FASTA files and, using Geneious Prime v. 2022.0.2, assigned taxonomic identities through megablast. Full taxonomic lineage was assigned using a custom R script based on the package *rentrez*[167]. Sequences were retained if similarity was equal to or greater than 85%[165]. A species-level identification was retained if the percent match was >= 99%. There is little information on appropriate cut-offs for assignment to genus or family using rRNA genes; 97% was used for genus and 95% was used for family to avoid retention of incorrect taxonomic IDs. Only order was assigned if percent matches were lower (85-94%). To construct the prey data set, phylum Arthropoda was selected from filtered sequences. Arthropod reads matching order Araneae were filtered to remove any sequences generated from the spider itself; while Araneae could include spiders eaten as prey, this was not confidently identifiable. Sequences in the order Hymenoptera were additionally removed from the prey dataset and treated as parasites because generalist spiders are rarely found to prey on Hymenoptera. Parasitism of Hawaiian Philodromidae is common[168] as is parasitism of their arthropod prey, making reads likely either secondary consumption or parasitism of the spider itself rather than prey reads.

Prey reads were assigned native or non-native status based on publicly available literature. Likewise, sequences identified as entomopathogenic fungi and parasitic wasps were assigned to native versus non-native status using publicly available literature. An order-level network grid was constructed to visualize differences in ginger-invaded sites and native forest using the package *bipartite*[169]. A network grid was also constructed to visualize the presence of different families of parasites in ginger-invaded sites and native forest.

4.2.5. Statistical analyses

Community matrices were constructed using data generated from all markers. Both ASV and order were used as taxonomic units; order-level matrices provided the coarsest view of diets while ASVs provided a species-level view. Matrices were populated with either binary incidence data or Hellinger-transformed relative read abundances. While read abundances have not been found to be true reflections of abundances, relative read abundances can provide useful

information in analyses because presence-absence data places higher emphasis on rare taxa and can increase the strength of compositional differences among communities[170].

Hill numbers[171] were calculated to quantify prey diversity for individual spiders using package vegan[46]. Welch's t-test was used to test the hypothesis of no differences in dietary prey diversity between spiders in ginger and native forest. Compositional differences were first assessed using taxonomic beta diversity, using function *beta* in package BAT[49]. Because of the low dietary overlap between any two individual spiders discovered using beta diversity (Supplementary Materials, Figure 2), sampling units were used for further compositional analyses. Beta diversity values were calculated between sampling units using package BAT[49]. Differences across sites grouped as between habitat and within each habitat were tested using ANOVA and Tukey's test was used to test if sites between habitats were more compositionally dissimilar than sites within either ginger-invaded or native forest sites. To assess overall differences between the diets of spiders in ginger sites versus native forests, distance matrices were calculated using Jaccard distance for incidence data and Hellinger-distance for relative read abundances. Non-metric dimensional scaling (NMDS) was performed using each distance matrix using a maximum of 1,000 random starts; k = 3 was used for ASV matrices and k = 2 for order matrices to achieve convergence. To test for group differences, permutational multivariate analysis of variance (PERMANOVA) was performed with ginger-invaded and native forest as the independent grouping variables using vegan. Multivariate homogeneity of group dispersions was also tested using package vegan.

Parasite frequency was calculated using the number of ASVs per spider within ginger and native forest sites that were identified as parasitic. Native and non-native status as well as most common arthropod host was determined by the literature for taxa identifiable to genus or species. Welch's t-test was used to test the hypothesis of no differences in parasitic load between ginger and native forest sites.

Analyses were conducted in R version 4.2.2. Data and code for analysis is available on GitHub and Dryad (see Data Accessibility).

4.3. Results

4.3.1. Summary of collections

168 Pagiopalus spp. were collected in total; abundances were nearly equal across ginger and native forest with 82 specimens collected from ginger sites and 86 specimens collected from native forest. There were on average 5.79 ± 0.55 spiders collected and 16.8 ± 2.4 spiders collected in total per transect.

4.3.2. Summary of molecular findings

Of the 1,370,873 reads retained following DADA2 and additional filtering, 1,059,644 reads were not prey reads; this reduction came largely from sequences of the spiders themselves (679,534 or 64.1% Philodromidae reads) followed by fungi (234,630 reads, 22.1%) (Supplementary Materials: Figure 3). Sequences of spiders were almost exclusively produced by the 28s marker pair (679,127 of 679,534 reads). Order Hymenoptera represented an additional 60,060 (5.7%) of

the non-prey reads. Spiders collected in ginger-invaded habitat produced more reads than spiders collected in native forest prior to filtering (1,111,019 reads versus 262,213 reads). This came from the presence of additional phyla and non-arthropod reads in higher abundances, particularly Basidiomycota which returned 136,932 reads from ginger and only 35,212 from native forest. Additionally, there were more reads identified as the spiders themselves in the samples from the ginger-invaded sites (518,456 versus 161,078), which may indicate higher sequencing coverage for unidentifiable reasons.

Following prey read curation, a total of 311,229 reads were retained from 82 spiders collected in ginger-invaded sites and 63 spiders collected in native forest for a total of 145 spiders. There was an average of 3,468.9 ± 449.0 reads from the spiders in invaded habitat and 700.8 ± 133.2 from spiders in the native forest transects, reflecting the higher sequencing depth for the spiders in the ginger-invaded sites. 16s returned the highest number of prey reads (205,236 reads) and ASVs (110 ASVs), followed by 28s (91,148 reads and 44 ASVs) and lastly 18s (14,845 reads and 10 ASVs). This was reflected in the sequencing success across samples; prey reads were produced for 112 specimens using 16s, 76 specimens using 28s and 17 specimens using 18s reads.

4.3.3. Summary of taxonomy

There were 23 species, 32 genera, and 31 families belonging to 9 orders detected from the gut content of the Pagiopalus across ginger and native forest sites (Supplementary Materials: Table 2). Only 23.2% and 37.8% of ASVs returned a confident species or genus ID respectively. This was improved at the family-level; 61.6% of ASVs returned family IDs at a 95% match. 16s sequences returned the lowest matches at 91.58% pairwise identity on average. While returning the lowest ASVs and retained specimens, 18s had the highest matches at 98.27%; 28s was near with an average match of 97.72%. The low percent identity matches for many ASVs biased taxonomic results at the species and genus level towards well-known taxa present in GenBank.

4.3.4. Prey diversity and abundance

Of the total 164 ASVs in the dataset, only 29 ASVs were shared across ginger and native-forest sites. Dietary richness was higher in ginger-invaded habitat than in native forest, with 66 total ASVs found in native forest compared to 127 total ASVs found in the spider diets in ginger-invaded sites. Taxonomic composition showed similar trends; 11 of 31 families, 10 of 32 genera, and 5 of 23 species were shared across ginger and native forest sites.

Hill numbers were used to quantify prey diversity between ginger-invaded and native forest spiders based on ASV and taxonomic identity. Spiders rather than sites were used to examine dietary breadth at the individual level. Mean values for dietary richness were determined to be significantly different between ginger-invaded habitat and native forest for both ASV and order (p-value < 0.01; Table 1), with spiders in ginger sites having higher diversity on average, representing a wider niche breadth than spiders in native forest.

4.3.5. Dietary composition

Beta diversity values based on individual spider comparisons were dominated by values of 1 (entirely dissimilar diets) (Supplementary Materials Figure 2). Because of this high dissimilarity,

sampling units were used to further explore dietary differences between ginger and native forest; this was done to allow assessment of the differences between habitat, rather the dietary diversity within Pagiopalus. For both order- and ASV-data, there were significant differences in compositional dissimilarity between groups – comparing diets within invaded habitat, diets within native forest, and diets between invaded and native forest (Figure 1). Using ASV data, the dissimilarity values between the diets in ginger-invaded habitat and native forest were significantly different than diets within either habitat type using both ASVs and order (p-value < 0.001). However, beta diversity was generally high using ASVs ($\bar{x}H = 0.944$, $\bar{x}I = 0.940$). Different trends emerged when using order; in particular, a higher dissimilarity of diets within ginger itself was detected. When using incidence data, the dissimilarity of diets within ginger was in fact higher ($\bar{x}I = 0.650$) than the dissimilarity of the diets across invaded and native habitat ($\bar{x}I = 0.629$). Diets within native forest were the most similar ($\bar{x}H = 0.444$, $\bar{x}I = 0.311$).

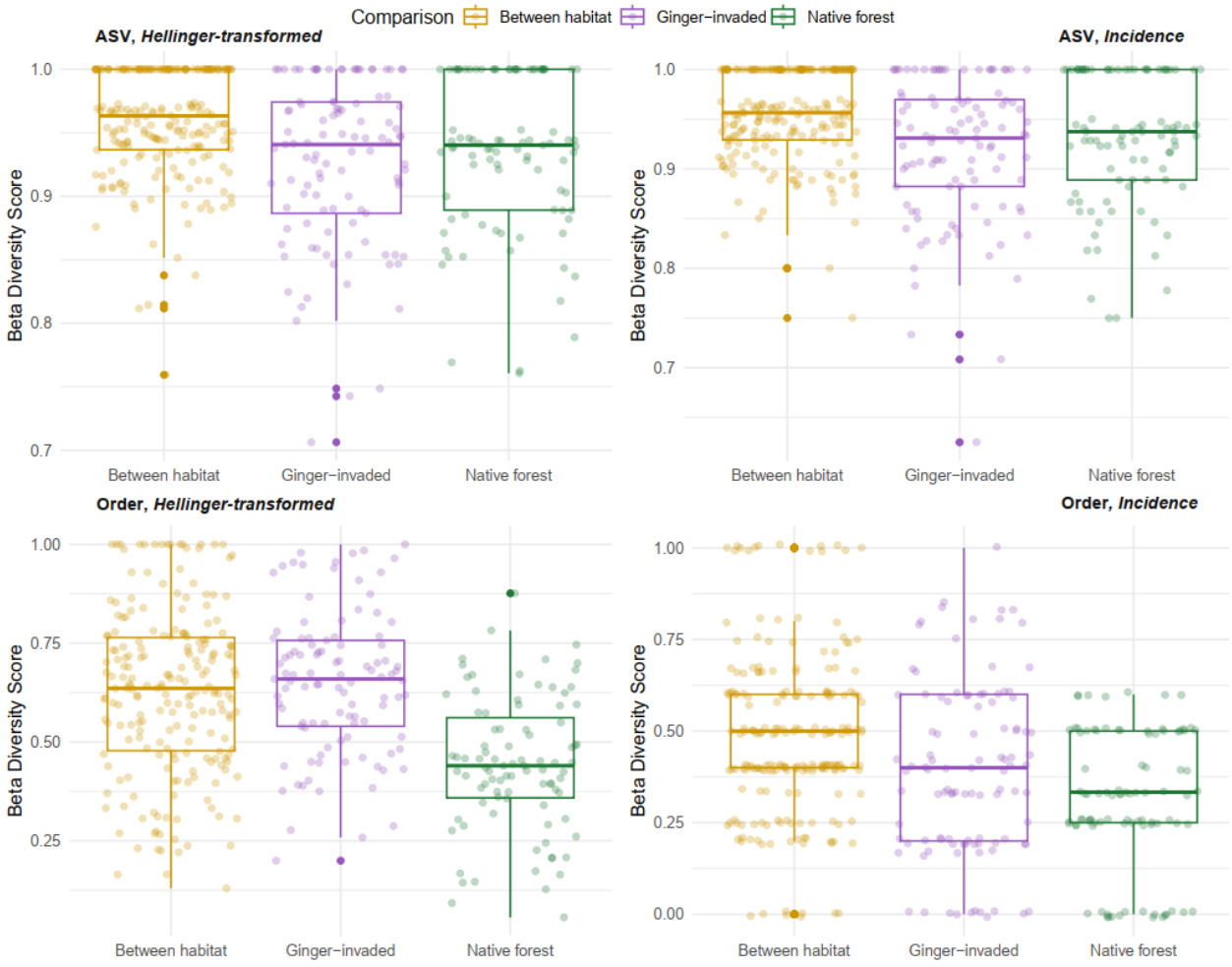


Figure 1 Dietary dissimilarities using beta diversity between invaded and native forest sites (yellow), within ginger-invaded sites (purple) and within native forest sites (green). Column 1 shows beta diversity calculated using incidence data, based on ASVs (top) and order (bottom). Column 2 shows beta diversity calculated using Hellinger-transformed reads, based on ASVs (top) and order (bottom).

No significant difference in group dispersions were detected for any matrix combination using PERMDISP2 (Supplementary Materials: Table 3). When performing PERMANOVA using Hellinger-distance, no significant differences emerged between the dietary composition of ginger-invaded sites and native forest, using ASV or order (Table 2). However, when using incidence data, significant differences were detected (Table 2) with the highest R2 value emerging when assessed by order. Visual assessment using ordination plots shows the higher overlap that is produced when using relative read abundances, while more separation exists when using incidence data (Figure 2).

	Mean difference	Ginger invaded	Native forest	t-statistic	p-value
ASV					
q = 0	0.782	2.885	2.103	3.253	0.0014**
q = 1	0.656	2.522	1.866	3.33	0.0011**
q = 2	0.575	2.309	1.734	3.352	0.001**
Order					
q = 0	0.432	1.846	1.414	3.635	4e-04***
q = 1	0.341	1.667	1.326	3.381	9e-04***
q = 2	0.298	1.578	1.28	3.25	0.0015**

Table 1. Results of the Welch's *t*-test for differences in Hill numbers (q = 0, 1, 2) between invaded and native habitat using ASV or order diversity. Values were significantly different at all values and taxonomic levels.

	F	MSE	SSE	Pr(>F)
ASV				
Hellinger	13.53	0.13020	0.2600	2.06e-06***
Incidence	14.09	0.04772	0.0954	1.22e-06***
Order				
Hellinger	154.90	17.46700	34.9000	<2e-16***
Incidence	166.70	15.66300	31.3000	<2e-16***

Table 2. Results of PERMANOVA testing for group differences in dietary composition between spiders collected from invaded and native habitat. Distances calculated using incidence data show significant differences

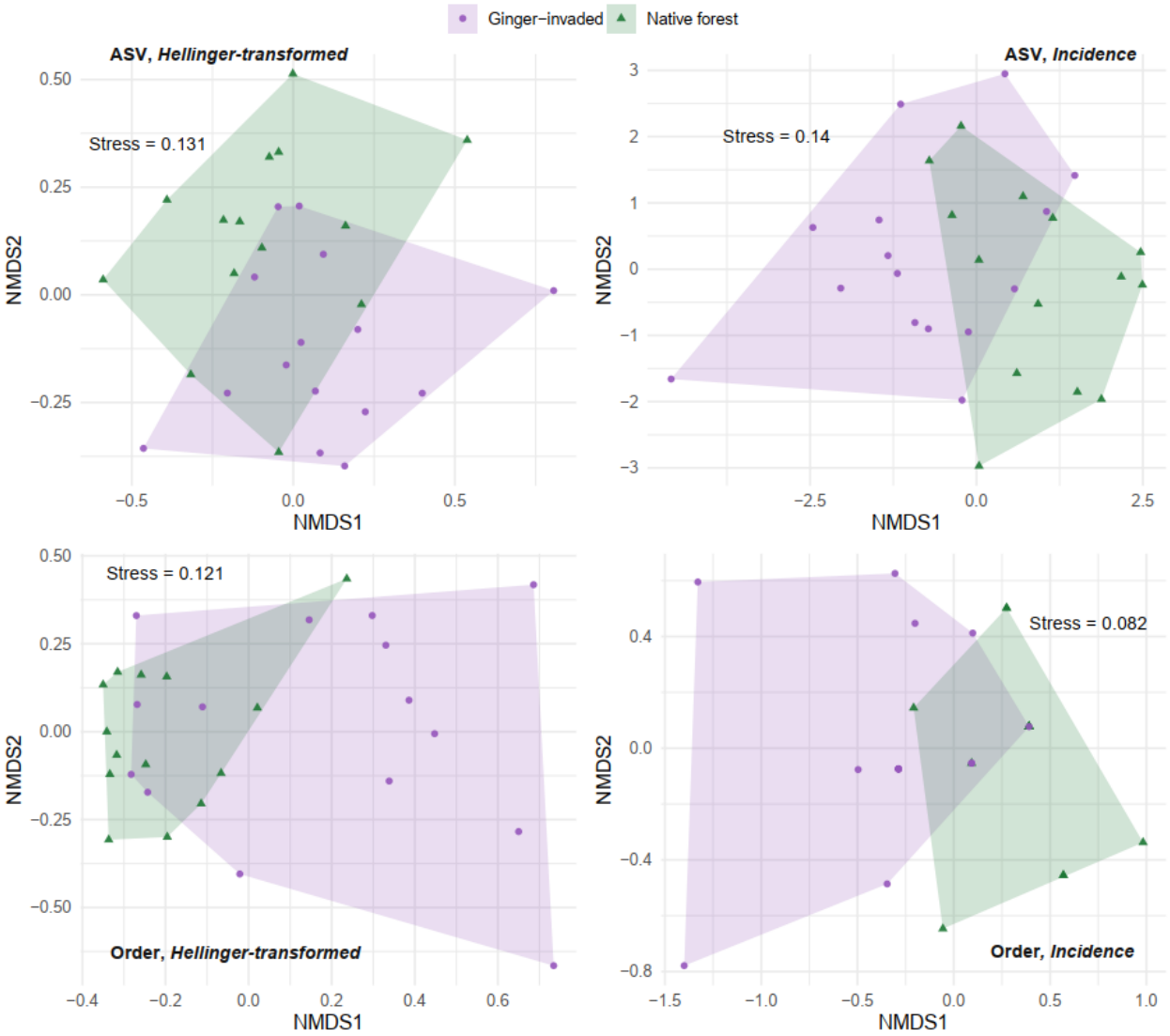


Figure 2. NMDS plots of dietary composition at each site, with shape and color representing sites in ginger-invaded habitat (purple points) and native forest (green triangles). Column 1 shows distances calculated using incidence data, based on ASVs (top) and order (bottom). Column 2 shows distances calculated using Hellinger-transformed reads, based on ASVs (top) and order (bottom).

There is high compositional turnover in the diets of spiders within the ginger habitat when we look at order-level diversity; this is reflected in the beta-diversity values (Figure 1) and in the NMDS plots (Figure 2), where the polygon encompassing ginger-invaded sites occupies a larger portion of the ordination space. The diets of spiders in ginger-invaded habitat are not dominated by any one order, with the most common prey (Hemiptera) detected in the diet of 43.6% of ginger spiders followed by Lepidoptera in 42.3% of spiders and Diptera in 37.1% of spiders. In contrast, spiders in native forest are consuming more similar diets consisting predominantly of Hemiptera, with 72.4% of spider diets containing Hemiptera. The second most common order, Lepidoptera, only occurs in 36.2% of spider diets. Entomobryomorpha was the fourth most common prey group in ginger-invaded habitat, detected in 27 spiders (34.6%) while Entomobryomorpha were only detected in 1 spider (1.7%) in native forest sites (Figure 3).

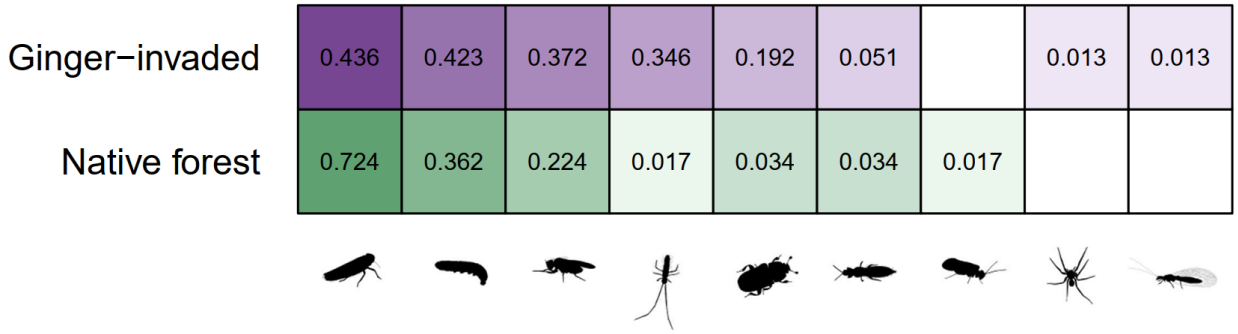


Figure 3. Proportion of spiders in ginger-invaded habitat and in native forest eating different orders of arthropod prey.

4.3.6. Native versus introduced prey items

Because of low BLAST matches, only 53 of 164 ASVs were identifiable as native versus non-native prey taxa. Therefore, most ASVs in the diets of spiders in ginger-invaded or native forest were not assigned an endemism status. With those that were identifiable, we find more non-native prey ASVs (20 ASVs) in ginger forest than in native forest (4 ASVs). Looking at the diets of individual spiders, we find that more spiders in ginger-invaded habitat are consuming diets consisting entirely of non-native prey (27 spiders) than diets consisting of native prey (21 spiders). 9 spiders were consuming both native and non-native prey (Figure 4). In native forest, 3 spiders were detected eating entirely non-native prey with an additional spider eating both native and non-native. All Entomobryomorpha detected in the diets of spiders are non-native. More unidentifiable sequences were detected in native forest; the native/non-native status in the diets of 33 of the 58 spiders from native forest were unknown while 21 of 78 spiders from ginger-invaded forest had unknown diets. This could relate to lack of presence in GenBank and indicate higher concentration of native prey taxa, while adventive or introduced taxa are more well-represented resulting in a higher level of ASV identification in ginger sites.

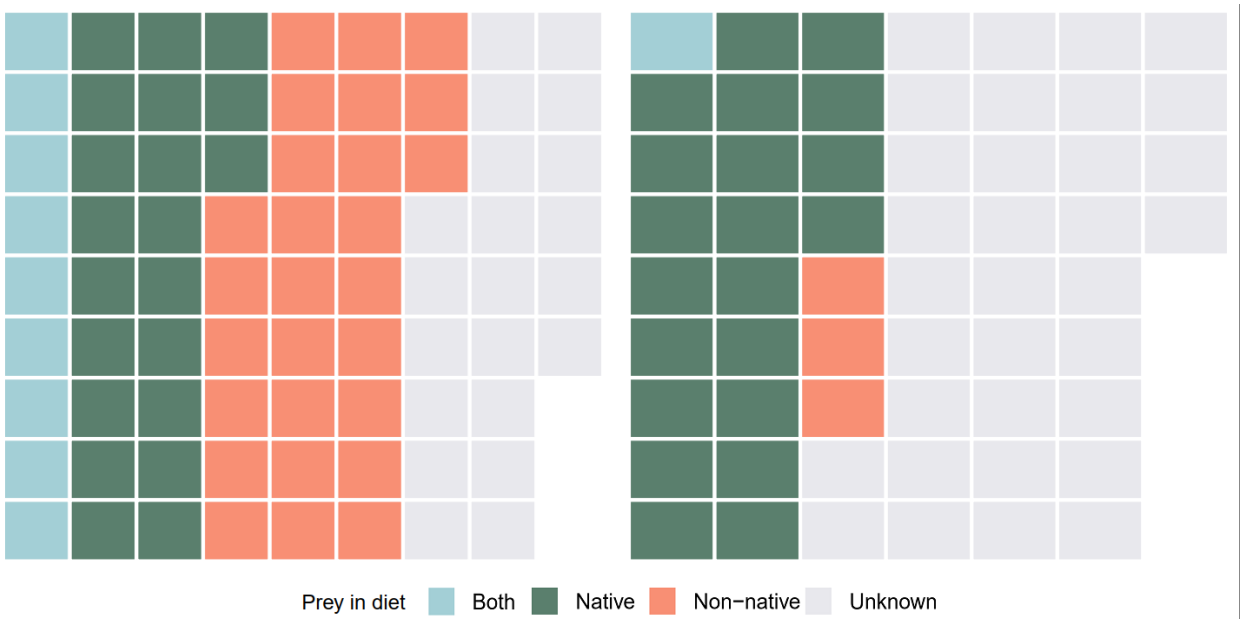
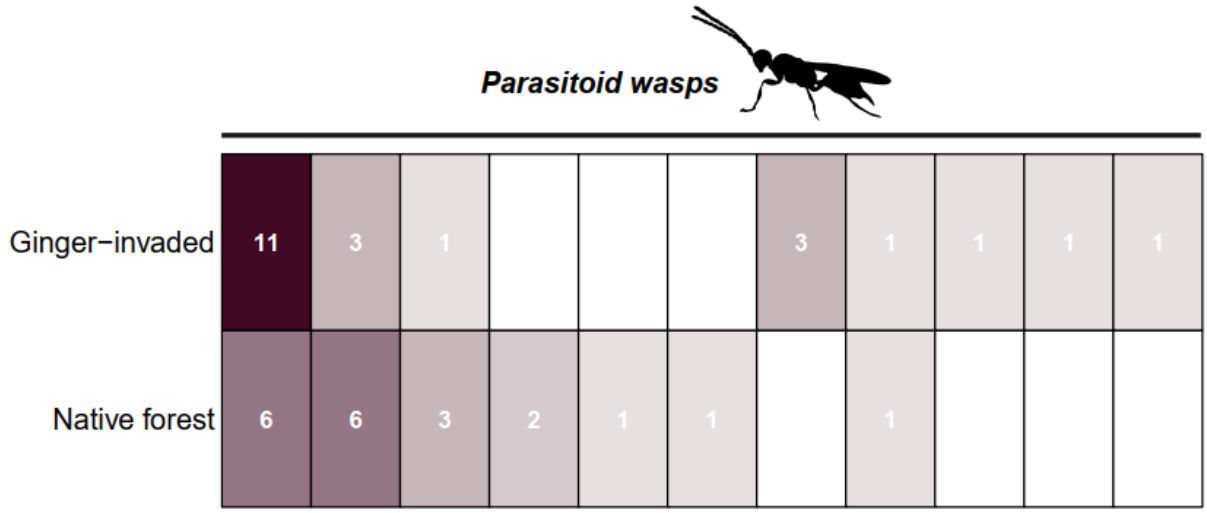


Figure 4. Dietary composition of each spider (represented by a single square), colored by whether prey in diet was entirely native, entirely non-native, or mixed in origins. Grey squares represent diets consisting of unidentifiable prey items.

4.3.7. Parasitism

49 spiders had reads from entomopathogenic fungi or parasitic wasps (Figure 5). Using number of ASVs associated with parasites in each individual spider, there was a slightly significant difference between ginger-invaded habitat and native forest sites, with spiders from ginger-invaded sites having an average of 0.58 ASVs associated with parasites compared to spiders from native forest having on average 0.35 ASVs associated with parasites (Welch t-test; t = -2.07, p-value = 0.0455). 11 spiders collected from ginger-invaded habitat had more than one ASV identified as parasitic compared to 4 from native forest.



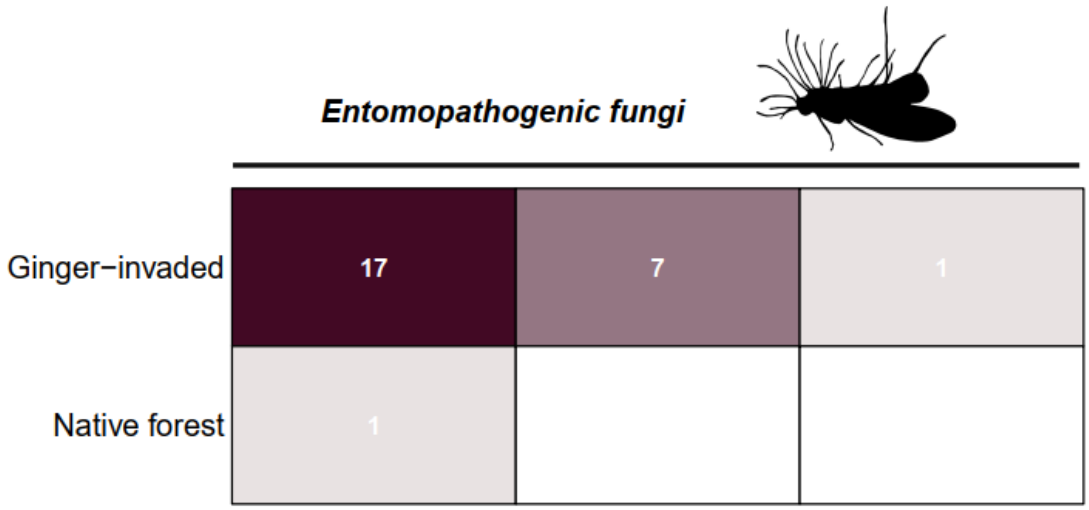


Figure 5. Number of spiders (in grid cells) detected with parasitoid wasp sequences and/or entomopathogenic fungal sequences in ginger-invaded habitat and native forest habitat. Color intensity relates to the number of spiders.

34 of the 49 spiders had sequences identified as hymenopteran. Of the identifiable species, all but one were identified as non-native (97.7% non-native). The same number of spiders in ginger-invaded habitat and native forest had parasitic wasp ASVs. It is worth noting that 18 spiders from ginger sites returned hymenopteran ASVs, in particular the parasitic Pteromalidae, but ASVs were removed due to BLAST matches under 85%; this reduced the number of spiders with parasitoids from 38 to 18 in ginger-invaded habitat, while only removing one record of parasitized spider from native habitat.

Braconid wasps, predominantly Rhopalophorus and Cotesia, were most common among sites, although detected in a higher number of spiders from ginger-invaded habitat (11) than native forest (6). This was followed by ichneumonid wasps, predominantly Ichneumon. Ichneumonid wasps were more common in native forest, detected in six spiders, than in ginger-invaded habitat, detected in three spiders. Two genera of ichneumonids were unique to native forest. Families Bethylidae, Eucharitidae, Eulophidae and Halictidae were unique to ginger-invaded habitat, with Eucharitidae detected in three spiders. Families Diapriidae, Pteromalidae, Signiphoridae and Sphecidae were detected only in only in native forest.

Entomopathogenic fungi were identified in 23 spiders in ginger sites and only 1 spider in native forest, a difference that was highly significant using number of ASVs per spider (Welch t-test; t = -4.9051, p-value = 4.411e-06). Beauveria was identified in 16 spiders in ginger, followed by Ophiocordyceps in 7 spiders. Most spiders (21 of 23) had a single fungal type. However, Beauveria and Ophiocordyceps were found co-occurring in two spiders in ginger forest. The single spider in native forest was identified with Gibellula, a known arachnid parasitic fungus. This genus was not detected in ginger sites, although it falls in the same family as Beauveria.

4.4. Discussion

The goal of our study was to examine if invasion by a plant which drastically alters the environment would alter the relative composition of native and non-native arthropod taxa, and their associated biotic interactions. Using Pagiopalus spiders as a means for sampling arthropods to reveal both dietary composition and presence of parasites, our results show major shifts in biotic interactions across native and invaded forest as well as an increased prevalence of non-native prey taxa in invaded habitat.

Despite the different habitat produced by ginger, we found Pagiopalus in nearly identical abundances. Previous findings showed the invasion of guava (*Psidium* spp.) in lowland forest of Hawaii is associated with an almost entirely endemic spider community[76,172] suggesting that vegetation does not constitute a major perturbation to generalist arthropods[173]. Similarly, native Hawaiian land snails have been found to prefer invasive ginger to native plant species suggesting no apparent negative effect when the understory plant assemblage shifts[174]. However, as discussed earlier, diversity metrics such as abundance are inadequate to assess the true impact of invasion on native taxa; an equal abundance of *Pagiopalus* in ginger does not eliminate the possibility that invasion produces compositional shifts and changes biotic interactions.

Our results showed that, in areas modified by invasive ginger, spiders are consuming a partially overlapping but distinct spectrum of prey items compared to spiders in native forest (Figure 2). Our results showed that native taxa were found in the diet of the endemic *Pagiopalus* in both native forest and ginger-invaded sites: *Campsicnemus* (Dolichopodidae) in 10 spiders from ginger sites and 3 from native forest, *Limonia* (Limoniidae) in 10 spiders from ginger sites and 4 from native forest. Hemipterans were detected in 42 of 58 spiders in native forest (72.4%), with the endemic genus *Nesophrosyne* (Hemiptera: Cicadellidae) detected in 13 spiders; hemipterans were also a common prey source in the spiders from invaded habitat but were detected in only 34 of 78 spiders spiders (43.6%). Like some other hemipteran families, the family Cicadellidae are specific to native plant species[175] (Bennett, pers comm). As the dominant order in the diets of spiders collected in native forest and as herbivores which are often tightly associated with native flora, these hemipterans likely represent an important prey source with which *Pagiopalus* spp. evolved. Because the host plants are absent from heavily invaded ginger sites, access to host-specific native prey may be limited for spiders in ginger-invaded forest, and only accessible in sites in close proximity to native habitat.

The lack of native prey taxa may be connected to the varied diets found in the spiders from ginger-invaded habitat which consisted of prey orders uncommonly found in the diets of spiders collected from native forest (Figure 3). In invaded sites, a more diverse prey community was detected in totality, reflected in the individual diets of spiders which showed wider breadth than spiders from native sites. Trophic dispersion has been detected in other studies following invasion, which can be followed by destabilization of the food network and significant alteration to the biotic community[176]. The increased dietary diversity and the unique taxa being consumed in ginger-invaded habitat point to changes in the broader arthropod community, although more general sampling would be needed to detect the extent of this change. Insertion of new interaction pathways may lead to destabilization and result in ecological state changes.

Differences between the diets of *Pagiopalus* in ginger and native-forest sites were partially driven by the presence of non-native taxa (Figure 4), specifically *Entomobryomorpha* (Collembola) which was detected in the diets of 28 spiders from ginger while being almost entirely absent in the diets of spiders from the native sites (Figure 3). The collembolan *Salina celebensis* was the most prevalent in ginger (detected in 18 of 82 spiders in ginger-invaded habitat or 22.0%) and was not detected in native forest. This species, introduced from Asia, is characteristic of moist understory vegetation, with its extraordinary abundance noted in previous studies[177–179]. The other species of Collembola was *Tomocerus sp*.; although not identifiable to species, *T. minor*, introduced from Europe has been known from Hawaii for > 50 years[177]. *Tomocerus* was found in a single spider in native forest in contrast to 15 spiders from ginger sites.

The much higher numbers of Collembola in invaded habitat may enhance the survival of native spiders such as *Pagiopalus*, simply because of their abundance; alternatively, they may serve to detract from their survival if they do not support the nutritional needs of the predator. Generalist predators are well documented to be dietarily selective when given a diverse set of prey, preferentially eating prey items which have the highest nutritional benefit[180–182]. To obtain nutritional balance, however, a mixed diet consisting of nutritionally imbalanced or even toxic prey may be consumed. Previous work has shown that, while being a common source of prey in cursorial spiders[183], Collembola have mixed nutritional benefits. Studies have found inclusion of certain species of collembolans (*Tomocerus bidentatus, Isotoma anglicana*) in the diets of spiders increases reproductive output or survival while other species (*Folsomia candida, Folsomia fimetaria, Isotoma trispinata*) drastically decrease reproductive output or result in high mortality[184–186]. These studies additionally note the possible toxicity of certain collembolans, resulting in a significant fitness cost[185]. Supplementation of Collembola and other prey not found in native forest may be a necessity for spiders in ginger-invaded sites because of the lack of abundant host-plant specific taxa; this could impart negative fitness costs. Further studies examining demographic structure of *Pagiopalus* spp. and quantifying the nutritional quality of common non-native prey items would be necessary to determine the effect.

Parasitism is one fitness cost that is identifiable using high throughput sequencing approaches[187]. Entomopathogenic fungus and hymenopteran taxa were detected in spiders collected from both ginger and native forest (Figure 5). We identified some well-known lepidopteran parasitoids that have been accidentally or purposefully introduced, including *Cotesia vestalis, Ichneumon xanthorius, Meteorus laphygmae*, and the scale parasite *Aphytis chrysomphali*; these parasitoids are found in both native forest and ginger-invaded sites. The major hymenopteran parasitoids were braconids, in particular *Microctonus* (Rhopalophorus) which is a well-known adventive species across the islands[188] and appears to parasitize beetles, notably chrysomelids[189]. A hymenopteran ASV detected in a single spider belonged to family Halticidae, the sweat bees, while all other families and genera were confirmed to be parasitic.

Infiltration of native forest by non-native parasitoids, in particular parasitoids of Lepidoptera, has been well documented[190]. Their prevalence in spiders from native forest, then, is not surprising. We do find a differing composition of parasitoid wasps in ginger-invaded habitat versus native

forest as well as increased parasite co-occurrence in spiders from ginger-invaded habitat. Additionally, the choice of percent match threshold altered our results. At an 80% threshold, 32 spiders in ginger-invaded habitat were detected with hymenopteran reads compared to 18 in native forest. At an 85% threshold, this was reduced to 17 spiders in ginger-invaded habitat and 17 spiders in native forest. This clearly demonstrates the influence of filtering decisions on the results in a study.

Entomopathogenic fungal reads were detected predominantly using 28s; while not the most widely used marker for fungal DNA barcoding, 28s has been shown to be relatively effective[166]. For the sequences detected in the spiders in our study, their high percent identity matches and known presence in Hawaii provides more support for our finding. All ASVs identified as entomopathogenic fungi produced matches above 95% and, therefore, the number of spiders found with fungi did not change as was observed in hymenopterans. Entomopathogenic fungi were detected in 22 spiders in ginger sites while only found in 1 spider from native forest. This result is consistent with previous work in New Zealand which has shown that fungivores are much more abundant in sites that have been invaded by ginger[191]. The entomopathogenic fungi *Gibellula*, the most common spider fungal pathogen[192], was detected in only one spider in native forest. Beauveria was the most commonly detected fungus in spiders from ginger-invaded sites, followed by Ophiocordyceps, a known parasite of beetle larvae[193]. *Beauveria* is a genus of cosmopolitan fungal pathogens, associated with arthropods and the surrounding habitat, including in the soil and on vegetation. It has a wide host range of over 17 arthropod orders which includes spiders[192,194], although infections have seldomly been documented[195]. In Hawaii, *B. bassiana* is used as a component of integrated pest management strategies in coffee plantations to control the coffee berry borer (*Hypothenemus hampei*), a major pest in Hawaii since 2014[196,197]. Multiple indigenous strains of *B. bassiana* are now found in Hawaii and detected in crops where there was no previous mycoinsecticide treatment[197,198].

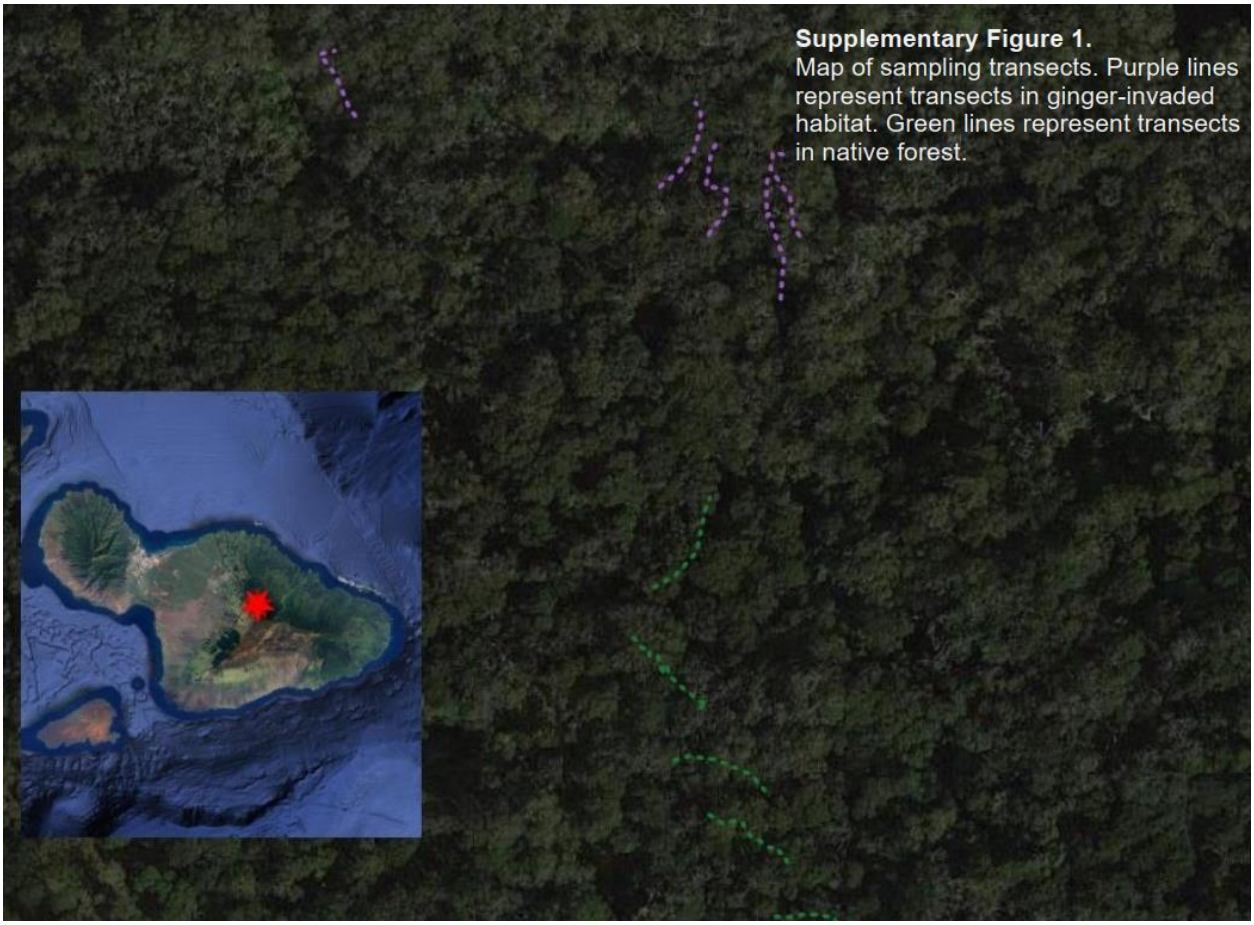
While we cannot comment on the strain or arthropod host with which *Beauveria* was carried, its prevalence, along with presence of other parasites, demonstrates a higher parasitic load found in the arthropods from ginger-invaded sites. Parasitism of prey can impart indirect effects on predators by altering prey density or prey behavior. Spiders could benefit from the secondary consumption of parasites by increasing the nutritional gains from a single prey item. While we cannot make any conclusions about the effect of higher parasitism on the spiders themselves, the higher detection of parasites in ginger-invaded habitat does demonstrate a cost for native arthropods, introducing new biotic interactions with potentially harmful taxa.

4.5. Conclusion

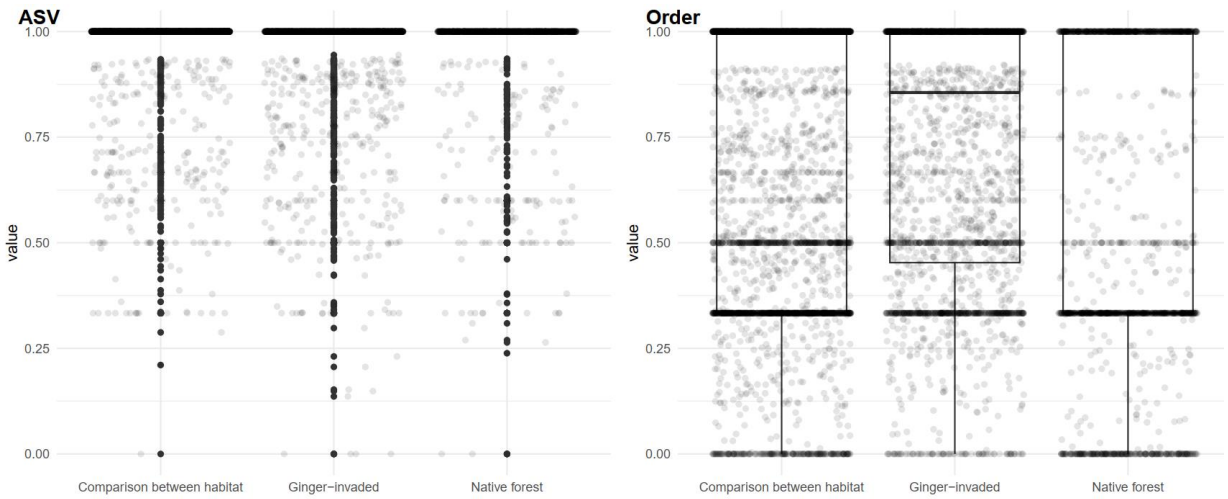
The combination of major dietary shifts driven by non-native taxa and the high prevalence of parasite reads from spiders in ginger sites indicates a prominent effect of a plant invasion on the relative proportion of non-native to native arthropods and the associated biotic interactions. The high density of non-native taxa and increases in both parasitoid wasps and entomopathogenic fungi, as indicated through the vehicle of the *Pagiopalus* spider gut, clearly demonstrate that the sites modified by plant invasion are associated with a transformation of the arthropod community. The importance of this work is in highlighting how entire communities and the

associated interactions are modified by a single invasive species that modifies the environment. Cascading effects of ecosystem alteration and the restructuring of biotic interactions may contribute to extinction debt in invaded systems, where the full consequences of invasion do not become evident for many years[199].

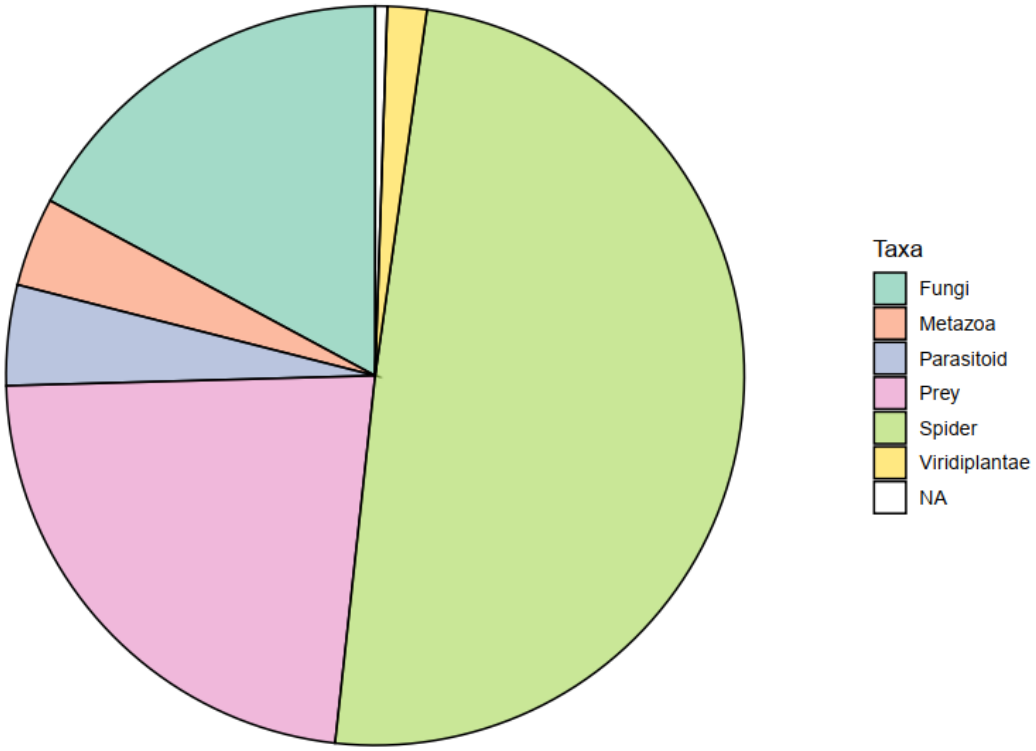
4.6. Supplementary Materials



Supplementary Materials, Figure 1: Maps of sampling locations within Hawaii



Supplementary Materials, Figure 2. Beta diversity values using ASVs versus order-level diversity.



Supplementary Materials, Figure 3. Taxonomic summary of all reads produced from sequencing. The majority of reads belonged to the predator spider itself, followed by prey items then fungi.

Marker		**Forward** 5`-3`	**Reverse** 5`-3`
16s		ATWACGCTGTTATCCCYAA	ARGACGAGAAGACCCYATA
28s		CCGTCTTGAAACACGGACCA	GWCCTCCATCAGGGTTTCCC
18s		AGCTCTTTCTYGATTCRGTGGGT, GTCTGGTTRATTCCGRTAACGAA	TTGAGCAATAACAGGTCTGTG

Supplementary Materials, Table 1. Primer pairs used in study, from Krehenwinkel et al. 2019. Two separate forward primers were utilized for 18s.

PERMDISP - ASV Hellinger

Statistic	N	Mean	St. Dev.	Min	Max
Df	2	14.000	18.385	1	27
Sum Sq	2	0.087	0.120	0.002	0.171
Mean Sq	2	0.004	0.003	0.002	0.006
F value	1	0.297		0.297	0.297
Pr(> F)	1	0.590		0.590	0.590

PERMDISP - ASV Incidence

Statistic	N	Mean	St. Dev.	Min	Max
Df	2	14.000	18.385	1	27
Sum Sq	2	0.024	0.034	0.00005	0.049
Mean Sq	2	0.001	0.001	0.00005	0.002
F value	1	0.026		0.026	0.026
Pr(> F)	1	0.873		0.873	0.873

PERMDISP - Order Hellinger

Statistic	N	Mean	St. Dev.	Min	Max
Df	2	14.000	18.385	1	27
Sum Sq	2	0.970	1.199	0.123	1.818
Mean Sq	2	0.095	0.039	0.067	0.123
F value	1	1.821		1.821	1.821
Pr(> F)	1	0.188		0.188	0.188

PERMDISP - Order Incidence

Statistic	N	Mean	St. Dev.	Min	Max
Df	2	14.000	18.385	1	27
Sum Sq	2	0.476	0.649	0.017	0.936
Mean Sq	2	0.026	0.012	0.017	0.035
F value	1	0.494		0.494	0.494
Pr(> F)	1	0.488		0.488	0.488

Supplementary Materials, Table 2. PERMDISP values using ASVs and order level identities, both incorporating read data by using Hellinger transformed reads, and using only incidence values.

In addition to deforestation discussed in **CHAPTER 3** and species invasion discussed in **CHAPTER 4**, changing disturbance regimes are an increasing threat to global biodiversity. In my final chapter, **CHAPTER 5**, I focus on arthropod community response to changing wildfire regimes across California and the importance of historical legacies associated with habitat types in dictating response

Chapter 5: The importance of habitat type and historical fire regimes in arthropod community response following large-scale wildfires

Anna J. Holmquist, Robert C. Markelz, Cierra C. Martinez, Rosemary G. Gillespie

5.1. Introduction

Disturbance is integral to the health and productivity of many ecosystems. The intermediate disturbance hypothesis (IDH) encapsulates this, in which diversity is predicted to be highest under intermediate levels of disturbance[200,201]. However, with climate change comes an increasing number of severe and devastating natural disasters that are shifting established disturbance regimes[202] and leading to rapid shifts in ecosystems[203,204]. Wildfires are one such example; while many ecosystems are adapted to cyclical fire regimes that have been maintained over millions of years[205,206], specific aspects of wildfire regimes are changing due to shifts in climate as well as increases in anthropogenic ignition[205,207,208]. In the Western USA, both fire frequency, area, and severity have seen sharp increases since the 1980s[209]. With fire activity predicted to continue intensifying alongside climate change, effective management requires that we understand how the regime shifts are impacting biotic communities[210].

Many habitat types and their associated biota rely on fire to maintain ecosystem stability. Historical regimes generally featured low severity fires at consistent intervals that generated a heterogeneous landscape; the resulting variability in resources and vegetation structure can create new niche space that increases species diversity[211]. Natural wildfire regimes produce a range of supporting, provisioning, regulating, and cultural ecosystem services for humans[202]. However, modern regimes featuring frequent and high intensity fires are driving forest conversion even in habitat types accustomed to fire such as in conifer-dominated forests[212] and pine-oak woodlands[213]. Intense wildfires can not only kill large expanses of trees but additionally alter soil quality, reduce seed reserves, and facilitate the establishment of non-native plant species. Such catastrophic wildfires result in changes to ecosystem structure and can cause the loss of important ecosystem services such as erosion control and carbon sequestration[214].

The effects of fire on local fauna are also influenced by severity. Plant-pollinator diversity has been shown to benefit from natural wildfire, with the highest richness of plants and insect pollinators found in sites that experienced low to moderate severity wildfires; however, this trend is lost in areas of high-severity burn[215]. A recent meta-analysis found a strong effect of high-severity fires on arthropod communities as a whole, consistently resulting in a reduction in richness and evenness[216]. Another meta-analysis found increases in pollinator abundance and richness after fire, but found short fire intervals to be detrimental to communities[217]. Both papers emphasize the importance of dynamics of the fire regime, such as fire intensity and burn interval, rather than burn itself, on the arthropod community. Habitat type (e.g. grasslands, desert, forest) has been identified as an additionally important factor in dictating the response of arthropod communities to fire[216] although most studies thus far have focused on forested areas. The response of vegetation and the rest of the biotic community is largely dependent on habitat type

and historical fire adaptation. While habitats adapted to fire can still be severely impacted by high-intensity fires, habitat types that rarely experience fire are at higher risk of reaching ecosystem tipping points following burn[214,218] Understanding broad patterns of community response following wildfire therefore requires assessment across diverse habitat types.

Because of the tight association between many insect species and the floral community as well as their importance in bottom-up structuring of the trophic network, insects can serve as excellent bioindicators of local ecosystem state following disturbance. Moreover, insects provide an opportunity to assess whole-community response through their ease of capture, providing a more generalizable picture of the effect of disturbance than single-taxon studies. Our study aims to address the differences in arthropod response to wildfire across diverse habitat types with different historical exposure to fire. Because of their importance in supporting plant succession as well as reliance on floral communities, we targeted pollinator communities using yellow pan traps and collected communities across burned and unburned sites in fall 2020 and spring 2021. We utilized protected natural areas that are part of the University of California Natural Reserve System (UCNRS), many of which were at least partially burned during the 2020 August Lightning Complex Fire. This fire system caused over 650 wildfires across the state, and burned 1,500,000 - 2,100,000 acres of land. It is one of many recent large-scale fires that have set records for California wildfire size[219]. The UC reserves, distributed across most of the major California ecoregions, offer a unique opportunity to examine fire effects on arthropods across a broad spatial scale in distinct habitat types that differ in historical fire regimes. The seven habitat types in our study are grassland, scrub, oak woodland, forest, redwood, chamise and serpentine. Each of these habitat types has a different pre-settlement fire interval estimate in years: grassland (<10), oak woodland (12), redwood (23), mixed pine forest (23), and scrub (60)[220]. This allowed us to specifically explore how differences in fire frequency by habitat type alter community response.

To assess specifically how the structure of communities changed, we adopted a metabarcoding approach to sequence all individuals in a given community. Metabarcoding techniques allow rapid classification of biotic communities without the need for taxonomic identifications. Adoption of a non-destructive extraction approach additionally allows *post-hoc* assessment of abundances and verification of results. We targeted pollinator communities specifically because of their importance in supporting plant succession as well as their reliance on floral communities. To do this, we used yellow pan traps to sample the across burned and unburned sites in fall 2020 and spring 2021 following the major Lightning Complex wildfires in California. We used these data to test a) how species richness values vary between burned and unburned sites, b) the extent of compositional differences between burned and unburned sites and c) the role of habitats with different historical fire regimes in shaping community response. Our specific hypotheses were:

1) Because multiple taxa capitalize on burn, alpha diversity will not differ by burn status. Differences in alpha diversity will instead be associated with season due changes in plant activity and the associated emergence of pollinators.

2) Community recovery is dependent on re-colonization, which is in part driven by dispersal capabilities. Because we are targeting pollinators, most of which are winged taxa and dispersive, we expect no strong compositional dissimilarity between habitat types across reserves. Compositional differences will be driven by season, habitat type, and burn status.

3) The extent of dissimilarity between burned and unburned sites will differ by habitat type, with habitat that rapidly recovers (grasslands) or habitat well-adapted to fire (oak woodlands) showing recovery by the spring, while habitat less accustomed to severe wildfire (scrub) will remain dissimilar when compared to unburned sites.

5.2. Methods

5.2.1. Field sites

The UC Natural Reserve System comprises 47,000 acres distributed across most of the major California ecosystems. Between August 15-16, 2020, an exceptionally fierce electrical storm crossed northern California, peaking at 200 lightning strikes in a 30-minute period. Over the following 72–96 hours, more than 12,000 lightning strikes sparked multiple wildfires around the Greater Bay Area. Many of the smaller fires merged into four large fire complexes, including two in Monterey County. Three of these fire complexes are considered to be the second, third, and fourth largest fires in California history[219]. These fires burned eight of the UC Natural Reserve System's 41 reserves, and consumed more than 20,000 acres of reserve lands. Affected NRS reserves include: Año Nuevo Island Reserve (CZU Lightning Complex), Blue Oak Ranch Reserve (SCU Lightning Complex), Landels-Hill Big Creek Reserve (Dolan Fire), Hastings Natural History Reservation (River Fire), McLaughlin Natural Reserve (LNU Lightning Complex), Point Reyes Field Station (Woodward Fire), Quail Ridge Reserve (LNU Lightning Complex), and Stebbins Cold Canyon Reserve (LNU Lightning Complex). A UC Davis property affiliated with Stebbins Cold Canyon, the Cahill Reserve, also burned extensively due to the LNU Lightning Complex. The current study sampled in six burned reserves: Quail Ridge, McLaughlin, Hastings, Blue Oak, Big Creek, and Ano Nuevo (Figure 1).

5.2.2. Field collection

A joint effort between (UCSC and UC Berkeley), funded by the University of California office of the president, was undertaken to sample plant and arthropod communities across the six burned reserves. At each reserve, burned and unburned sites were selected within 8 habitat types: oak woodland, redwood, scrub, chamise, grassland, forest, and serpentine. Arthropods were collected at each site using yellow pan traps filled with water and added soap to break surface tension. Traps were moved to whirlpaks following the sampling period. Samples were sorted and stored at -20°F until further processing.

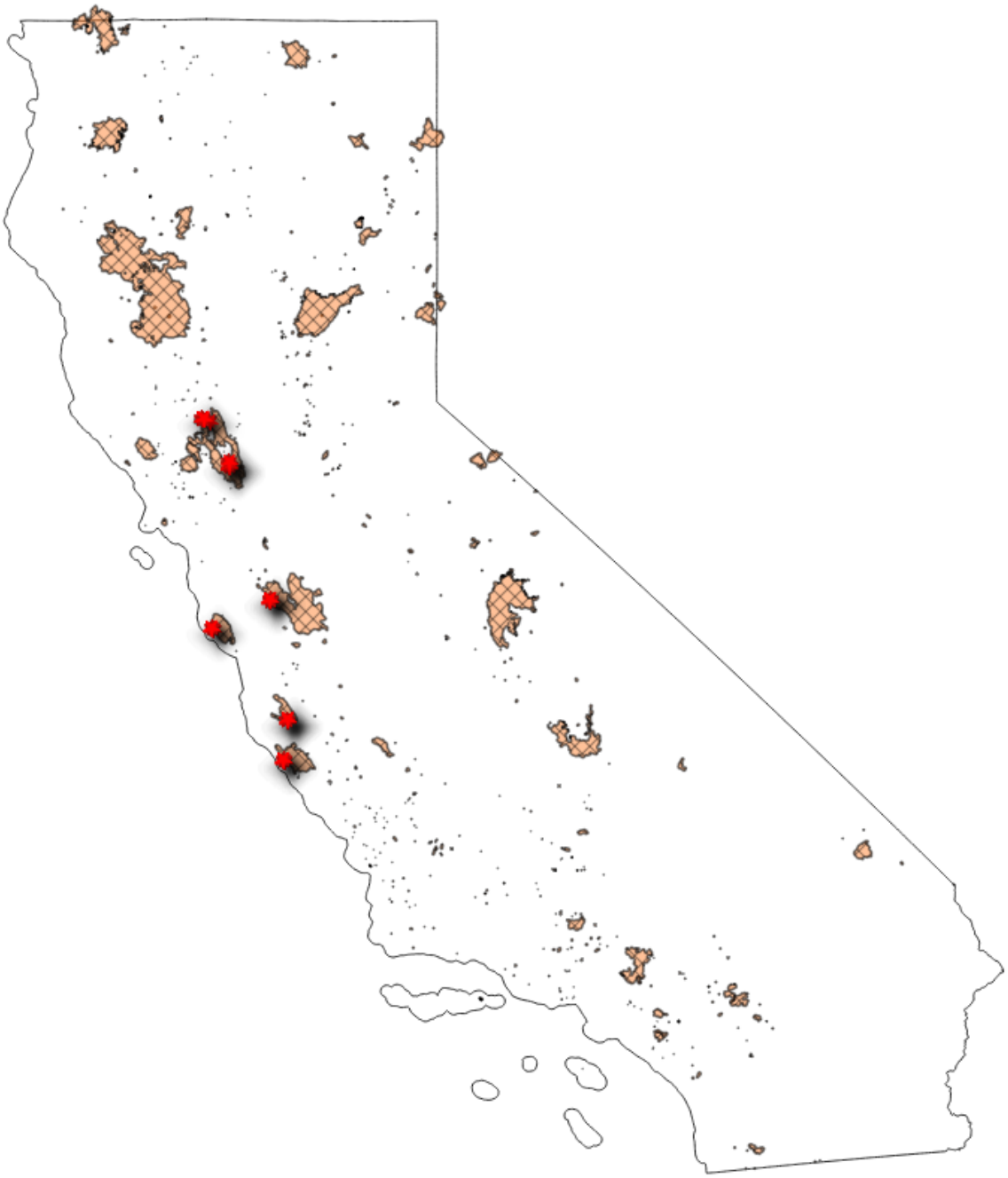


Figure 1. Map showing the California Lightening Complex fires of 2020, as well as points showing the areas sampled in our study.

5.2.3. Molecular procedures

Samples were counted then subdivided into hard-bodied and soft-bodied arthropods to allow varied lysing time. Once sorted, samples were stored in 95% EtOH in 96-well plates at -20°. A BenchSmart 96 (Mettler-Toledo Rainin, Colombus, OH USA) was used to remove EtOH after spinning plates down briefly. Remaining EtOH was evaporated using a vacuum centrifuge, and checked at regular intervals until fully evaporated. Extraction protocols were modified from the Qiagen PureGene extraction kit (Qiagen, Hilden, Germany). Hard-bodied insects were incubated for 24 hours at 55° in cell lysis buffer and Proteinase K while soft-bodied insects were incubated for 3 hours. Following incubation, lysate was transferred to a second plate and RNAse added, followed by an additional incubation period at 37°C for 30 minutes. Plates were then placed on ice and, once cool, Protein Precipitation Solution was added. Plates were spun for 10 minutes at 2000rcf to form protein pellets, then supernatant was transferred to a solution of isopropanol and glycogen. Plates containing specimens were washed repeatedly to remove lysate and stored in 70% EtOH for future taxonomic work. Samples were incubated at room temperature overnight then spun down to form DNA pellets. Isopropanol was removed and pellets were washed with cold 70% EtOH. Ethanol was removed and, once remaining ethanol evaporated, elution buffer was added and then plates incubated at 65° for 1 hour. Plates were then aliquoted and stored in -20° until further use.

Amplification of COI was performed using the Qiagen multiplex kit (Qiagen, Hilden, Germany). Two sets of primers were used - MCO and BF3 (Table 1). Primers additionally had a 5` TruSeq tail for indexing. Amplification was performed on 96-well plates and consisted of 5ul of the Qiagen PCR MasterMix (MM) (Qiagen, Hilden, Germany)., 3ul of H20, 0.5ul of each primer, and 1ul of template DNA. Three PCR replicates were performed. Annealing temperature was 46° and ran for 30 cycles. A negative PCR control was included on each plate, consisting of MM, H20, and primers. PCR products were visualized using gel electrophoresis on a 3% agarose gel. A dual indexing strategy was implemented using a second round of PCR to attach 8bp indexes. Annealing temperature for indexing PCR was 55° and ran for 6 cycles. PCR products were visualized once again using a 3% agarose gel and ran at low voltage to allow clear visualization of length to confirm the addition of indexes to each amplicon. Final libraries were constructed by pooling PCR products proportionally based upon band strength. Further processing was conducted at Berkeley QB3 Genomics (QB3, Berkeley, CA USA).

Forward	Sequence 5′-3′	Reverse	Sequence 5′-3′
mlCOIintF[121]	GGWACWGGWTGAACWGTWTAYCCYCC	Fol-degen-rev[83]	TANACYTCNGGRTGNCCRAARAAYCA
BF3[123]	CCHGAYATRGCHTTYCCHCG	BR2[123]	GCHCCHGAYATRGCHTTYCC

Table 1. Primer pairs used to amplify COI.

5.2.4. Bioinformatic processing

Reads were trimmed of primer adapters using CutAdapt[38]. Reads were processed using the DADA2 denoising algorithm[39] to form amplicon sequence variants (ASVs). The removeBimeraDenovo function included in the DADA2 function was used to remove chimeras. The BF3 primer pair produced low quality sequences with over half being discarded through normal filtering procedures, and was not included in further analysis. The PCR negative controls were utilized for decontamination, using the package decontam [163,164]. Following

decontamination, the curation algorithm LULU[41] was used to remove erroneous sequences. Sequences were aligned in Geneious Primer Prime v.2022.0.2 to detect any remaining pseudogenes, identified by stop codons or shifts in the reading frame. Following removal, BLAST was performed using megablast. A custom script in R using package rentrez[167] was then used to generate full lineage information from accession numbers. Only phylum Arthropoda was retained, and sequences with below 80% percent identity matches were also removed. The final filtering step retained only ASVs found in two of three replicates. Operational taxonomic units (OTUs) were generated using *otu* in package kmer[43], clustered at 97% with kmer = 5. This cluster threshold was used to examine taxa at a coarser level, closer to genus than species.

5.2.5. Statistical analysis

Quail Ridge had no unburned sites and was excluded from analysis. There were a low number of sites in redwood (2 burned and 1 unburned), chamise (3 burned and 2 unburned), forest (2 burned and 2 unburned), and serpentine (2 burned and 2 unburned); scrub, grassland and oak woodland habitat types were best sampled and the data was narrowed to include only these habitat types. As explained in chapter 3, the BF3/BR2 primer pair was found to generate high errors and so the amplicons produced from the MCO primer pair was used (Table 1). Due to this, we chose to use only the data produced from the widely used MCO primer set. Reads were summed for each ASV detected across replicates for a particular sample and a community matrix was created. Because reads do not translate directly to abundance count, read data was transformed using Hellinger standardization; this method has been widely adopted in metabarcoding. Standardization was performed using *decostand* in vegan[46].

We used OTU richness, the Shannon index and the Simpson index to calculate alpha diversity metrics, with the standardized read matrix as input. F-tests were performed to compare variances; the non-parametric Mann-Whitney U test was used to assess median differences between alpha diversity of burned and unburned sites. For compositional comparisons, sites with three or more OTUs were used. All analyses were performed using both incidence and Hellinger-standardized read data. Nonmetric multidimensional scaling was performed using Hellinger distance, by applying Euclidean distance to the Hellinger-standardized data; this was done using metaMDS in vegan with up to 1,000 random starts and 3 dimensions. The argument *noshare* was utilized which creates extended dissimilarities for sites that share no species. Outlier samples with entirely distinct communities prevented convergence and were removed from NMDS (Supplementary Materials, Figure 2), but all statistical tests were run using the data with and without the outlier data and are included in the supplementary materials (Supplemenatry Materials, Table 1 & Table 2). Permutational multivariate analysis of variation (PERMANOVA) was used to test group differences based on burn, habitat type, reserve and year using the Hellinger distances. This was done using *adonis2* in vegan[46]. Differences in group dispersions were tested using PERMDISP2 with function *betadisper*. Additionally, we split the data by habitat type, and performed NMDS, PERMANOVA and PERMDISP2 for each individual habitat type.

To calculate compositional differences in an additional way, we used the *beta* function in the package BAT[49], which partitions beta diversity into differences driven by species replacement and by richness differences. We performed the non-parametric Kruskal-Wallis test to assess if burned and unburned sites had higher beta diversity values than sites within burned habitat and sites within unburned habitat. We compared differences in beta diversity against distances

between sites within reserves to examine spatial autocorrelation and the key components driving dissimilarities. Spatial distance between sites was calculated using QGIS v.3.22.4 and the standard setting under distance matrix calculation. We performed linear regression to assess the association with distance. Using ANOVA, we tested the difference between beta diversity values across habitat types.

Data processing and statistical analysis was performed in the R programming environment (R Core Team, 2022). Figures were created using packages ggplot, and venn.

5.3. Results

5.3.1. Data summary

6,461,595 reads were produced after primer removal and initial filtering by CutAdapt. Filtering, clustering and chimeral removal with DADA2 produced 5,215,076 reads. After implementing LULU, removing potential contaminants using decontam, removing NuMts, selecting only phylum Arthropoda, and removing ASVs not found in two of the three replicates for a sample, the dataset consisted of 2,167,382 reads and 584 ASVs, clustered into 448 OTUs. Taxonomic identities were assigned based on percent identity; 12 orders (> 85%), 33 families (> 92%), 94 genera (> 97%) and 51 species (>99%) were identified using BLAST.

The most common order by OTU in the fall was Diptera and in the spring was Hymenoptera (Figure 2). 17 families were shared across burned and unburned sites. There were 9 unique families detected in unburned sites and 6 unique families detected in burned sites. Two arachnid families were detected in burned sites (Acariformes: Erythraeidae and Araneae: Lycosidae). In addition, two unique families of dipterans (Chironomidae and Syrphidae), one unique family of lepidopterans (Lycaenidae) and one unique family of hymenopterans (Megaspilidae) were identified in burned sites. Four families of flies, two families of wasps, one family of butterflies, one family of beetles and one family of spiders were the distinct families in unburned habitat. However, only 133 of 584 OTUs were confidently identified to family with percent identity matches above 95.

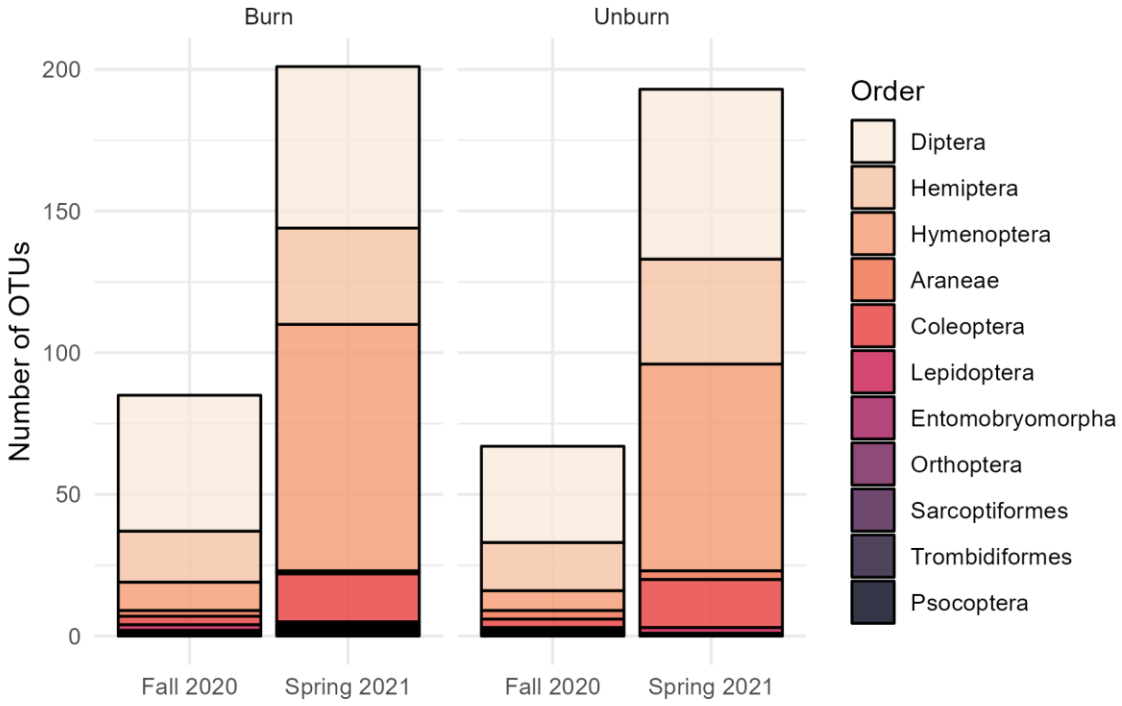


Figure 2. Box plots of burned and unburned sites, showing the number of OTUs belonging to each order in Fall 2020 and Spring 2021. Composition was similar across burned and unburned sites. There were major increases in Hymenoptera OTUs in the spring.

5.3.2. Alpha diversity

Variances between burned and unburned sites were not significantly different for any of the calculated alpha diversity values. Values in alpha diversity were not significantly different between burned sites and unburned sites (p-value = 0.315 *richness*; 0.252 *Shannon*; 0.217 *Simpson).* There were significant differences by season, with an increase in richness in the spring (p-value = 0.0006, *richness*; 0.0004, *Shannon*; 0.0009, *Simpson).* All habitats saw increases in richness in the spring (Figure 3).

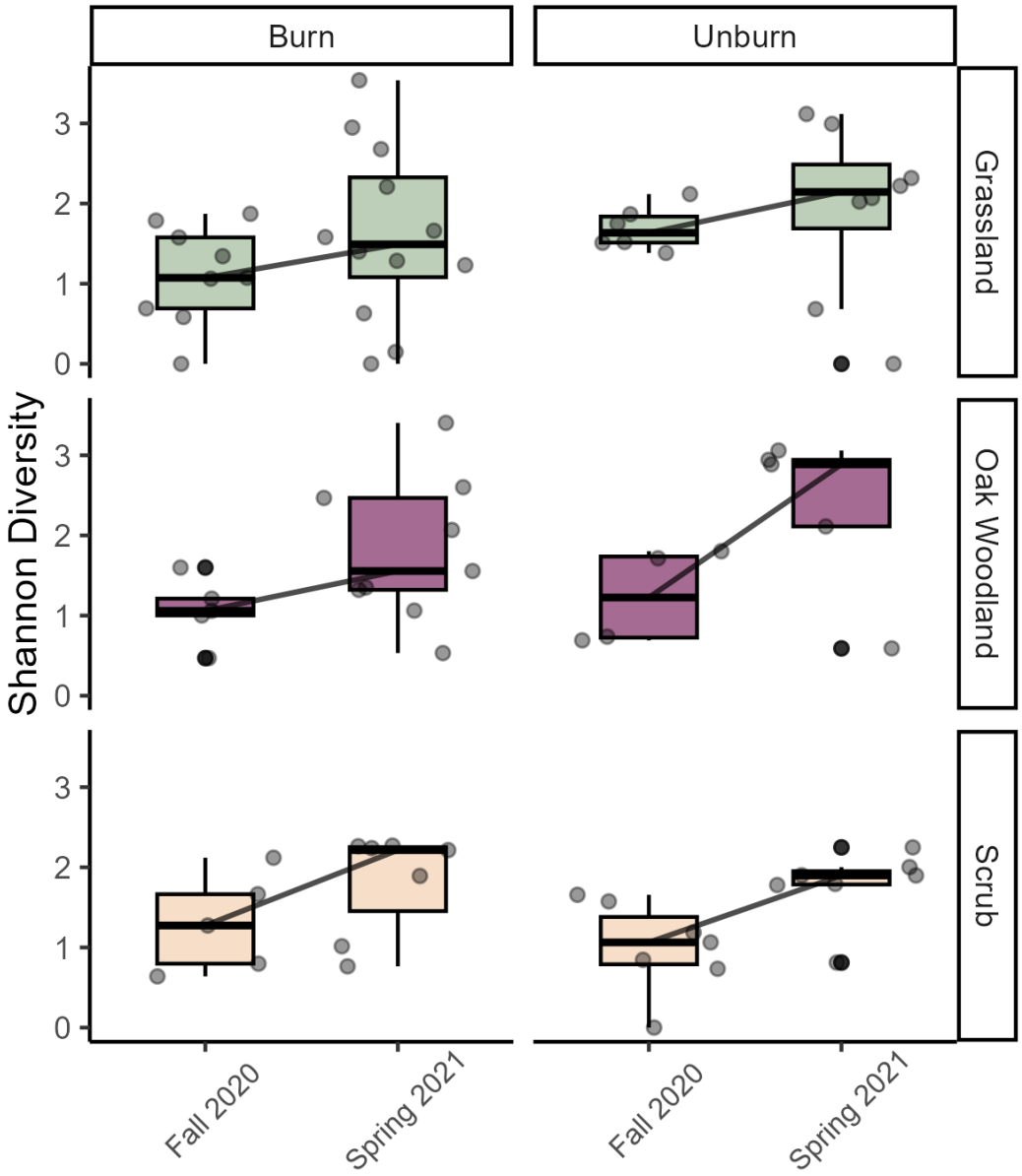


Figure 3. Differences in richness across fall and spring for three different habitat types (grassland top, oak woodland middle and scrub bottom) in burned sites (left) and unburned sites (right).

5.3.3. Predictors of community compositional differences

Results were similar between Hellinger-standardized read data and Jaccard distances. Because of this, results presented are from Hellinger-standardized data; results from incidence data are included in the supplementary materials. Two outlier samples were removed to allow convergence of nMDS; the outliers were both included and excluded from statistical tests and did not change results (Supplementary Materials, Figure 1, Table 2 & Table 2).

There were not significant dispersion differences between burned and unburned sites (F = 0.355, Pr = 0.554; Figure 4). Multiple predictors and interactions between predictors were significant. Reserve was the strongest explanatory variable of group variation (R2 = 0.0925, F = 1.98, Pr < 0.001) followed by an interaction between reserve and season (R2 = 0.0799, F = 1.71, Pr < 0.001). There was a significant interaction between burn status and reserve as well (R2 = 0.0587, F = 1.26, Pr < 0.001) followed by an interaction between habitat and reserve (R2 = 0.0572, F = 1.22, Pr < 0.001). Season individually explained less variation in the response (R2 = 0.034) but was the strongest predictor of group differences based on the F-statistic (F = 2.895, Pr < 0.001; Figure 5). Burn status was not a significant explanatory variable by itself, and only showed significance when included in interactions.

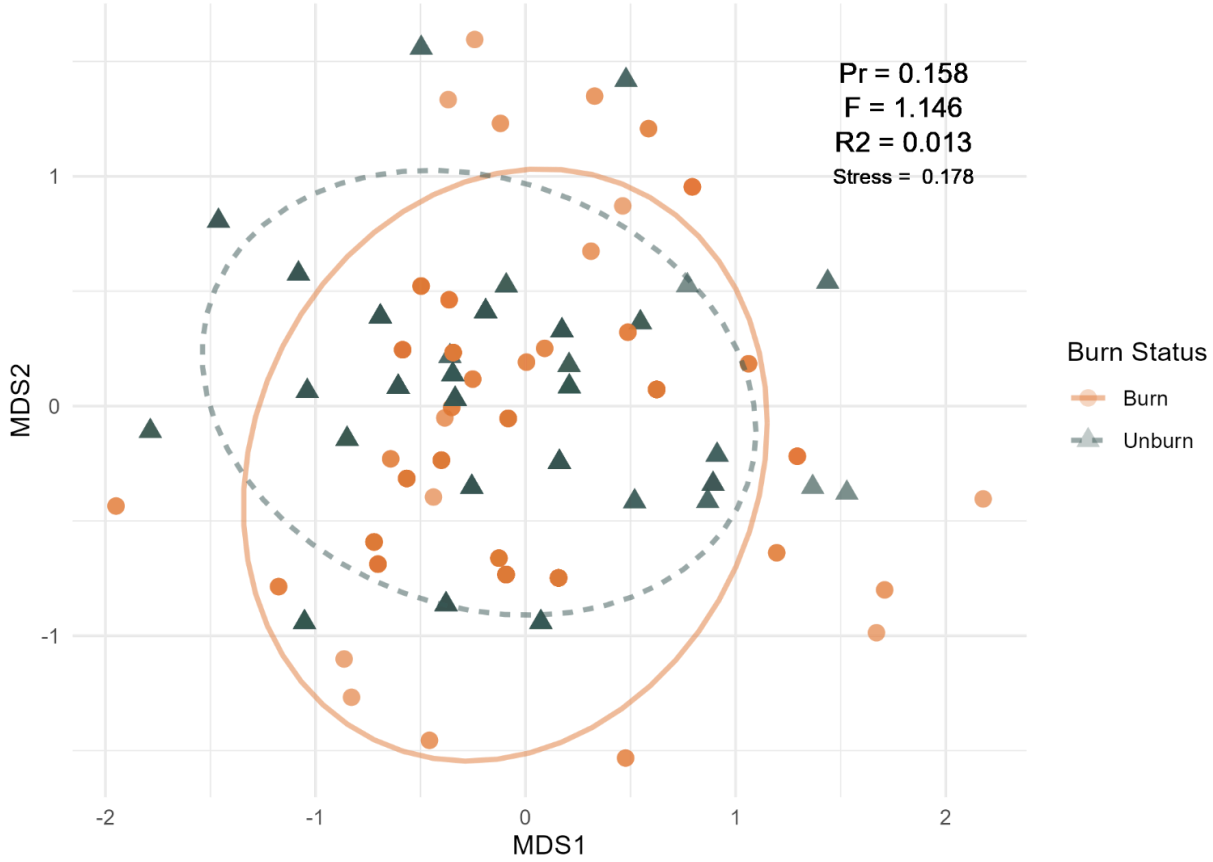


Figure 4. NMDS plot showing similarity in sites, colored by burn status with burned sites in orange and unburned sites in green. There was no significant difference by group.

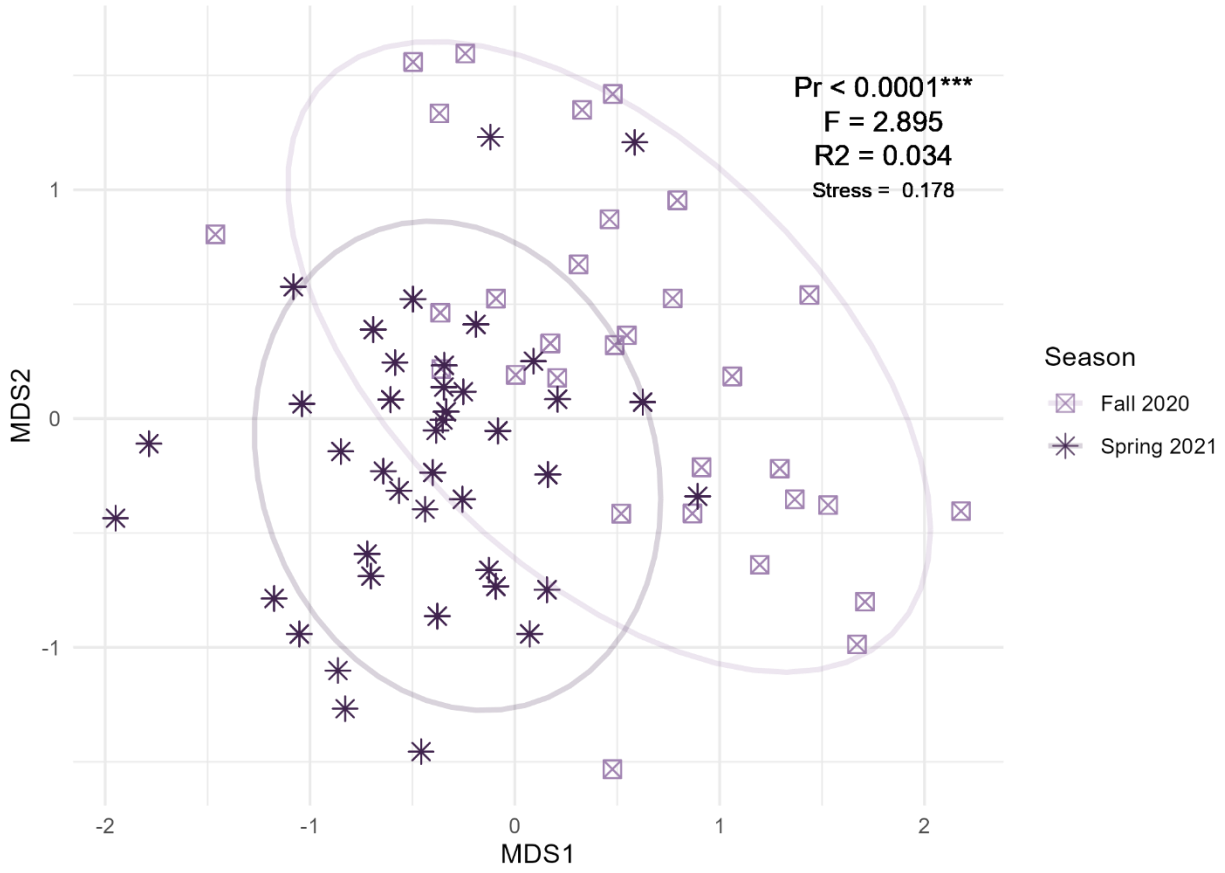


Figure 5. NMDS plot showing similarity in sites, colored by season with fall sites in light purple and spring sites in dark purple. There was a significant difference by season.

The median beta diversity value was 0.964 between burned and unburned sites found within the same reserve, habitat type, and during the same season. When comparing sites within unburned habitat, the median beta diversity score was 0.908 and when comparing sites within burned habitat, the median beta diversity score was 0.924. The differences between comparisons, however, were not significant. The major component of beta diversity was species replacement with a mean of 0.769 of the total beta diversity values. In fall 2020, there was a slightly significant decrease in beta diversity across distance (Adjusted R2 = 0.140, t = -2.285, Pr = 0.031), driven by replacement (Adjusted R2 = 0.133, t = -2.236, Pr = 0.035) (Figure 6). This trend was lost in spring 2021; instead, there was a slightly significant association with the richness component of beta diversity across distance (Adjusted R2 = 0.077, t = -2.024, Pr = 0.051), in which sites closer to one another differ compositionally in a larger part from differences in species richness, rather than only species differences. This decreased over distance, with the farther sites having beta diversity values largely composed of species replacement. Visually, there is a decrease in replacement over distance, although not significant likely due to a few intermediate-distance sites with a richness component (Figure 6).

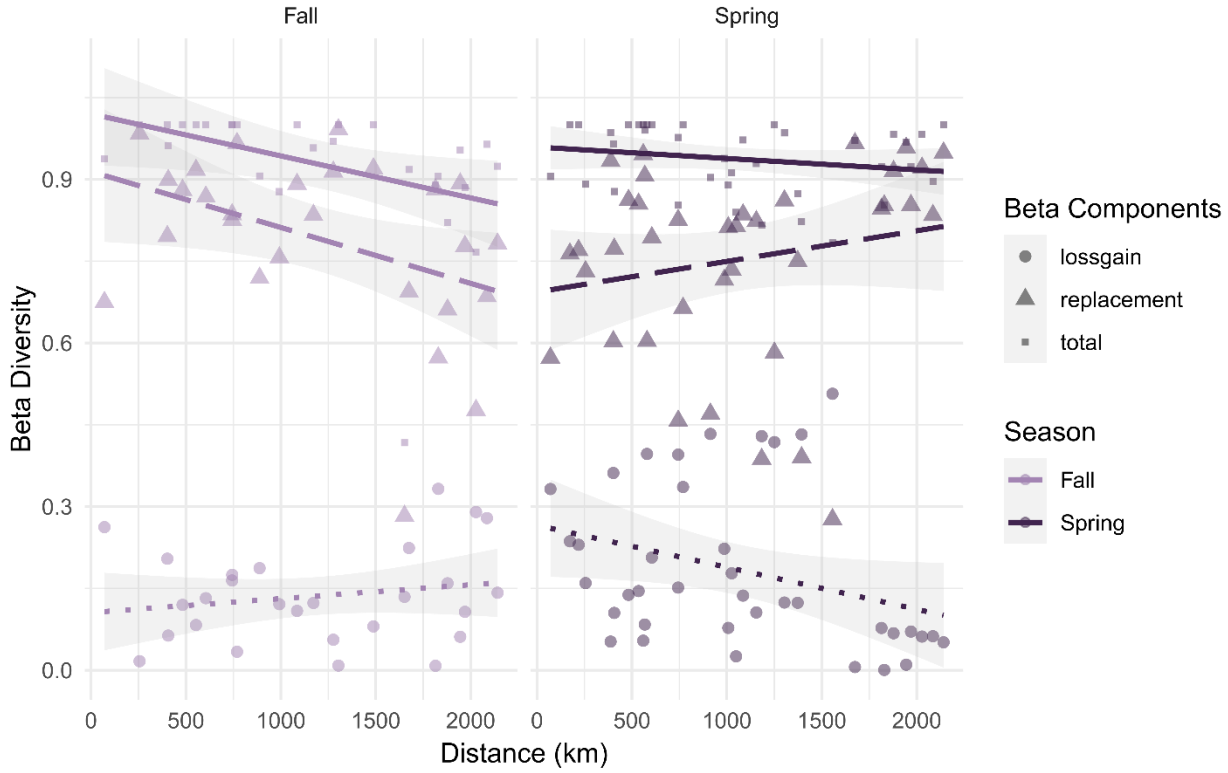


Figure 6. In the fall (left), beta diversity values decreased with distance, largely due to decreases in the replacement component. This differed in the spring (right), where the importance of replacement increased with distance.

5.3.4. Response by habitat type

Beta diversity between burned and unburned sites within habitat types did not vary, nor did they significantly change from fall to spring. In grassland, burn status was not explanatory in either season (PERMANOVA). However, there were significant dispersion differences in the spring season (Pr = 0.006), which is also noticeable in the ordination plot (Figure 7B). Oak woodland sites showed a similar pattern, without significant differences between groups in either year. This is seen in ordination space by interspersing of points. Unlike grassland sites, no dispersion differences were detected in oak woodlands; sites show a similar spread in ordination space (Figure 7A, B). In scrub habitat, burn was not explanatory in the fall but was a significant explanatory variable in the spring (R2 = 0.103, F = 1.38, Pr = 0.029; Figure 7E, F). Points separate clearly in ordination space into two groups, especially noticeable in the spring. There were no detectable dispersion differences in either season in scrub habitat. In scrub, the lowest median beta diversity, or the highest compositional similarity, was between burned habitats (median = 0.881) while the highest median beta diversity score was comparing burned and unburned sites (median = 0.972). This contrasted the other habitat types, with which the highest median beta diversity score was comparing between the burned sites (median = 0.833 *grassland*; median = 0.990 *oak woodlands*) and the lowest median beta diversity was comparing between unburned sites (median = 0.958 *grassland*; median = 0.918 *oak woodlands*) (Figure 6).

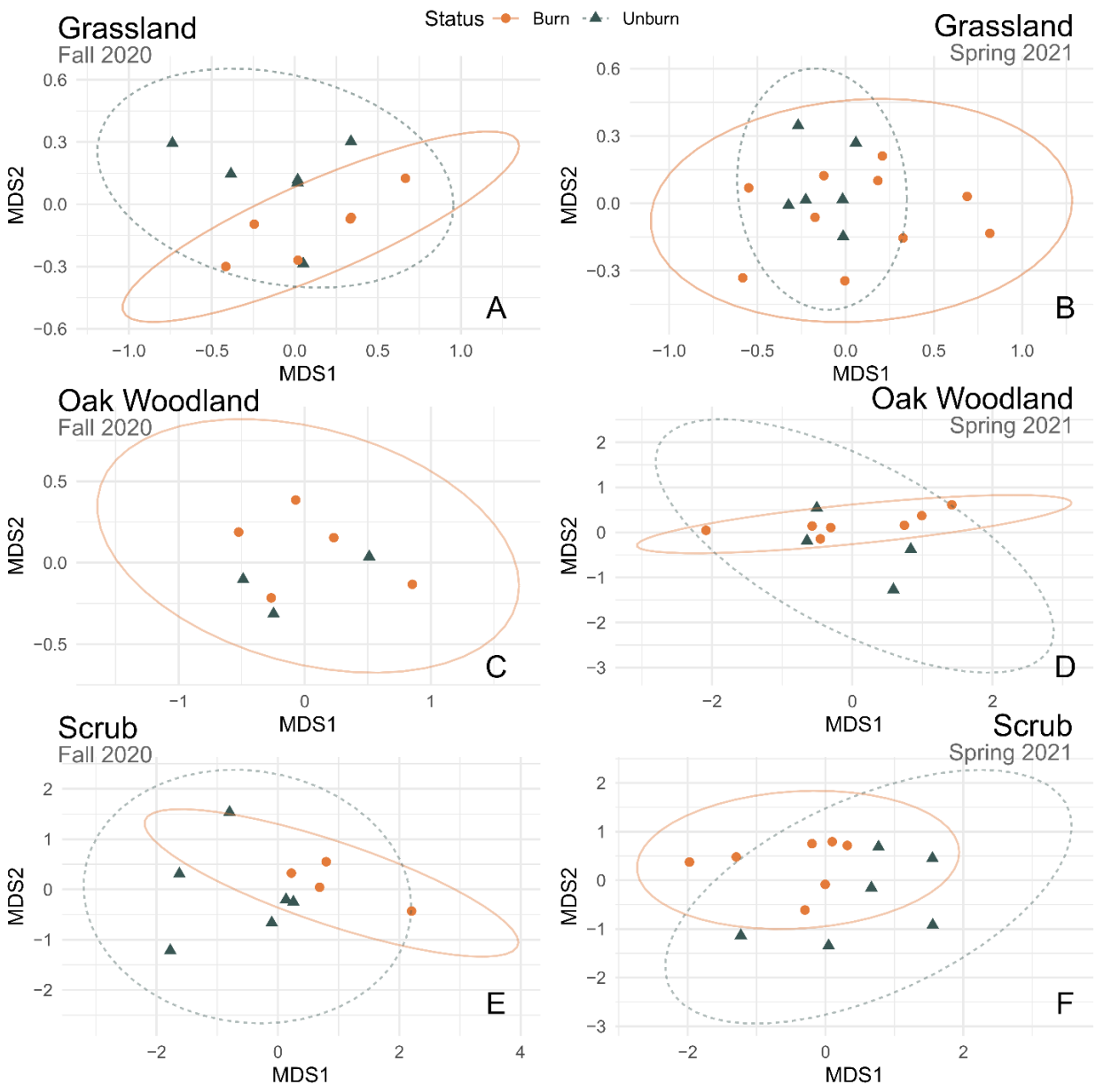


Figure 7. NMDS plots by habitat, with the top plots (A, B) showing grasslands, the middle plots (C, D) showing oak woodland and the bottom plots (E, F) showing scrub. Grassland sites became more similar in spring (B), although burned sites showed significantly different dispersion. Oak woodlands were similar with burned and unburned clustering together in both season (C, D). Significant differences existed between community compositions of burned and unburned sites in the scrub habitat in spring (F).

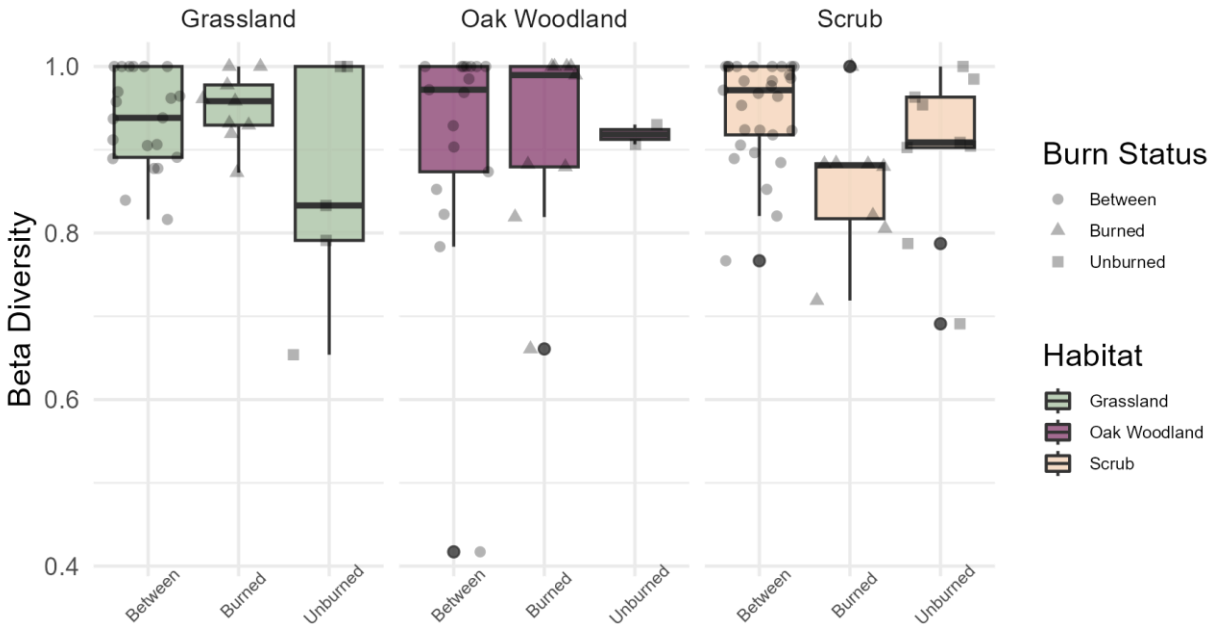


Figure 8. Beta diversity values split by comparisons between burned and unburned sites, comparisons between sites only within burned habitat and comparison between sites only in unburned habitat. In scrub habitat, burned sites showed the lowest dissimilarity. There were also comparisons between sites in unburned habitats that had low dissimilarity scores, seen in both grasslands and scrub.

5.4. Discussion

In this study, we utilized metabarcoding of whole arthropod communities collected from burned and unburned sites to assess the effect of large-scale wildfire across habitat types. Cyclic wildfires can foster biodiversity and many studies detect similar or higher species richness values in burned sites compared to unburned[221,222]. Our results align with this common finding, with no significant difference in alpha diversity values between burned and unburned sites. Instead, there were strong seasonal differences. This supports our first hypothesis, in which alpha diversity values would not show differences by burn but would show differences by season. Certain arthropod species are pyrophilous, drawn to burned habitat by smoke or heat[223]. Immediate colonization during primary succession usually occurs by these fire-associated species followed by colonization by taxa that benefit indirectly from fire effects[224,225]. The emergence of plant species in the spring, be it species commonly detected in the original ecosystem pre-burn or new species that capitalize on the burn, increases pollinator richness at all sites.

While maintenance of biodiversity itself is important, patterns of species replacement and reordering are essential to explore if we are to understand how ecosystems change under novel disturbance. Burn status did not emerge as a significant grouping factor when comparing across all sites. Instead, we detected a highly significant difference in community composition based on reserve. Distance is then the most crucial factor in maintaining high diversity, with each reserve hosting a distinct pool of arthropods. This is true despite most taxa we collected being winged and presumably more capable of dispersal. Our second hypothesis predicted high connectivity across reserves within similar habitat types, but this finding counters our hypothesis. The high spatial turnover in diversity between reserves speaks to California's classification as a

biodiversity hotspot[3]. California is known for its incredibly diverse flora with around 6,900 native plants, most of which are endemic[226]. Because insect diversity is often correlated with plant diversity[227], we can expect that California has a high abundance of endemic insects with the distributions of many species' closely associated with different plants. Protection of habitat more broadly is therefore essential in maintaining biodiversity.

To assess the effect of burn exclusively, we compared the compositions of sites within the same reserve, habitat type and season. Burn status was not a significant grouping factor itself when comparing all sites, but when variability due to reserve differences was taken into account with an interaction term, burn status did emerge as an explanation of variation. Within reserves and the same habitat types, distance was significantly associated with beta diversity differences in the spring. Even on such a small scale (less than 3000 meters), dispersal seems to be a limiting factor in recovery. The closer sites share more species-like groupings, with differences driven by richness, while the farther sites have beta diversity almost entirely driven by replacement. Interestingly, this trend is reversed in the fall. This may be explainable by the reduced alpha diversity in general, with taxa that remain active in the fall being detected more broadly. The most abundant order in the fall was Diptera (flies), which are less likely reliant on plant communities. Additionally, flies could potentially be more dispersive than Hymenoptera (ants, bees and wasps), which was the dominant order in the spring and function more as the primary pollinators. The increased dissimilarity between sites in the spring could indicate that the successional trajectory is moving away from pre-fire conditions rather than towards. Long-lasting compositional differences are noted in other systems[228]; these major changes in species composition can lead to an overall functionally distinct community[229,230]. In places such as California with highly endemic taxa restricted to certain habitat types and demonstrating high spatial turnover, conversion of areas to novel ecosystems is a serious threat to maintenance of biodiversity.

The tight association between insects and flora makes consideration of habitat type crucial when measuring compositional response. We found a significant interaction between habitat and reserve in explaining compositional differences, emphasizing the importance of habitat type in biotic community structure. More dispersive and generalist insects may be favored following fire as they can capitalize on a novel plant community and colonize quickly[224]. Depending on dynamics of the wildfire, native vegetation can re-establish and flourish or instead be entirely missing from the plant community. Wildfires in forested areas can create a heterogeneous landscape with open areas that promote high flowering plant density and an associated increase in pollinators[231]. Alternatively, frequent fires can cause invasion of non-native plants, particularly grasses, that can displace native species[232,233]. Insects with particular specializations on endemic plant flora may not recover due to loss of host plants or competitive exclusion by opportunistic species that establish following fire.

The way a habitat will respond is driven both by the historical adaptation to fire and the particular dynamics of the wildfire. Grasslands are one habitat type that quickly flourishes after burn, regardless of severity. Within grasslands, non-native species of grasses are noted to become more prevalent following large scale disturbance. Our results show that arthropod communities found in burned grassland sites resemble unburned sites by the following? spring. This has been found in other groups, such as spiders, which show a similarly rapid recovery time-frame in grasslands following disturbance[229] . When comparing compositions within the burned sites, we see higher dissimilarities than observed between unburned sites. This may

reflect the higher richness found in some of the burned grassland sites, and possibly reflects the benefits fire can have on community diversity in certain habitat types.

Oak woodlands are another Californian habitat type reliant on regular wildfires in short intervals, which reduce density of conifer competitors and other fire intolerant species. Historical fire suppression caused major compositional changes to oak woodland habitat but prescribed burn strategies in more recent years have allowed these habitats to flourish once more. Oak woodlands are thought to be resilient even in the face of high severity burns, although this may impact the long-term development of the trees[234]. Our results in oak woodland sites showed no strong explanatory nature of burned and unburned in either fall or spring. This aligns with the long history of fire in these habitat types, consisting of communities that are well-adapted and resilient to fire. Prescribed burn management over the last few decades has supported the naturally occurring biodiversity of oak woodlands. As in grasslands, the highest beta diversity scores in oak woodlands are when comparing between burned sites and when comparing between burned and unburned sites. Similarly, this may reflect the beneficial nature of fires in adapted systems.

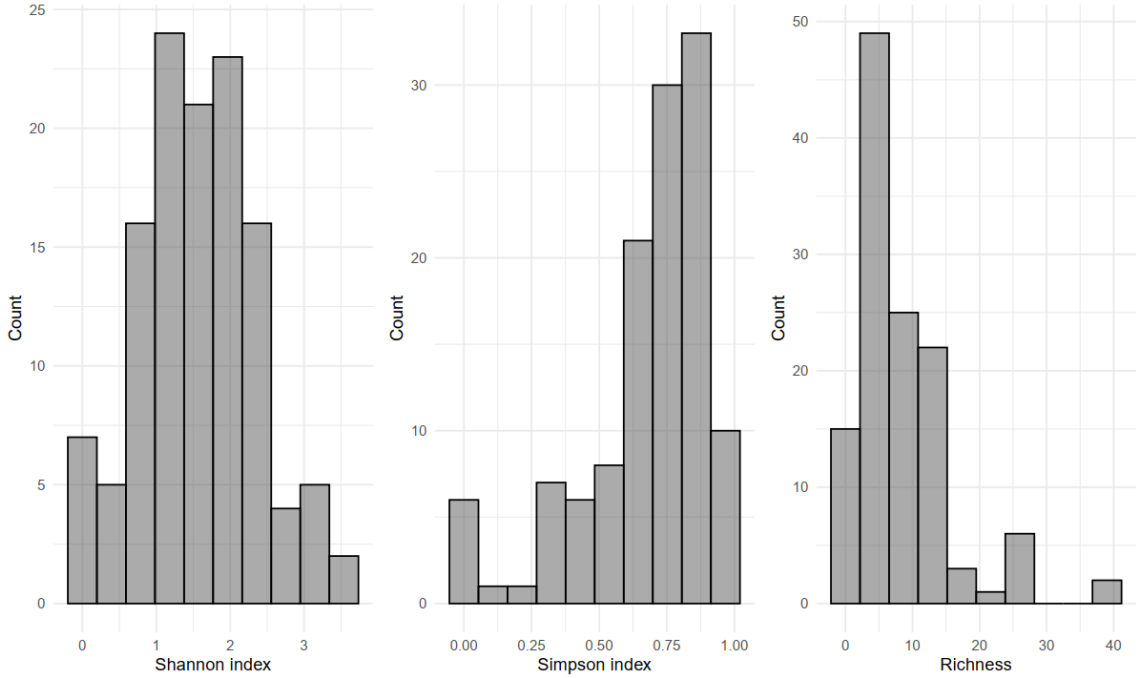
While prescribed fire has been a regular part of management, application of standard fire management strategies in habitats with alternate fire regimes like scrubland will have consequential ecological effects[235]. Fires are infrequent in scrubland, with intervals between 325 - 450 years for low sagebrush[236]. Despite this irregularity, scrublands were the most frequently burned habitat type from 2000 to 2020 in California[235]. Because early successional stages are more sensitive to perturbations, fire prior to robust recovery will alter the successional trajectory[237]. Our results show that arthropod communities in burned scrub become more dissimilar than unburned sites in spring. Contrasting both grasslands and oak woodlands, burned sites show the lowest beta diversity scores between one another, potentially indicative of a homogenization process cause by novel disturbance regimes. This may be driven by the openness created by fire in scrub habitat that prompts establishment of non-native grasses and creates habitat that promotes establishment of less common species that are more widespread. Species that serve as bioindicators in scrub habitat, such as the greater sage-grouse (*Centrocercus urophasianus*), are strongly affected by large-scale wildfires, possibly because of increases in predator populations which benefit from the new vegetation[238]. Predator-prey densities also change, altering trophic networks[239]. The now shorter fire intervals (less than five years) will likely prevent full recovery of scrubland and permanently alter ecosystem structure, shifting to more grassland-like habitat.

Our study emphasizes the importance of studying the effect of wildfires across a broad spatial scale and diverse habitat types to better understand the nuanced community responses of ecosystems. Climate change alone will alter the composition of biotic communities, in turn changing the presence of fuel which may put infrequently burned habitats at higher risk for fire. As is the case with novel environmental changes, taxa must adapt, disperse or go extinct. For taxa dependent on fire, their adaptations are associated with aspects of the regime itself rather than general fire pressure, making new wildfire dynamics a novel pressure[205]. The limited distribution and extent of habitat such as scrublands provides little refugia for the endemic and endangered species located in scrublands threatened by wildfire. Empty niche space created by dispersing or locally extinct taxa will prompt establishment of non-native or opportunistic taxa and may permanently alter the successional trajectory of the habitat. Management methods which can assist in recovery following wildfires are essential to maintain such habitat types.

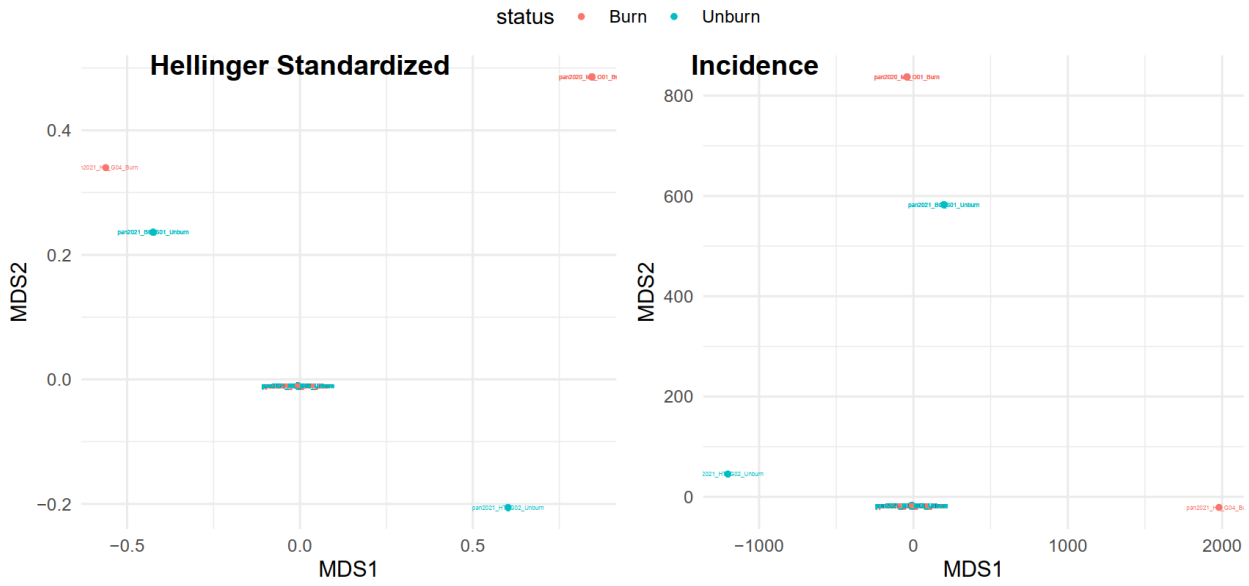
5.5. Conclusions

There are many future steps needed to make ecologically-informed management decisions. While our study built knowledge on whole community responses, an ecosystem level approach is necessary, one which incorporates biotic and abiotic factors as well as the network of interactions. More functional information is also crucial to determine the cost of compositional differences. Combining these data with attributes of the fire itself, associated with both historical and current fire regimes, will allow us to identify threatened ecosystems and move towards supporting recovery.

5.6. Supplementary Materials



Supplementary Figure 1. Histograms of diversity values using Shannon, Simpson and OTU richness.



Supplementary Figure 2. Outlier samples with few OTUs that matched no other samples, causing convergence issues in NMDS.

Predictor	SumOfSqs	R2	F	Pr(>F)
year	2.1660164	0.02687548	2.5602735	0.000999001
status	0.9836265	0.01220463	1.1626657	0.108891109
habitat	2.4344191	0.03020576	1.4387653	0.000999001
reserve	6.3635555	0.07895767	1.8804616	0.000999001
year:status	0.9872088	0.01224908	1.1669000	0.081918082
year:habitat	2.2010716	0.02731044	1.3008547	0.002997003
status:habitat	1.8856528	0.02339679	1.1144391	0.100899101
year:reserve	5.6111939	0.06962252	1.6581351	0.000999001
status:reserve	4.0178322	0.04985242	1.1872889	0.008991009
habitat:reserve	3.9943578	0.04956116	1.1803521	0.009990010
year:status:habitat	1.9522288	0.02422285	1.1537862	0.043956044
year:status:reserve	3.6665315	0.04549356	1.0834779	0.091908092
year:habitat:reserve	3.7032873	0.04594962	1.0943394	0.069930070
status:habitat:reserve	2.5350106	0.03145388	0.9988106	0.489510490
year:status:habitat:reserve	1.7141061	0.02126827	1.0130534	0.414585415
Residual	36.3784208	0.45137586	NA	NA
Total	80.5945198	1.00000000	NA	NA

Supplementary Table 1. Results of PERMANOVA using Euclidean distance matrix and inclusion of outliers.

Predictor	SumOfSqs	R2	F	Pr(>F)
year	1.3127523	0.03300788	3.228144	0.000999001
status	0.5209520	0.01309883	1.281055	0.025974026
habitat	1.2889699	0.03240989	1.584830	0.000999001
reserve	3.1862222	0.08011446	1.958782	0.000999001
year:status	0.4926112	0.01238623	1.211363	0.085914086
year:habitat	1.0459904	0.02630041	1.286079	0.004995005
status:habitat	0.9230291	0.02320867	1.134894	0.102897103
year:reserve	2.9792902	0.07491136	1.831567	0.000999001
status:reserve	2.0334191	0.05112835	1.250077	0.005994006
habitat:reserve	1.9683661	0.04949265	1.210085	0.004995005
year:status:habitat	0.8712574	0.02190692	1.071239	0.230769231
year:status:reserve	1.8830717	0.04734801	1.157649	0.015984016
year:habitat:reserve	1.7853202	0.04489014	1.097555	0.074925075
status:habitat:reserve	1.1757126	0.02956215	0.963718	0.672327672
year:status:habitat:reserve	0.8175907	0.02055752	1.005254	0.451548452
Residual	17.4863199	0.43967652	NA	NA
Total	39.7708750	1.00000000	NA	NA

Supplementary Table 2. Results of PERMANOVA using Jaccard dissimilarity matrix and inclusion of outliers.

CONCLUSION

The current biodiversity crisis calls for rapid and extensive documentation of species diversity on a global scale. Without understanding the extent, structure and formation of biodiversity, we cannot properly protect it. The discovery of a substantial number of previously unknown spider species in a relatively small sample size in Sulawesi highlights the vast extent of undocumented biodiversity in Wallacea. This reinforces the need for increased exploration and comprehensive surveys in other hyperdiverse regions that remain poorly studied. The improvement of sequencing techniques has allowed rapid classification of species as well as has made it feasible to generate large multi-locus data sets for hundreds of individual spiders in a timely and cost-efficient manner. By expanding our knowledge of species richness and composition using high-throughput barcoding approaches, we can enhance conservation efforts and develop targeted strategies to protect threatened ecosystems.

Furthermore, the insights gained from studying *in-situ* diversification of tetragnathid spiders in relation to elevation offer valuable guidance for conservation planning. As climate change continues to impact habitats globally, understanding the ecological and evolutionary processes that shape species distributions becomes increasingly crucial. By recognizing the importance of elevational associations and considering historical factors, we can identify key areas for conservation focus and prioritize the protection of vulnerable species and habitats.

One of the major barriers to documenting and monitoring whole communities in the past has been the incredible taxonomic effort needed to identify the thousands of arthropods that may be collected in a single site. The application of metabarcoding techniques to assess entire arthropod communities provides a powerful tool to rapidly assess communities and then enable monitoring of ecosystem health and evaluating the impacts of anthropogenic disturbances in near real time. By quantifying changes in community structure following stressors such as species invasions and wildfires, we can better understand the cascading effects on ecosystem functioning and the potential consequences for ecosystem services. This knowledge can inform management strategies aimed at mitigating the negative impacts and promoting the resilience of ecosystems in the face of global change.

Overall, my dissertation research contributes to the growing body of knowledge in the biodiversity sciences and highlights the need for interdisciplinary approaches that integrate genomics, ecology, and evolutionary biology. By leveraging technological advancements and enhancing our understanding of ecological processes, we can work towards a more sustainable future that preserves the remarkable diversity of life on Earth.

REFERENCES

1. Cowie, R. H., Bouchet, P. & Fontaine, B. The Sixth Mass Extinction: fact, fiction or speculation? *Biol. Rev.* **97**, 640–663 (2022).
2. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
3. Norman, M. Biodiversity Hotspots Revisited. *BioScience* **53**, 916 (2003).
4. Warren, B. H. *et al.* Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. *Ecol. Lett.* **18**, 200–217 (2015).
5. McGuire, J. A. *et al.* Species Delimitation, Phylogenomics, and Biogeography of Sulawesi Flying Lizards: A Diversification History Complicated by Ancient Hybridization, Cryptic Species, and Arrested Speciation. *Syst. Biol.* syad020 (2023) doi:10.1093/sysbio/syad020.
6. Evans, B. J., Supriatna, J., Andayani, N. & Melnick, D. J. DIVERSIFICATION OF SULAWESI MACAQUE MONKEYS: DECOUPLED EVOLUTION OF MITOCHONDRIAL AND AUTOSOMAL DNA. *Evolution* **57**, 1931–1946 (2003).
7. Nugraha, A. M. S. & Hall, R. Late Cenozoic palaeogeography of Sulawesi, Indonesia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **490**, 191–209 (2018).
8. Driller, C. *et al.* Stop and Go – Waves of Tarsier Dispersal Mirror the Genesis of Sulawesi Island. *PLOS ONE* **10**, e0141212 (2015).
9. Frantz, L. A. F. *et al.* Synchronous diversification of Sulawesi's iconic artiodactyls driven by recent geological events. (2018).
10. Nyffeler, M. & Birkhofer, K. An estimated 400–800 million tons of prey are annually killed by the global spider community. *Sci. Nat.* **104**, 30 (2017).
11. World Spider Catalog. (2023) doi:10.24436/2.

12. Fukami, T. Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects. *Annu. Rev. Ecol. Evol. Syst.* **46**, 1–23 (2015).
13. HilleRisLambers, J., Adler, P. B., Harpole, W. S., Levine, J. M. & Mayfield, M. M. Rethinking Community Assembly through the Lens of Coexistence Theory. *Annu. Rev. Ecol. Evol. Syst.* **43**, 227–248 (2012).
14. Moore, J. C. Predicting tipping points in complex environmental systems. *Proc. Natl. Acad. Sci.* **115**, 635–636 (2018).
15. Dudney, J. & Suding, K. N. The elusive search for tipping points. *Nat. Ecol. Evol.* **4**, 1449–1450 (2020).
16. Kolbert, E. *The Sixth Extinction: An Unnatural History*. (A&C Black, 2014).
17. Jordán, F. Keystone species and food webs. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 1733–1741 (2009).
18. Paine, R. T. Food Web Complexity and Species Diversity. *Am. Nat.* **100**, 65–75 (1966).
19. Mills, L. S., Soulé, M. E. & Doak, D. F. The Keystone-Species Concept in Ecology and Conservation. *BioScience* **43**, 219–224 (1993).
20. Gamfeldt, L., Hillebrand, H. & Jonsson, P. R. Multiple Functions Increase the Importance of Biodiversity for Overall Ecosystem Functioning. *Ecology* **89**, 1223–1231 (2008).
21. Joppa, L. N., Roberts, D. L., Myers, N. & Pimm, S. L. Biodiversity hotspots house most undiscovered plant species. *Proc. Natl. Acad. Sci.* **108**, 13171–13176 (2011).
22. Riedel, A. & Narakusumo, R. P. One hundred and three new species of Trigonopterus weevils from Sulawesi. *ZooKeys* **828**, 1–153 (2019).
23. Tixier, M.-S. & Kreiter, S. Arthropods in biodiversity hotspots: the case of the Phytoseiidae (Acari: Mesostigmata). *Biodivers. Conserv.* **18**, 507–527 (2009).

24. Costello, M. J., May, R. M. & Stork, N. E. Can We Name Earth's Species Before They Go Extinct? *Science* **339**, 413–416 (2013).

25. Bellard, C. *et al.* Vulnerability of biodiversity hotspots to global change. *Glob. Ecol. Biogeogr.* **23**, 1376–1386 (2014).

26. Trew, B. T. & Maclean, I. M. D. Vulnerability of global biodiversity hotspots to climate change. *Glob. Ecol. Biogeogr.* **30**, 768–783 (2021).

27. de Kerdrel, G. A., Andersen, J. C., Kennedy, S. R., Gillespie, R. & Krehenwinkel, H. Rapid and cost-effective generation of single specimen multilocus barcoding data from whole arthropod communities by multiple levels of multiplexing. *Sci. Rep.* **10**, 78 (2020).

28. Shokralla, S. *et al.* Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Sci. Rep.* **5**, 9687 (2015).

29. Valentini, A., Pompanon, F. & Taberlet, P. DNA barcoding for ecologists. *Trends Ecol. Evol.* **24**, 110–117 (2009).

30. Beng, K. C. *et al.* The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. *Sci. Rep.* **6**, 24965 (2016).

31. Oliverio, A. M., Gan, H., Wickings, K. & Fierer, N. A DNA metabarcoding approach to characterize soil arthropod communities. *Soil Biol. Biochem.* **125**, 37–43 (2018).

32. Thomsen, P. F. & Sigsgaard, E. E. Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. *Ecol. Evol.* **9**, 1665–1679 (2019).

33. Hsieh, C.-H., Huang, C.-G., Wu, W.-J. & Wang, H.-Y. A rapid insect species identification system using mini-barcode pyrosequencing. *Pest Manag. Sci.* **76**, 1222–1227 (2020).

34. Meusnier, I. *et al.* A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* **9**, 214 (2008).

35. Rowley, D. L. *et al.* Vouchering DNA-barcoded specimens: test of a nondestructive extraction protocol for terrestrial arthropods. *Mol. Ecol. Notes* **7**, 915–924 (2007).

36. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrigenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. **3**, 294–299 (1994).

37. Jusino, M. A. *et al.* An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Mol. Ecol. Resour.* **19**, 176–190 (2019).

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

39. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).

40. Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226 (2018).

41. Frøslev, T. G. *et al.* Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nat. Commun.* **8**, 1188 (2017).

42. Mahé, F., Rognes, T., Quince, C., Vargas, C. de & Dunthorn, M. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* **3**, e1420 (2015).

43. Wilkinson, S. kmer: an R package for fast alignment-free clustering of biological sequences. (2018).

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

45. Paradis, E. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419–420 (2010).

46. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).

47. Yan, L. Package 'ggvenn'. *CRAN* (2021).

48. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

49. Cardoso, P., Rigal, F. & Carvalho, J. C. BAT – Biodiversity Assessment Tools, an R package for the measurement and estimation of alpha and beta taxon, phylogenetic and functional diversity. *Methods Ecol. Evol.* **6**, 232–236 (2015).

50. Xu, S. *et al.* Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. *iMeta* **1**, e56 (2022).

51. Wang, L.-G. *et al.* Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Mol. Biol. Evol.* **37**, 599–603 (2020).

52. M, B. The Hersiliidae of the Oriental Region including New Guinea. Taxonomy, phylogeny, zoogeography (Arachnida, Araneae). *Spixiana* 1–96 (1993).

53. National Library of Medicine - National Institutes of Health. https://www.nlm.nih.gov/?_gl=1*2a74wv*_ga*MTUwNzQ3ODU1LjE2ODM0ODY2NzU.*_ga_7147EPK006*MTY4MzQ4NjY3NS4xLjAuMTY4MzQ4NjY3NS4wLjAuMA..*_ga_P1FPTH9PL4*MTY4MzQ4NjY3NS4xLjAuMTY4MzQ4NjY3NS4wLjAuMA..

54. Perrigo, A., Hoorn, C. & Antonelli, A. Why mountains matter for biodiversity. *J. Biogeogr.* **47**, 315–325 (2020).

55. Rahbek, C. The Elevational Gradient of Species Richness: A Uniform Pattern? *Ecography* **18**, 200–205 (1995).

56. Guo, Q. *et al.* Global variation in elevational diversity patterns. *Sci. Rep.* **3**, 3007 (2013).

57. Merckx, V. S. F. T. *et al.* Evolution of endemism on a young tropical mountain. *Nature* **524**, 347–350 (2015).

58. Ali, J. R. & Heaney, L. R. Wallace's line, Wallacea, and associated divides and areas: history of a tortuous tangle of ideas and labels. *Biol. Rev.* **96**, 922–942 (2021).

59. Wallace, A. R. *The Malay Archipelago: The Land of the Orang-utan and the Bird of Paradise; a Narrative of Travel, with Studies of Man and Nature*. (Courier Corporation, 1869).

60. Wallace, A. R. *Island Life*. (Macmillan & Co, London, 1880).

61. Wallace, A. R. On the Physical Geography of the Malay Archipelago. *J. R. Geogr. Soc. Lond.* **33**, 217–234 (1863).

62. Wallace, A. R. On the Zoological Geography of the Malay Archipelago. *J. Proc. Linn. Soc. Zool.* **4**, 172–184 (1860).

63. Huxley, T. H. On the classification and distribution of the Alectoromorphae and Heteromorphae. *Proc. Zool. Soc. Lond.* 294–319 (1868).

64. Lydekker, R. *A geographical history of mammals*. (Cambridge University Press, Cambridge, 1896).

65. Heilprin, A. *The Geographical and Geological Distribution of Mammals*. (Keegan Paul, Trench & Co, London, 1887).

66. Rueda, M., Rodríguez, M. Á. & Hawkins, B. A. Identifying global zoogeographical regions: lessons from Wallace. *J. Biogeogr.* **40**, 2215–2225 (2013).

67. Ficetola, G. F., Mazel, F. & Thuiller, W. Global determinants of zoogeographical boundaries. *Nat. Ecol. Evol.* **1**, 1–7 (2017).

68. Lohman, D. J. *et al.* Biogeography of the Indo-Australian Archipelago. *Annu. Rev. Ecol. Evol. Syst.* **42**, 205–226 (2011).

69. Hall, R. Southeast Asia: New Views of the Geology of the Malay Archipelago. *Annu. Rev. Earth Planet. Sci.* **45**, 331–358 (2017).

70. Hawkins, M. T. R. *et al.* Evolutionary history of endemic Sulawesi squirrels constructed from UCEs and mitogenomes sequenced from museum specimens. *BMC Evol. Biol.* **16**, 80 (2016).

71. von Rintelen, T., Stelbrink, B., Marwoto, R. M. & Glaubrecht, M. A Snail Perspective on the Biogeography of Sulawesi, Indonesia: Origin and Intra-Island Dispersal of the Viviparous Freshwater Gastropod Tylomelania. *PLoS ONE* **9**, e98917 (2014).

72. von Rintelen, T., Bouchet, P. & Glaubrecht, M. Ancient lakes as hotspots of diversity: a morphological review of an endemic species flock of Tylomelania (Gastropoda: Cerithioidea: Pachychilidae) in the Malili lake system on Sulawesi, Indonesia. *Hydrobiologia* **592**, 11–94 (2007).

73. Letsch, H. *et al.* Transgressing Wallace's Line brings hyperdiverse weevils down to earth. *Ecography* **43**, 1329–1340 (2020).

74. Eldridge, R. A., Achmadi, A. S., Giarla, T. C., Rowe, K. C. & Esselstyn, J. A. Geographic isolation and elevational gradients promote diversification in an endemic shrew on Sulawesi. *Mol. Phylogenet. Evol.* **118**, 306–317 (2018).

75. Gillespie, R. G. GEOGRAPHICAL CONTEXT OF SPECIATION IN A RADIATION OF HAWAIIAN TETRAGNATHA SPIDERS (ARANEAE, TETRAGNATHIDAE). *J. Arachnol.* **33**, 313–322 (2005).

76. Gillespie, R. G. Hawaiian Spiders of the Genus Tetragnatha: I. Spiny Leg Clade. *J. Arachnol.* **19**, 174–209 (1991).

77. Cotoras, D. D. *et al.* Co-occurrence of ecologically similar species of Hawaiian spiders reveals critical early phase of adaptive radiation. *BMC Evol. Biol.* **18**, 100 (2018).

78. Blackledge, T. A. & Gillespie, R. G. Convergent evolution of behavior in an adaptive radiation of Hawaiian web-building spiders. *Proc. Natl. Acad. Sci.* **101**, 16228–16233 (2004).

79. Simberloff, D. S. & Wilson, E. O. Experimental Zoogeography of Islands. A Two-Year Record of Colonization. *Ecology* **51**, 934–937 (1970).

80. Čandek, K., Agnarsson, I., Binford, G. J. & Kuntner, M. Biogeography of Long-Jawed Spiders Reveals Multiple Colonization of the Caribbean. *Diversity* **13**, 622 (2021).

81. Krehenwinkel, H., Kennedy, S. R., Rueda, A., Lam, A. & Gillespie, R. G. Scaling up DNA barcoding – Primer sets for simple and cost efficient arthropod systematics by multiplex PCR and Illumina amplicon sequencing. *Methods Ecol. Evol.* **9**, 2181–2193 (2018).

82. Gibson, J. *et al.* Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proc. Natl. Acad. Sci.* **111**, 8007–8012 (2014).

83. Yu, D. W. *et al.* Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol. Evol.* **3**, 613–623 (2012).

84. Fonseca, V. G. *et al.* Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nat. Commun.* **1**, 98 (2010).

85. Machida, R. J. & Knowlton, N. PCR Primers for Metazoan Nuclear 18S and 28S Ribosomal DNA Sequences. *PLOS ONE* **7**, e46180 (2012).

86. Colgan, D. Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolutio. *Aust. J. Zool.* (1998).

87. Carlsen, T. *et al.* Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecol.* **5**, 747–749 (2012).

88. van der Valk, T., Vezzi, F., Ormestad, M., Dalén, L. & Guschanski, K. Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. *Mol. Ecol. Resour.* **20**, 1171–1181 (2020).

89. Wheeler, W. C. *et al.* The spider tree of life: phylogeny of Araneae based on target-gene analyses from an extensive taxon sampling. *Cladistics* **33**, 574–616 (2017).

90. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

91. Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**, 286–298 (2008).

92. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

93. Sánchez, R. *et al.* Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Res.* **39**, W470–W474 (2011).

94. Attaluri, P. K., Christman, M. C., Chen, Z. & Lu, G. SeqMaT: A sequence manipulation tool for phylogenetic analysis. *Bioinformation* **5**, 400–401 (2011).

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

96. Magalhaes, I. L. F., Azevedo, G. H. F., Michalik, P. & Ramírez, M. J. The fossil record of spiders revisited: implications for calibrating trees and evidence for a major faunal turnover since the Mesozoic. *Biol. Rev.* **95**, 184–217 (2020).

97. Garrison, N. L. *et al.* Spider phylogenomics: untangling the Spider Tree of Life. *PeerJ* **4**, e1719 (2016).

98. Responte, M. *et al.* Northward geographic diversification of a kleptoparasitic spider *Argyrodes lanyuensis* (Araneae, Theridiidae) from the Philippine Archipelago to Orchid Island. *Ecol. Evol.* **11**, 11241–11266 (2021).

99. Xu, X. *et al.* Phylogenetic placement and species delimitation of the crab spider genus Phrynarachne (Araneae: Thomisidae) from China. *Mol. Phylogenet. Evol.* **173**, 107521 (2022).

100. Ezard, T., Fujisawa, T. & Barraclough, T. Splits: Species' limits by threshold statistics. **R package version 1.0-18/45**, (2013).

101. Emerson, B. C. & Gillespie, R. G. Phylogenetic analysis of community assembly and structure over space and time. *Trends Ecol. Evol.* **23**, 619–630 (2008).

102. Cadotte, M. W. & Tucker, C. M. Should Environmental Filtering be Abandoned? *Trends Ecol. Evol.* **32**, 429–437 (2017).

103. Kraft, N. J. B. *et al.* Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.* **29**, 592–599 (2015).

104. Münkemüller, T. *et al.* Dos and don'ts when inferring assembly rules from diversity patterns. *Glob. Ecol. Biogeogr.* **29**, 1212–1229 (2020).

105. Overcast, I. *et al.* A unified model of species abundance, genetic diversity, and functional diversity reveals the mechanisms structuring ecological communities. *Mol. Ecol. Resour.* **21**, 2782–2800 (2021).

106. Ning, D. *et al.* A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat. Commun.* **11**, 4717 (2020).

107. Ponisio, L. C. *et al.* A Network Perspective for Community Assembly. *Front. Ecol. Evol.* **7**, (2019).

108. Baselga, A., Gómez-Rodríguez, C. & Vogler, A. P. Multi-hierarchical macroecology at species and genetic levels to discern neutral and non-neutral processes. *Glob. Ecol. Biogeogr.* **24**, 873–882 (2015).

109. Emerson, B. C. & Gillespie, R. G. Phylogenetic analysis of community assembly and structure over space and time. *Trends Ecol. Evol.* **23**, 619–630 (2008).

110. Dini-Andreote, F., Stegen, J. C., van Elsas, J. D. & Salles, J. F. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc. Natl. Acad. Sci.* **112**, E1326–E1332 (2015).

111. Hiller, A. E. *et al.* Niche conservatism predominates in adaptive radiation: comparing the diversification of Hawaiian arthropods using ecological niche modelling. *Biol. J. Linn. Soc.* **127**, 479–492 (2019).

112. Pyron, R. A., Costa, G. C., Patten, M. A. & Burbrink, F. T. Phylogenetic niche conservatism and the evolutionary basis of ecological speciation. *Biol. Rev.* **90**, 1248–1262 (2015).

113. Emerson, B. C. & Kolm, N. Species diversity can drive speciation. *Nature* **434**, 1015–1017 (2005).

114. García-Navas, V., Martínez-Núñez, C. & Christidis, L. Imprints of historical and ecological factors in the phylogenetic structure of Australian Meliphagides assemblages. *J. Biogeogr.* (2023) doi:10.1111/jbi.14581.

115. Gerhold, P., Cahill Jr, J. F., Winter, M., Bartish, I. V. & Prinzing, A. Phylogenetic patterns are not proxies of community assembly mechanisms (they are far better). *Funct. Ecol.* **29**, 600–614 (2015).

116. Chase, J. M. Community assembly: when should history matter? *Oecologia* **136**, 489–498 (2003).

117. MacArthur, R. & Wilson, E. *The Theory of Island Biogeography*. (Princeton University Press, 1967).

118. Gillespie, R. G. & Roderick, G. K. Arthropods on Islands: Colonization, Speciation, and Conservation. *Annu. Rev. Entomol.* **47**, 595–632 (2002).

119. Noroozi, J. *et al.* Hotspots within a global biodiversity hotspot - areas of endemism are associated with high mountain ranges. *Sci. Rep.* **8**, 10345 (2018).

120. Krehenwinkel, H. *et al.* Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Sci. Rep.* **7**, 17668 (2017).

121. Leray, M. *et al.* A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front. Zool.* **10**, 34 (2013).

122. Kennedy, S. *et al.* Richness and resilience in the Pacific: DNA metabarcoding enables parallelized evaluation of biogeographic patterns. *Mol. Ecol.* **n/a**, (2022).

123. Elbrecht, V. *et al.* Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ* **7**, e7745 (2019).

124. Song, H., Buhay, J. E., Whiting, M. F. & Crandall, K. A. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc. Natl. Acad. Sci.* **105**, 13486–13491 (2008).

125. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).

126. Hansen, M. C. *et al.* High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**, 850–853 (2013).

127. Gotelli, N. J. Null Model Analysis of Species Co-Occurrence Patterns. *Ecology* **81**, 2606–2621 (2000).

128. Gotelli, N. J. & Entsminger, G. L. Swap Algorithms in Null Model Analysis. *Ecology* **84**, 532–535 (2003).

129. Anderson, J. M. The enigma of soil animal species diversity. in *Progress in Soil Zoology: Proceedings of the 5th International Colloquium on Soil Zoology Held in Prague September 17–22, 1973* (ed. Vaněk, J.) 51–58 (Springer Netherlands, 1975). doi:10.1007/978-94-010-1933-0_5.

130. Bardgett, R. D. Causes and consequences of biological diversity in soil1. *Zoology* **105**, 367–375 (2002).

131. Decaëns, T., Jiménez, J. J., Gioia, C., Measey, G. J. & Lavelle, P. The values of soil animals for conservation biology. *Eur. J. Soil Biol.* **42**, S23–S38 (2006).

132. Ashford, O. S. *et al.* Litter manipulation and the soil arthropod community in a lowland tropical rainforest. *Soil Biol. Biochem.* **62**, 5–12 (2013).

133. Cole, R. J., Holl, K. D., Zahawi, R. A., Wickey, P. & Townsend, A. R. Leaf litter arthropod responses to tropical forest restoration. *Ecol. Evol.* **6**, 5158–5168 (2016).

134. Stelbrink, B., Albrecht, C., Hall, R. & von Rintelen, T. THE BIOGEOGRAPHY OF SULAWESI REVISITED: IS THERE EVIDENCE FOR A VICARIANT ORIGIN OF TAXA ON WALLACE'S "ANOMALOUS ISLAND"?: BIOGEOGRAPHY OF SULAWESI. *Evolution* **66**, 2252–2271 (2012).

135. Harvey, M. S. Short-range endemism amongst the Australian fauna: some examples from non-marine environments. *Invertebr. Syst.* **16**, 555–570 (2002).

136. Edward, K. L. & Harvey, M. S. Short-range endemism in hypogean environments: the pseudoscorpion genera Tyrannochthonius and Lagynochthonius (Pseudoscorpiones : Chthoniidae) in the semiarid zone of Western Australia. *Invertebr. Syst.* **22**, 259–293 (2008).

137. Sømme, L. Adaptations of Terrestrial Arthropods to the Alpine Environment. *Biol. Rev.* **64**, 367–407 (1989).

138. Purać, J. *et al.* Cold Adaptation Responses in Insects and Other Arthropods: An "Omics" Approach. in *Short Views on Insect Genomics and Proteomics: Insect Proteomics, Vol.2* (eds. Raman, C., Goldsmith, M. R. & Agunbiade, T. A.) 89–112 (Springer International Publishing, 2016). doi:10.1007/978-3-319-24244-6_4.

139. Lee, V. M. J., Kuntner, M. & Li, D. Ballooning behavior in the golden orbweb spider Nephila pilipes (Araneae: Nephilidae). *Front. Ecol. Evol.* **3**, (2015).

140. Hulme, P. E. *et al.* Grasping at the routes of biological invasions: a framework for integrating pathways into policy. *J. Appl. Ecol.* **45**, 403–414 (2008).

141. Hulme, P. E. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J. Appl. Ecol.* **46**, 10–18 (2009).

142. Sinclair, J. S., Lockwood, J. L., Hasnain, S., Cassey, P. & Arnott, S. E. A framework for predicting which non-native individuals and species will enter, survive, and exit human-mediated transport. *Biol. Invasions* **22**, 217–231 (2020).

143. Jones, C. G., Lawton, J. H. & Shachak, M. Organisms as Ecosystem Engineers. *Oikos* **69**, 373–386 (1994).

144. Richardson, D. M. *Fifty Years of Invasion Ecology: The Legacy of Charles Elton*. (John Wiley & Sons, 2011).

145. Ehrenfeld, J. G. Effects of Exotic Plant Invasions on Soil Nutrient Cycling Processes. *Ecosystems* **6**, 503–523 (2003).

146. Weidenhamer, J. D. & Callaway, R. M. Direct and Indirect Effects of Invasive Plants on Soil Chemistry and Ecosystem Function. *J. Chem. Ecol.* **36**, 59–69 (2010).

147. Ruckli, R., Rusterholz, H.-P. & Baur, B. Invasion of Impatiens glandulifera affects terrestrial gastropods by altering microclimate. *Acta Oecologica* **47**, 16–23 (2013).

148. Schirmel, J., Bundschuh, M., Entling, M. H., Kowarik, I. & Buchholz, S. Impacts of invasive plants on resident animals across ecosystems, taxa, and feeding types: a global assessment. *Glob. Change Biol.* **22**, 594–603 (2016).

149. Bezemer, T. M., Harvey, J. A. & Cronin, J. T. Response of Native Insect Communities to Invasive Plants. *Annu. Rev. Entomol.* **59**, 119–141 (2014).

150. Harvey, J. A. & Fortuna, T. M. Chemical and structural effects of invasive plants on herbivore–parasitoid/predator interactions in native communities. *Entomol. Exp. Appl.* **144**, 14–26 (2012).

151. Keeler, M. S., Chew, F. S., Goodale, B. C. & Reed, J. M. Modelling the impacts of two exotic invasive species on a native butterfly: top-down vs. bottom-up effects. *J. Anim. Ecol.* **75**, 777–788 (2006).

152. Schlaepfer, M. A., Sherman, P. W., Blossey, B. & Runge, M. C. Introduced species as evolutionary traps. *Ecol. Lett.* **8**, 241–246 (2005).

153. Keeler, M. S. & Chew, F. S. Escaping an evolutionary trap: preference and performance of a native insect on an exotic invasive host. *Oecologia* **156**, 559–568 (2008).

154. Litt, A. R., Cord, E. E., Fulbright, T. E. & Schuster, G. L. Effects of Invasive Plants on Arthropods. *Conserv. Biol.* **28**, 1532–1549 (2014).

155. Mattos, K. J. & Orrock, J. L. Behavioral consequences of plant invasion: an invasive plant alters rodent antipredator behavior. *Behav. Ecol.* **21**, 556–561 (2010).

156. Suarez, A. V., Richmond, J. Q. & Case, T. J. Prey Selection in Horned Lizards Following the Invasion of Argentine Ants in Southern California. *Ecol. Appl.* **10**, 711–725 (2000).

157. Green, P. T. *et al.* Invasional meltdown: Invader–invader mutualism facilitates a secondary invasion. *Ecology* **92**, 1758–1768 (2011).

158. Simberloff, D. & Von Holle, B. Positive Interactions of Nonindigenous Species: Invasional Meltdown? *Biol. Invasions* **1**, 21–32 (1999).

159. Harvey, J. A., Bukovinszky, T. & van der Putten, W. H. Interactions between invasive plants and insect herbivores: A plea for a multitrophic perspective. *Biol. Conserv.* **143**, 2251–2259 (2010).

160. Pereira, M. J., Eleutério, T., Meirelles, M. G. & Vasconcelos, H. C. Hedychium gardnerianum Sheph. ex Ker Gawl. from its discovery to its invasive status: a review. *Bot. Stud.* **62**, 11 (2021).

161. Minden, V., Hennenberg, K. J., Porembski, S. & Boehmer, H. J. Invasion and management of alien Hedychium gardnerianum (kahili ginger, Zingiberaceae) alter plant species composition of a montane rainforest on the island of Hawai'i. *Plant Ecol.* **206**, 321–333 (2010).

162. Vorsino, A. E. *et al.* Modeling Hawaiian Ecosystem Degradation due to Invasive Plants under Current and Future Climates. *PLOS ONE* **9**, e95427 (2014).

163. The Nature Conservancy of Hawaii. Waikamoi Preserve, East Maui, Hawai'i. Long Range Management Plan, 2013-2018. *Dep. Land Nat. Resour. Nat. Area Partnersh. Program* (2011).

164. Cardoso, P., Pekár, S., Jocqué, R. & Coddington, J. A. Global Patterns of Guild Composition and Functional Diversity of Spiders. *PLOS ONE* **6**, e21710 (2011).

165. Krehenwinkel, H. *et al.* Multiplex PCR targeting lineage-specific SNPs: A highly efficient and simple approach to block out predator sequences in molecular gut content analysis. *Methods Ecol. Evol.* **10**, 982–993 (2019).

166. Zhao, P., Luo, J. & Zhuang, W.-Y. Practice towards DNA barcoding of the nectriaceous fungi. *Fungal Divers.* **46**, 183–191 (2011).

167. Winter, D. J. *rentrez: An R package for the NCBI eUtils API.* https://peerj.com/preprints/3179 (2017) doi:10.7287/peerj.preprints.3179v2.

168. Suman, T. Spiders of the Family Thomisidae in Hawaii. (University of Hawai'i at Manoa, 1967).

169. Dormann, C. F., Gruber, B. & Fründ, J. Introducing the bipartite Package: Analysing Ecological Networks. **8**, (2008).

170. Laporte, M. *et al.* Proper environmental DNA metabarcoding data transformation reveals temporal stability of fish communities in a dendritic river system. *Environ. DNA* **3**, 1007–1022 (2021).

171. Alberdi, A. & Gilbert, M. T. P. A guide to the application of Hill numbers to DNA-based diversity analyses. *Mol. Ecol. Resour.* **19**, 804–817 (2019).

172. Gillespie, R., Rivera, M. A. & Garb, J. Sun, surf and spiders: taxonomy and phylogeography of Hawaiian Araneae. 41–51 (1998).

173. Gillespie, R. G. Naiveté and Novel Perturbations: Conservation of Native Spiders on an Oceanic Island System. *J. Insect Conserv.* **3**, 263–272 (1999).

174. Meyer, W. M. Native Hawaiian Succineids Prefer Non-Native Ginger (Hedychium spp.) Plant Species in the Kohala Mountains, Hawaii: Conservation Ramifications. *Am. Malacol. Bull.* **30**, 147–151 (2012).

175. Bennett, G. M. & O'Grady, P. M. Host–plants shape insect diversity: Phylogeny, origin, and species diversity of native Hawaiian leafhoppers (Cicadellidae: Nesophrosyne). *Mol. Phylogenet. Evol.* **65**, 705–717 (2012).

176. Wainright, C. A., Muhlfeld, C. C., Elser, J. J., Bourret, S. L. & Devlin, S. P. Species invasion progressively disrupts the trophic structure of native food webs. *Proc. Natl. Acad. Sci.* **118**, e2102179118 (2021).

177. Christiansen, K. & Bellinger, P. Biogeography of Hawaiian Collembola: The simple principles and complex reality. *Orient. Insects* **28**, 309–351 (1994).

178. Gruner, D. S. & Taylor, A. D. Richness and species composition of arboreal arthropods affected by nutrients and predators: a press experiment. *Oecologia* **147**, 714–724 (2006).

179. Gruner, D. S., Taylor, A. D. & Forkner, R. E. The effects of foliar pubescence and nutrient enrichment on arthropod communities of Metrosideros polymorpha (Myrtaceae). *Ecol. Entomol.* **30**, 428–443 (2005).

180. Michalko, R. *et al.* Reforestations of Tropical Forests Alter Interactions Between Web-Building Spiders and Their Prey. *Ecosystems* **24**, 1962–1975 (2021).

181. Rendon, D., Taylor, P. W., Wilder, S. M. & Whitehouse, M. E. A. Does prey encounter and nutrient content affect prey selection in wolf spiders inhabiting Bt cotton fields? *PLOS ONE* **14**, e0210296 (2019).

182. Toft, S. Prey Choice and Spider Fitness. *J. Arachnol.* **27**, 301–307 (1999).

183. Birkhofer, K. & Wolters, V. The global relationship between climate, net primary production and the diet of spiders. *Glob. Ecol. Biogeogr.* **21**, 100–108 (2012).

184. Møller Marcussen, B., Axelsen, J. A. & Toft, S. The value of two Collembola species as food for a linyphiid spider. *Entomol. Exp. Appl.* **92**, 29–36 (1999).

185. Toft, S. & Wise, D. H. Growth, development, and survival of a generalist predator fed single- and mixed-species diets of different quality. *Oecologia* **119**, 191–197 (1999).

186. Rickers, S., Langel, R. & Scheu, S. Dietary Routing of Nutrients from Prey to Offspring in a Generalist Predator: Effects of Prey Quality. *Funct. Ecol.* **20**, 124–131 (2006).

187. Traugott, M. & Symondson, W. O. C. Molecular analysis of predation on parasitized hosts. *Bull. Entomol. Res.* **98**, 223–231 (2008).

188. Nishida, G. Hawaiian terrestrial arthropod checklist. *Honol. Hawaii Hawaii Biol. Surv. Bish. Mus. Tech. Rep. 22* **4**, (2002).

189. Beardsley, J. W. A Review of the Hawaiian Braconidae (Hymenoptera). (1961).

190. Henneman, M. L. & Memmott, J. Infiltration of a Hawaiian Community by Introduced Biological Control Agents. *Science* **293**, 1314–1316 (2001).

191. Bassett, I. E. Impacts on invertebrate fungivores: a predictable consequence of ground-cover weed invasion? *Biodivers. Conserv.* **23**, 791–810 (2014).

192. Shrestha, B. *et al.* Spider-pathogenic fungi within Hypocreales (Ascomycota): their current nomenclature, diversity, and distribution. *Mycol. Prog.* **18**, 983–1003 (2019).

193. Wang, Y. *et al.* Environmental DNA of preservative ethanol performed better than water samples in detecting macroinvertebrate diversity using metabarcoding. *Divers. Distrib.* **27**, 1989–2002 (2021).

194. Zimmermann, G. Review on safety of the entomopathogenic fungi Beauveria bassiana and Beauveria brongniartii. *Biocontrol Sci. Technol.* **17**, 553–596 (2007).

195. Meyling, N. V. & Eilenberg, J. Ecology of the entomopathogenic fungi Beauveria bassiana and Metarhizium anisopliae in temperate agroecosystems: Potential for conservation biological control. *Biol. Control* **43**, 145–155 (2007).

196. Greco, E. B., Wright, M. G., Burgueño, J. & Jaronski, S. T. Efficacy of Beauveria bassiana applications on coffee berry borer across an elevation gradient in Hawaii. *Biocontrol Sci. Technol.* **28**, 995–1013 (2018).

197. Hollingsworth, R. G. *et al.* Incorporating Beauveria bassiana Into an Integrated Pest Management Plan for Coffee Berry Borer in Hawaii. *Front. Sustain. Food Syst.* **4**, (2020).

198. Castrillo, L. A. *et al.* Genetic diversity among naturally-occurring strains of Beauveria bassiana associated with the introduced coffee berry borer, Hypothenemus hampei, (Coleoptera: Curculionidae) on Hawai‘i Island. *J. Invertebr. Pathol.* **175**, 107456 (2020).

199. Kuussaari, M. *et al.* Extinction debt: a challenge for biodiversity conservation. *Trends Ecol. Evol.* **24**, 564–571 (2009).

200. Connell, J. H. Diversity in Tropical Rain Forests and Coral Reefs: High diversity of trees and corals is maintained only in a nonequilibrium state. *Science* **199**, 1302–1310 (1978).

201. Roxburgh, S. H., Shea, K. & Wilson, J. B. The Intermediate Disturbance Hypothesis: Patch Dynamics and Mechanisms of Species Coexistence. *Ecology* **85**, 359–371 (2004).

202. Pausas, J. G. & Keeley, J. E. Wildfires as an ecosystem service. *Front. Ecol. Environ.* **17**, 289–295 (2019).

203. Johnstone, J. F. *et al.* Changing disturbance regimes, ecological memory, and forest resilience. *Front. Ecol. Environ.* **14**, 369–378 (2016).

204. Seidl, R. *et al.* Forest disturbances under climate change. *Nat. Clim. Change* **7**, 395–402 (2017).

205. Pausas, J. G. & Keeley, J. E. A Burning Story: The Role of Fire in the History of Life. *BioScience* **59**, 593–601 (2009).

206. Pausas, J. G. Evolutionary fire ecology: lessons learned from pines. *Trends Plant Sci.* **20**, 318–324 (2015).

207. Jager, H. I. *et al.* Resilience of terrestrial and aquatic fauna to historical and future wildfire regimes in western North America. *Ecol. Evol.* **11**, 12259–12284 (2021).

208. Core Writing Team, IPCC. Climate change 2007: synthesis report. 104 (2007).

209. Parks, S. A. & Abatzoglou, J. T. Warmer and Drier Fire Seasons Contribute to Increases in Area Burned at High Severity in Western US Forests From 1985 to 2017. *Geophys. Res. Lett.* **47**, e2020GL089858 (2020).

210. Schoennagel, T. *et al.* Adapt to more wildfire in western North American forests as climate changes. *Proc. Natl. Acad. Sci.* **114**, 4582–4590 (2017).

211. Taillie, P. J. *et al.* Interacting and non-linear avian responses to mixed-severity wildfire and time since fire. *Ecosphere* **9**, e02291 (2018).

212. Coop, J. D. *et al.* Wildfire-Driven Forest Conversion in Western North American Landscapes. *BioScience* **70**, 659–673 (2020).

213. Barton, A. M. Intense wildfire in southeastern Arizona: transformation of a Madrean oak–pine forest to oak woodland. *For. Ecol. Manag.* **165**, 205–212 (2002).

214. Lecina-Diaz, J., Martínez-Vilalta, J., Alvarez, A., Vayreda, J. & Retana, J. Assessing the Risk of Losing Forest Ecosystem Services Due to Wildfires. *Ecosystems* **24**, 1687–1701 (2021).

215. Ponisio, L. C. *et al.* Pyrodiversity begets plant–pollinator community diversity. *Glob. Change Biol.* **22**, 1794–1808 (2016).

216. Bieber, B. V. *et al.* Increasing prevalence of severe fires change the structure of arthropod communities: Evidence from a meta-analysis. *Funct. Ecol.* **n/a**, (2022).

217. Carbone, L. M., Tavella, J., Pausas, J. G. & Aguilar, R. A global synthesis of fire effects on pollinators. *Glob. Ecol. Biogeogr.* **28**, 1487–1498 (2019).

218. Moritz, M. A. *et al.* Climate change and disruptions to global fire activity. *Ecosphere* **3**, art49 (2012).

219. Cal Fire. Top 20 Largest California Wildfires. (2022).

220. Keeley, J. E. & Safford, H. D. Fire as an Ecosystem Process. in *Ecosystems of California* 27–46 (University of California Press, 2016). doi:10.1525/9780520962170-007.

221. Pausas, J. G. & Parr, C. L. Towards an understanding of the evolutionary role of fire in animals. *Evol. Ecol.* **32**, 113–125 (2018).

222. Yekwayo, I., Pryke, J. S., Gaigher, R. & Samways, M. J. Only multi-taxon studies show the full range of arthropod responses to fire. *PLOS ONE* **13**, e0195414 (2018).

223. Saint-Germain, M., Drapeau, P. & Buddle, C. M. Persistence of pyrophilous insects in fire-driven boreal forests: population dynamics in burned and unburned habitats. *Divers. Distrib.* **14**, 713–720 (2008).

224. Koltz, A. M. *et al.* Global change and the importance of fire for the ecology and evolution of insects. *Curr. Opin. Insect Sci.* **29**, 110–116 (2018).

225. Boulanger, Y. & Sirois, L. Postfire Succession of Saproxylic Arthropods, with Emphasis on Coleoptera, in the North Boreal Forest of Quebec. *Environ. Entomol.* **36**, 128–141 (2007).

226. Burge, D. O. *et al.* Plant Diversity and Endemism in the California Floristic Province. *Madroño* **63**, 3–206 (2016).

227. Novotny, V. *et al.* Why Are There So Many Species of Herbivorous Insects in Tropical Rainforests? *Science* **313**, 1115–1118 (2006).

228. Morris, J. L. *et al.* Stable or seral? Fire-driven alternative states in aspen forests of western North America. *Biol. Lett.* **15**, 20190011 (2019).

229. Podgaiski, L. R. *et al.* Spider Trait Assembly Patterns and Resilience under Fire-Induced Vegetation Change in South Brazilian Grasslands. *PLOS ONE* **8**, e60207 (2013).

230. Polchaninova, N., Savchenko, G., Ronkin, V., Drogvalenko, A. & Putchkov, A. Summer fire in steppe habitats: long-term effects on vegetation and autumnal assemblages of cursorial arthropods. *Hacquetia* **18**, 213–231 (2019).

231. Galbraith, S. M., Cane, J. H., Moldenke, A. R. & Rivers, J. W. Wild bee diversity increases with local fire severity in a fire-prone landscape. *Ecosphere* **10**, e02668 (2019).

232. Reilly, M. J. *et al.* Repeated, high-severity wildfire catalyzes invasion of non-native plant species in forests of the Klamath Mountains, northern California, USA. *Biol. Invasions* **22**, 1821–1828 (2020).

233. Hunter, M. E. *et al.* Establishment of non-native plant species after wildfires: effects of fuel treatments, abiotic and biotic factors, and post-fire grass seeding treatments. *Int. J. Wildland Fire* **15**, 271–281 (2006).

234. Nemens, D. G., Varner, J. M., Kidd, K. R. & Wing, B. Do repeated wildfires promote restoration of oak woodlands in mixed-conifer landscapes? *For. Ecol. Manag.* **427**, 143–151 (2018).

235. Calhoun, K. L. *et al.* Spatial overlap of wildfire and biodiversity in California highlights gap in non-conifer fire research and management. *Divers. Distrib.* **28**, 529–541 (2022).

236. Baker, W. L. Fire and Restoration of Sagebrush Ecosystems. *Wildl. Soc. Bull.* **34**, 177–185 (2006).

237. Kröel-Dulay, G. *et al.* Increased sensitivity to climate change in disturbed ecosystems. *Nat. Commun.* **6**, 6682 (2015).

238. Anthony, C. R., Foster, L. J., Hagen, C. A. & Dugger, K. M. Acute and lagged fitness consequences for a sagebrush obligate in a post mega-wildfire landscape. *Ecol. Evol.* **12**, e8488 (2022).

239. Holbrook, J. D. *et al.* Occupancy and abundance of predator and prey: implications of the fire-cheatgrass cycle in sagebrush ecosystems. *Ecosphere* **7**, e01307 (2016).