The Hitchhiker's Guide to Terrestriality:
Exploring Sea-to-Land Life History Transitions in Decapod Crabs

By

VICTORIA MELISSA MORGAN
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Population Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Richard K. Grosberg, Co-Chair

_____

Rachael A. Bay, Co-Chair

_____

Andrew Whitehead

Committee in Charge

2022

ABSTRACT

Much of Earth's biodiversity currently exists on land and in the air, but all life began in the oceans. Nevertheless, sea-to-land transitions are rare and physiologically challenging; few marine groups have successfully invaded land and diversified there. Decapod crabs are a notable exception; at least 11 families have colonized terrestrial environments. Furthermore, these lineages occupy different transitional stages: some lineages live intertidally and spend short periods of time above water, while others live up to several kilometers inland and only seasonally return to the ocean to hatch eggs that then undergo normal planktonic development. No matter where they exist on this spectrum of terrestriality, they all have successfully overcome significant osmotic challenges and evolved solutions to address the vast physical differences between seawater and air. The evolutionary pathways to overcoming these obstacles are largely unexplored, and little is known about the genomic basis of these impressive phenotypic and behavioral changes. For my dissertation, I sought to characterize the physiological, behavioral, and genomic changes associated with the transition from marine to terrestrial habitats in gecarcinid crabs that exhibit varying grades of terrestrial adaptation.

In my first chapter, I reclassified extant decapod land crab diversity by designing a novel framework that assigned crabs to specific "grades" based on the association between their habitats and the key traits they possess that permit them to survive in their particular habitat. In this framework, I also considered the fact that brachyuran and anomuran crabs colonized terrestrial habitats independently, and that both infraorders may have transitioned onto land either via marine environments or freshwater ones. This framework presents testable hypotheses for the sequence of key trait evolution in the land crabs, and is the first framework of its kind that classifies land crab diversity in a phylogenetic and evolutionary context.

In my second chapter, we measured tissue-specific differential gene expression in two congeneric land crab sister species displaying different degrees of terrestrial adaptation, *Tuerkayana celeste* and *T. magna*, and a highly terrestrial confamilial species, *Gecarcoidea natalis* after placing the crabs into increasingly severe desiccation conditions. We found that while most of the differentially expressed genes were more likely to be conserved across all three species, genes from families expanded in one or more lineage or genes not shared across species also appear to play a critical role in how land crabs from different terrestrial grades adapt to the unique selective challenges that accompany a terrestrial life.

In my third chapter, we sought to understand the genomic basis of red-blue color polymorphisms in *Birgus latro*, the coconut crab. Coconut crabs are the largest terrestrial arthropods on Earth and exist in two color morphs, the adaptive significance of which is currently unknown. We investigated whether sequence-level variation in the gene crustacyanin, which has been implicated in exoskeletal coloration in many other malacostracan crustaceans, played a similar role in this system using whole genome resequencing and transcriptomics. We found that while there were no significant differences in F_{st} between single nucleotide polymorphisms within crustacyanin genes and those randomly selected from across the genome, the presence of 40 paralogs of crustacyanin in the coconut crab genome might play some role in the mechanistic basis of coloration in this system.

This body of work also produced novel genomic resources for the land crab study system (*i.e.*, four transcriptomes and one ultra long-read genome) that open the door for future studies that seek to understand the genomic and transcriptomic basis of terrestrial adaptations in this, and potentially other, biological systems.

ACKNOWLEDGEMENTS:

I would like to express my deepest gratitude to my academic advisors, Dr. Richard Grosberg and Dr. Rachael Bay, for their intellectual contributions and extensive personal and professional support over the past seven years. It is no exaggeration on my part to say that I would not be submitting this document today had it not been for their care and support through some of the most personally challenging years of my life.

I'd also like to thank my qualification exam committee members, Dr. Andrew Whitehead, Dr. Geerat Vermeij, Dr. Annie Schmitt, Dr. Anne Todgham, and Dr. Eric Sanford for their intellectual advice and feedback on my scientific ideas, and Dr. Tim Caro for introducing me to some of the most amazing and charismatic animals on the planet. I'd like to thank everyone in my Population Biology community for their care and endless support, and all my lab mates, both past and present.

I also want to thank my mother and grandmother for everything they did to raise me. Although both passed away during my time as a graduate student, I know they would have been so extremely proud of me for reaching this milestone, and I love them both dearly.

Finally, I'd like to thank my husband, Joshua Watson-Zink, for literally everything. Thank you for building me up when I was down, for believing in me when I did not believe in myself, for always being ready with a big smile and an even bigger hug when I had my many rough days. Thank you, thank you, thank you – I truly would not have been able to accomplish any of this without you.

PUBLISHED IN *ARTHROPOD STRUCTURE AND DEVELOPMENT*

Making the Grade: Physiological adaptations to terrestrial environments in decapod crabs

Victoria M. Watson-Zink[1†]

[1]Department of Evolution and Ecology
University of California,
Davis
CA, 95616, USA

[†]Corresponding Author: Victoria M. Watson-Zink
vmmorgan@ucdavis.edu

Abstract: All extant macroscopic terrestrial diversity has evolved from the ancestors of a small group of successful terrestrial colonizers, but in a few lineages this transition has independently occurred multiple times in spite of the significant functional challenges it presents. Decapod crabs have transitioned from marine to terrestrial environments at least ten times, occupy diverse habitats, and display varying degrees of terrestriality. Previous attempts to categorize land crab diversity relied on single traits, did not explicitly distinguish between brachyuran and anomuran lineages, and did not separate lineages that colonized land via freshwater or marine environments. As a result, critical phylogenetic and ecological constraints were missing from these earlier classifications. In this paper, I reclassify terrestriality in the land crabs by designating four transition pathways that reflect deep phylogenetic relationships between the two decapod crab infraorders and the route-specific nature of this transition. I then describe the adaptive traits that evolved in response to six primary terrestrial selective challenges. I conclude by proposing six grades of terrestriality in this system that describe observable trait-by-environment associations, and propose studies that can test the hypothetical sequence of trait evolution and the nature of convergence in the land crabs using phylogenomic and transcriptomic tools.

Keywords: sea-to-land transitions, terrestrial adaptations, evolution, land crabs, decapods

1. Introduction

1.1. Marine-to-terrestrial transitions across the Tree of Life

All life began in the oceans, but out of the 1.5 million known macroscopic species on Planet Earth, only 15% of these species currently reside in the sea (May 1994). Since the large majority of known species (85%) occur in terrestrial habitats (Grosberg et al. 2012), at some point in evolutionary history, the marine ancestors of all extant terrestrial species transitioned out of the oceans on to land. Overall, transitions between the marine and the terrestrial realms are rare and present major evolutionary challenges. Yet, incredibly, all diversity on land, and therefore, the large majority of Earth's macroscopic metazoan species, originated from a small handful of successful terrestrial colonizers.

Recent evidence employing protein sequence comparisons and molecular clock analyses suggests that the ancestors of the earliest land plants (Embryophyta) left the oceans during the middle Cambrian – early Ordovician period (between 515.2 Ma – 473.5 Ma) (Morris et al. 2018). The first amphibious arthropods, the millipedes, emerged as early as the mid-Cambrian (Fernandez et al. 2018), while the ancestor of the arachnids colonized land in the early Ordovician (~485 Ma) (Fernandez et al. 2020) and the pancrustacean ancestor of terrestrial insects colonized land during the early Silurian (444-419 Ma) (Engel & Grimaldi 2004). These vascular plant and arthropod biotas then evolved extensively together in the late Silurian (~425 Ma) (Brookfield et al. 2020). By the late Devonian (383-359 Ma), terrestrial plants began to increase dramatically in size and complexity (characterized by true roots, leaves, and taller, more "tree-like" forms). Because the colonization and diversification of animals in terrestrial environments is trophically linked to that of plants, this particular development fueled the diversification of the insects and eventually the evolution of insect flight (Speer 1995). The

ancestor to the terrestrial vertebrates (*Tiktaalik roseae*) also left the marine realm around this time (~375 Ma) (Shubin et al. 2013), and terrestrial mollusks and isopods followed soon after in the late Carboniferous (307-299 Ma) (Vermeij & Dudley 2000; Broly et al. 2013).

Since the emergence of these early pioneers, very few marine groups have followed in their paths, except for the remaining terrestrial crustaceans (*i.e.,* the amphipods, and the brachyuran and anomuran crabs), which invaded land as recently as the Cenozoic (<66 Ma) (Vermeij & Dudley 2000, Copilaş-Ciocianu et al. 2020). In this paper, I integrate physiological, behavioral, and evolutionary perspectives to categorize the colonization and resulting radiation of the terrestrial and semi-terrestrial land crabs.

1.2. Defining marine-to-terrestrial transitions in the decapod crabs

Decapod land crabs are an exceptional example of an ancestrally marine lineage that has repeatedly and independently invaded terrestrial habitats. At least ten families have colonized land (Bliss 1968) via several different routes (*i.e.,* via the littoral zone, mudflats, sandflats, and mangrove forests or after having first invaded freshwater habitats via estuaries) (Table 1, Figure 1). Furthermore, these lineages occupy different transitional stages: some lineages live intertidally and spend only short periods of time above water (*e.g.,* many Grapsidae), whereas others live up to several kilometers inland and only seasonally return to the ocean *en masse* to release eggs that then undergo normal planktonic development (*e.g.,* Gecarcinidae) (Burggren & McMahon 1988). Still other species have taken the final step to divorce themselves from the sea by unmooring their early developmental stages from the ocean completely; these species suppress most of their ancestral larval stages within the egg and hatch as juvenile land crabs (*e.g.* many species of *Geosesarma*) (Ng & Tan 1995). In a system with so many independent evolutionary origins and degrees of terrestriality, what does it mean to be "terrestrial"? Which

species qualify as land crabs? And how many times, and by which evolutionary pathways, has some form of terrestriality evolved in this clade?

Because of the high degree of functional and ecological convergence, the apparent gradient of terrestrial landscapes that the crabs occupy, and the many diverse terrestrial phenotypes displayed by the decapod crabs, many attempts have been made to define what "terrestriality" means in this clade. It has proven a difficult challenge. As Bliss (1968) noted, "It is apparent that one can draw no sharp line between terrestrial and semi-terrestrial decapod crustaceans nor between semi-terrestrial and aquatic forms." The most notable of these classification attempts include Powers et al. (1983) and Hartnoll (1988) (also see Bliss 1968, Warner 1977, Rebach & Dunham 1983, Atkinson & Taylor 1988, Greenaway 1998). Although each attempt to define terrestriality has brought the field closer to circumscribing the multivariate nature of terrestriality in decapods, it has been more than two decades since the question was last assessed by Greenaway (1998). Since then, major advances in our understanding of decapod physiology, species-specific adaptations to terrestriality, character-state evolution, and phylogenetic relationships makes it timely to re-examine how terrestriality may have evolved in this clade. In particular, without complete and robust knowledge of the phylogenetic relationships in decapods, analysis of the evolution of terrestriality is simply impossible. Earlier classifications were based on species lists that were not nearly as complete as they are today (currently, Ng et al.'s *Systema Brachyuroum* (2008) provides a very thorough checklist of all described extant brachyuran species), and because deep molecular phylogenetic analyses were not available when these earlier classifications were published, they also lacked robust evolutionary and phylogenetic frameworks.

Hartnoll's grades of terrestrial adaptation, as presented in *Biology of the Land Crabs*

(1988), is arguably the most prominent of these attempts, but because it was focused largely on the traditionally defined gecarcinid "land crabs" and the coenobitid hermit crabs, many highly terrestrial freshwater-derived lineages were given relatively minimal treatment or excluded from the classification altogether. Hartnoll's classification also ranked decapod families from T1 (a rank containing the least terrestrially-adapted land crabs) to T5 (containing the most terrestrially-adapted land crabs), based mostly on their relative dependence on standing fresh and salt water resources. As a character, independence from standing water is very difficult to quantify because of the highly diverse adaptations different lineages employ to combat desiccation stress; but, more significantly, using this metric to rank the clades ignores many other important biological, physiological, ecological, and behavioral traits associated with the marine to terrestrial transition, as well as the selective drivers associated with increasingly terrestrial lifestyles. Furthermore, while Hartnoll's discussion acknowledged in passing that brachyuran and anomuran land crabs most likely independently evolved their terrestrial habits, and that there were at least two apparent ecological routes onto land (directly via the high intertidal or after first adapting to freshwater environments), these critical phylogenetic and environmental considerations were not incorporated into the T1 - T5 designations themselves. Since phylogenetic history and the ecological route taken onto land ultimately determine which adaptive traits and strategies are evolutionarily possible for each clade, this was perhaps a missed opportunity.

In this paper, I introduce a new trait-based framework for analyzing the evolution of terrestriality in the decapod crabs by describing their marine-to-terrestrial transition in terms of six grades of terrestriality that incorporate much of our current knowledge of land crab taxonomy, systematics, biology, physiology, behavior, and ecology. I first propose four distinct transition pathways (TP) of land crabs to incorporate general phylogenetic constraints and the

route-specific nature of terrestrial adaptations in this system in my classifications. I then briefly describe the primary selective challenges faced by ancestrally-marine arthropod lineages as they transition into terrestrial habitats, and then explore the suite of physiological and behavioral adaptations that have evolved in response to each major challenge across the four TPs. Finally, I categorize the adaptations across the six grades of terrestriality and identify which adaptations appear to be key traits associated with each grade's occupation. It is important to note that although the terrestrial grades display directionality of a sort, they are not meant to imply orthogenesis towards a terrestrial optimum. By proposing the ordered grades, I instead intend to present a model of sequential trait evolution in this system that flows from the logic of function – one that can be tested using modern phylogenetic tools and approaches. By so doing, I aim to more fully capture the breadth and diversity of extant terrestrial crab species, present phylogenetically testable hypotheses about how adaptive traits may have evolved in this system, and create a foundation for future studies that seek to determine how many times, and by which adaptive pathways, terrestriality has arisen in this ecologically and evolutionarily labile system.

2. Brachyuran and anomuran terrestrial Transition Pathways (TP)

It is widely hypothesized that land crabs transitioned onto land via two distinct routes: either directly from the marine realm via the high-intertidal zone, intertidal mudflats, sandflats, or mangrove forests, or after first occupying freshwater habitats via estuaries (Hartnoll 1988). In both scenarios, the crabs evolved traits that allowed them to spend increasing amounts of time emersed. Therefore, I categorize lineages that have adapted to terrestrial habitats via the high-intertidal zone, mudflats, sandflats, and mangrove forests as transition pathway 1 (TP:1) lineages, and those that adapted via estuaries as transition pathway 2 (TP:2) lineages.

In this paper however, I make an additional distinction that separates land crabs across the two decapod infraorders that contain all extant crabs (*i.e.,* Brachyura, which contains all the "true crabs," and Anomura) since these two ancient clades diverged long before either became terrestrial. The crown ages for Brachyura and Anomura are between 250 - 230 Ma (early Triassic) (Wolfe et al. 2019), but the earliest fossil anomuran land crab (family Coenobitidae) dates back to the early Miocene (<23 Ma) (Greenaway 2003), while brachyurans likely began colonizing land during the early Cenozoic (<66 Ma) (Vermeij & Dudley 2000). Therefore, I place anomuran land crabs in the TP:1a and TP:2a subcategories, and brachyuran land crabs in the TP:1b and TP:2b subcategories.

These four major transition pathways, explored below, aim to encompass all of the decapod lineages that partially or fully transitioned on to land, and provide a framework for categorizing adaptations to each selective challenge associated with terrestrialization (summarized in Table 2; relationships shown in Fig. 1):

TP:1a - Anomuran crabs that colonized land via marine environments (including intertidal mudflats, sandflats, mangrove forests, etc.) (*e.g.*, Paguridae Latreille 1802; Porcellanidae Haworth 1825; Coenobitidae Dana 1851; and Diogenidae Ortmann 1892). While some clades of diogenid, pagurid, and porcellanid crabs show some minor terrestrial tendencies (*e.g*., some occupy intertidal and mangrove habitats, or they can briefly tolerate emersion (Greenaway 2003)), since they are not typically active out of water and none penetrate farther onto land, only the terrestrial adaptations in the coenobitid land crabs will be considered.

TP:1b - Brachyuran crabs that colonized land via marine environments (including intertidal mudflats, sandflats, mangrove forests, etc.) (*e.g.*, some Ocypodidae Rafinesque 1815; Grapsidae MacLeay 1838; Gecarcinidae MacLeay 1838; some Sesarmidae Dana 1851; Mictyridae Dana 1851; and some Varunidae H. Milne-Edwards 1853). While all the listed clades are thoracotremes, this grouping is not monophyletic according to the most recent Brachyuran Tree of Life (Tsang et al. 2014). Therefore, the presence of terrestrial representatives across these five non-monophyletic clades likely indicates multiple independent origins of terrestriality in this group (Figure 1).

TP:2a - Anomuran crabs that colonized land via freshwater environments (*e.g.*, Diogenidae and Aeglidae Dana 1852). There is a single species of freshwater hermit crab that is native to Vanuatu (*Clibanarius fonticola*; McLaughlin & Murray 1990) and a family (Aeglidae) of 63 species of freshwater anomuran crabs that are all restricted to Southern South America (*i.e.* Chile, Brazil, Bolivia, Uruguay, Paraguay, and Argentina) (Bond-Buckup et al. 2008). Because little is known about their biology or physiology, and all are primarily freshwater crabs with no known terrestrial representatives, for the purposes of this paper, trait analyses for this subcategory will be excluded.

TP:2b - Brachyuran crabs that colonized land via freshwater environments (*e.g.*, Trichodactylidae H. Milne-Edwards 1835; Sesarmidae, some Varunidae, some Ocypodidae, Pseudothelphusidae Ortmann 1893; Potamidae Ortmann 1896; and Gecarcinucidae Rathbun 1904). According to the Brachyuran Tree of Life (Tsang et al. 2014), these clades are also non-monophyletic, suggesting that there were multiple

origins of terrestriality in this TP as well (Figure 1).

3. Overview of the Six Grades of Terrestriality

The Six Grades of Terrestriality model describes observable trait-by-habitat associations across the extant lineages of land crabs. Each grade corresponds to a particular arrangement of key traits within a given TP that make living in a particular grade possible. The hypothesis builds on land crab habitat types ordered along a hypothetical environmental gradient of increasing desiccation stress that extends from lower intertidal or estuarine habitats to xeric habitats (which are arguably the most challenging terrestrial landscapes for ancestrally marine arthropod lineages). I classify the adaptive traits observed across all three Transition Pathways in response to the six primary selective challenges (discussed in detail in sections 5.1 – 5.6 below), and identify six distinct grades of terrestriality (*i.e.,* six suites of trait-by-habitat associations) among the extant land crabs (Table 3).

It is important to note that the arrangement of adaptive traits has been done somewhat rationally; I did not consider implausible combinations (*e.g.*, living in an arid zone is not reasonably possible for a crab that has not yet evolved an alternative respiratory structure); as a result, the Six Grades of Terrestriality represent a hypothesis for how these adaptive traits may have sequentially evolved in the land crab system. It is important to recognize that the transition pathways do not represent an evolutionary transition series, and that lineages do not necessarily have to evolve sequentially from one grade to the next. The order of transitions may differ among lineages, and the suites of traits currently associated with particular grades may not have, and most likely did not, evolve as well-integrated packages. Nevertheless, with the application of deep phylogenetic and phylogenomic approaches, particularly for the brachyuran crabs (see

Tsang et al. 2014; Wolfe et al. 2019; and Ma et al. 2019), and advances in phylogenetic models of trait evolution, the actual sequence of adaptive trait evolution among the more or less terrestrial members of the Brachyura will, at last, become clear.

4. Primary Selective Challenges

For an ancestrally-marine arthropod lineage to successfully colonize land and diversify there, it must overcome at least six primary selective challenges, many of which are functionally interrelated:

1. Osmotic and ionic regulation
2. Nitrogenous waste excretion
3. Aerial respiration
4. Desiccation and thermoregulation
5. Ecdysis
6. Terrestrial early development

An extensive discussion of the six primary selective challenges and their corresponding terrestrial adaptations is beyond the scope of this paper (but see Richardson & Araujo (2015) for a thorough review across all decapods); instead, I will very briefly review each primary selective challenge and its associated adaptations to both contextualize the Six Grades of Terrestriality model in land crabs and to identify instances of putative convergent trait evolution in this system. By analyzing how different lineages have adapted to terrestrial environments, it also becomes apparent that even within clades, sometimes multiple functional solutions evolved in response to

each of the selective challenges.

In addition to these six primary challenges, the physical differences between air and water as media pose other functional challenges related to visual and olfactory signal transmission and reception (see Stensmyr et al. 2005; Krieger et al. 2015; and Chou et al. 2020), and locomotion due to differences in gravitational loading while emersed (see Taylor 2018 for an excellent discussion of this challenge, in particular). Inhabiting a new realm also presents challenges with nutrient acquisition that have led to many of these terrestrial lineages becoming almost completely herbivorous (see Greenaway & Linton 1995 for a case study on *Gecarcoidea natalis*, which has a diet that is more than 90% lignified plant matter), while most marine crabs are detritivores. Inclusion of these and other functional challenges and their corresponding adaptive responses would open new dimensions to the analysis presented here. However, for the time being, there are currently not enough relevant trait data across the decapod Tree of Life to go beyond mere storytelling.

5.1 Osmotic and ionic regulation

A species' osmoregulatory status is perhaps the fundamental metric for determining its relative degree of terrestrial adaptation. The most terrestrially adapted land crabs across all three primary TPs are obligate strong osmoregulators, while the majority of marine crabs, which inhabit aquatic habitats with diverse salinity conditions, tend to be either stenohaline or euryhaline osmoconformers (Lignot & Charmantier 2015). Between these two extremes lie alternative osmoregulatory states (*i.e.*, weak osmoregulation and facultative strong osmoregulation); therefore, a species' osmoregulatory state critically determines its ability to survive in terrestrial habitats and utilize freshwater while on or near land.

In terms of marine crabs, when environmental salinity changes rapidly, physiologically appropriate osmotic conditions in the hemolymph and tissues of these crabs are maintained by modifying the intracellular concentrations of free amino acids and phosphoric compounds to minimize osmotic gradients at the cell membrane surface. This process, however, which is believed to be the ancestral trait for the decapod crabs, is relatively slow and inefficient (Mantel & Farmer 1983; Lignot & Charmantier 2015).

On the other hand, a few marine species and all freshwater species are euryhaline osmoregulators, and can maintain stable osmotic conditions in their hemolymph and tissues across a wide range of tolerable environmental salinities (Lignot & Charmantier 2015). Many of these species can switch reversibly between osmoconformation and osmoregulation, depending on their environment and the length of exposure to varying salinity levels (*i.e.*, facultative osmoregulation). In dilute environments, they hyper-osmoregulate, but in highly saline environments, they remain iso-osmotic via osmoconformation. The TP:1 species that are facultative osmoregulators are termed "weak osmoregulators" because they can only briefly tolerate hypo-osmotic environments, but cannot reside in or migrate to full freshwater. The TP:2 species in this group are termed "strong osmoregulators" because they are fully adapted to freshwater and have a high capacity for salt absorption via their gills. Because of this, however, they can only tolerate increased environmental salinities for short periods of time and cannot return to full-strength seawater (Lignot & Charmantier 2015).

Finally, since the concentrations of freely available salts and ions in the medium drops off precipitously between water and air, the land crabs must, by necessity, be strong osmoregulators and must even tolerate desiccation while on land. They maintain constant osmotic conditions by hyper-osmoregulating when they are exposed to dilute seawater,

freshwater, and while on land, and by hypo-regulating when exposed to seawater (Lignot & Charmantier 2015).

TP:1a hermit crabs inhabit a variety of environments: some dwell on the beach and drink seawater, whereas others live farther inland and primarily drink freshwater. But all are weak osmoregulators of sorts that maintain hemolymph that is iso-osmotic to the reservoir of water they carry in their shells. The shell reservoir serves as an intermediate environment between their internal environment and external water sources, and hermit crabs use a variety of physiological and behavioral tactics (*e.g.*, drinking water, flushing their shells with freshwater or saltwater, urinating, and re-ingesting expelled fluids) to maintain relatively constant reservoir osmolarities (Greenaway 2003). The coconut crab (*Birgus latro*) is the only exception; as it does not carry a mollusk shell into adulthood, these crabs regulate their hemolymph directly against the environment, and are remarkably tolerant of hyperosmotic hemolymph for extended periods (Greenaway 2003).

TP:1b crabs display a range of osmoregulatory states depending on their environments. Crabs that reside intertidally, being only periodically exposed to air, tend to be weak osmoregulators (*e.g.*, *Pachygrapsus crassipes*) or facultative strong osmoregulators, depending on the season (as in *Hemigrapsus nudus* and *H. oregonensis*) (Dehnel 1962). Farther inland, TP:1b crabs exhibit obligate strong osmoregulation and many tolerate desiccation and utilize freshwater (as in *Discoplax celeste*, see Ng & Davie 2012).

TP:2b crabs that reside in brackish environments are facultative strong osmoregulators, whereas those that are adapted to full freshwater or terrestrial habitats are all obligate strong osmoregulators (Lignot & Charmantier 2015) (Table 2).

Additionally, since most crabs, regardless of environment, produce a urine that is

isosmotic to their hemolymph, to avoid losing salts at high rates in their urine, some freshwater and all terrestrial crabs further process their urine post-renally, They use their gills and their guts to reabsorb key ions and microminerals (Wolcott & Wolcott 1990; Morris & Ahern 2003). Since salts play a critical role in the movement of water, this ion reabsorption via the gills is a particularly important protection against desiccation for terrestrial crabs, because it ultimately reduces the amount of water that is lost during urination, while also conserving critical ions. Furthermore, while the nephropores of most marine and freshwater crabs are located distally in relation to the mouth, in land crabs, they are located more proximally (in *Ocypode quadrata*, they lie above the third maxillipeds, allowing the urine to trickle down into the mouthparts), or in more terrestrial species, are even situated posterior to the third maxillipeds, which more readily permits urine drinking (as in *Gecarcinus lateralis*, see Wolcott & Wolcott 1990).

5.2 Nitrogenous waste excretion

Excretory mode is a trait closely linked to osmoregulation that determines a species' relative degree of terrestrial adaptation. Ammonia is a waste byproduct of protein and nucleic acid metabolism, and it is highly toxic to most animals (Weihrauch et al. 2004). There are three main strategies in the Metazoa for releasing this waste: ammoniotely, where nitrogenous waste is released as soluble ammonia, ureotely, where nitrogenous waste is excreted as urea, and uricotely, where nitrogenous waste is released as uric acid. In terms of relative toxicity, energetic cost, and solubility, uric acid is the least toxic and least soluble form (therefore requiring little water to excrete), but it is the most costly form to produce (although recent studies suggest that xanthine dehydrogenase may play some role in reducing the overall production cost of uric acid in this system, see Linton et al. 2017). Ammonia, on the other hand, is the most toxic and most

soluble form (therefore requiring plenty of water to excrete), but the least energetically costly to produce. Urea falls between those two extremes in terms of relative toxicity, solubility, and energetic cost.

Only two modes of nitrogenous waste excretion exist in the decapod crabs, ammoniotely and uricotely. For an animal that lives in an aquatic habitat with ready access to water, ammoniotely would be the most efficient strategy, and ammoniotely is indeed the ancestral excretory state for the decapod crabs. Surprisingly, almost all extant crab species, regardless of habitat, still possess this ancestral mode of excretion even though it is particularly disadvantageous in terrestrial habitats where desiccation is a persistent challenge (Weihrauch et al. 2004). There is a single notable exception: *Birgus latro*, the coconut crab, has independently evolved uricotely (specifically purinotely), and it releases its nitrogenous waste as a urate and guanine purine pellet in its feces (Linton et al. 2005; Linton et al. 2017). However, the main nitrogenous excretory product of the remaining TP:1a terrestrial hermit crabs remains unknown. They do not appear to be purinotelic (their feces contain very little purine), but like *B. latro*, they do possess urate deposits in their viscera (Linton et al. 2017). They have no detectable urea in their blood, only trace amounts of ammonia, and somewhat elevated uric acid concentrations, but this might result from the visceral urate deposits (Greenaway 2003).

Terrestrial crabs across TP:1b and TP:2b, however, have evolved a suite of adaptations that allow them to maintain ammoniotely while avoiding desiccation. This allows us to distinguish between ancestral ammoniotely (*i.e.,* low tolerance for ammonia in the hemolymph, and no storage excretion or postrenal urine reprocessing capabilities) and derived ammoniotely (*i.e*., high tolerance for ammonia in the hemolymph and storage excretion and postrenal processing capabilities) (Table 2).

Terrestrial crabs can tolerate relatively high levels of ammonia in their hemolymph (0.1 - 2 mmol/L), compared to aquatic species (100 - 200 μmol/L)(Linton et al. 2017). In this respect, *Geograpsus grayi*, a highly terrestrial and carnivorous species, is especially notable, tolerating hemolymph ammonia concentrations up to 1.92 mmol/L (Greenaway & Nakamura 1991).

During dry periods, when drinking water is scarce (which can last as long as three years for *Austrothelphusa transversa*, the Australian arid zone crab (Lignot & Charmantier 2015)), land crabs can also drastically reduce their urine output to conserve water, and can even stop producing urine altogether (Linton et al. 2017). To avoid the toxicity of accumulated ammonia, it is hypothesized that their mitochondria convert it into less toxic, non-essential amino acid intermediates in a process termed transamination. The crabs then accumulate these urate deposits, which appear as white spherical deposits in and around their visceral tissues. While the exact function of these visceral uric acid stores has yet to be determined, leading hypotheses support a major role in storage excretion (other hypotheses claim that they function as emergency nitrogen stores for the crabs and are catabolized during times of high nitrogen demand or inadequate dietary intake (Linton et al. 2017)). Once water becomes available again, these uric acid intermediates are catabolized into urea in the tissues and secreted into the hemolymph, then absorbed by the gills, where they are hydrolyzed to form ammonia. This ammonia is then secreted into the urine during its reprocessing in the branchial chamber, and finally excreted (Linton et al. 2017).

5.3 Aerial respiration

The third trait associated with a species' relative degree of terrestriality is its primary respiratory organ. The presence of any type of branchiostegal lung, a "rebreather system," or a

dorsal respiratory membrane in an obligate air-breathing crab is associated with a greater degree of terrestriality, but since there appear to be phylogenetic and allometric constraints that determine which strategy evolves in a given taxon, I consider all of the described adaptations as "equally terrestrial" adaptations to the aerial respiration selective challenge (Table 2).

All aquatic and some air-breathing crabs rely on phyllobranchiate gills to respire. The gills consist of many thin, closely packed, cuticularized plates (the lamellae) that are arranged in pairs along the gill axes, and crabs can have up to nine pairs of gills in their branchial chamber (Farrelly & Greenaway 1992). Water flows between the lamellae and provides structural support to maximize gas exchange. Hemolymph circulates through the gill lamellae, and when these tissues are kept moist, oxygen can diffuse into the hemolymph and circulate around the body. Regardless of habitat, for nearly all crabs, the branchial chambers with the gills are ventilated much in the same way: the tiny respiratory appendages from the second maxillae, the scaphognathites, regularly propel water or air in through the carapace, over the gills, and out of small pores near the mouth (O'Mahoney & Full 1984). A notable exception to this has been observed in *Austrothelphusa transversa*, which ventilates its lungs through the alternating muscular expansion and contraction of its epibranchial chambers (Greenaway & Taylor 1976; Greenaway et al. 1983).

Typically, without the buoyant support of water, gills tend to collapse, trapping water between the lamellae. This ultimately limits their respiratory efficiency and can lead to hypoxia, hypercapnia, and even death, if they remain emersed for extended periods of time (Farrelly & Greenaway 1992). To overcome this stress, many amphibious and terrestrial brachyuran crabs of both freshwater and marine origin have adapted by thickening the cuticle layer around their lamellae, thereby making them more rigid, and by also adding chitinous ridges to specific

regions of the gill backbone, which mechanically separate the lamellae (Díaz & Rodríguez 1977; Morris 2002). The more terrestrial of these species have also both reduced the overall size and quantity of their gills and possess an accessory respiratory organ, the branchiostegal lung, that enhances aerial respiration (Burggren & McMahon 1988). In the presence of a branchiostegal lung, the gills then primarily take on an osmoregulatory role (Farrelly & Greenaway 1992).

In some TP:1b and TP:2b taxa (*e.g.*, Gecarcinidae and Trichodactylidae, respectively), the branchial chamber itself is typically greatly enlarged to accommodate the branchiostegal lung, and the membrane lining the chamber is smooth and simply vascularized (Díaz & Rodríguez 1977; Burggren 1992; Farrelly & Greenaway 1993). In other TP:1b crabs (*e.g.*, Ocypodidae), the branchial chamber is slightly enlarged and the vascularized lining possesses moderate folding that further increases the surface area of the branchiostegal lung. In yet other TP:2b clades (*e.g.*, Pseudothelphusidae), instead of enlarging the overall volume of the branchial chamber, the surface area of the branchiostegal lung is instead increased via elaborate tufting of the branchial membrane and increased vascularization. In general, both brachiostegal lung types are extremely efficient at oxygen uptake, especially when compared to the relative performance of the gills (although *G. natalis* has gills and a branchiostegal lung that are equally efficient at gas exchange, which is a notable exception (Farrelly & Greenaway 1994)). Generally, as chamber volume decreases, overall complexity of the membrane increases, but there remains an allometric relationship between the volume of the chamber and the mass of the animal that appears to constrain the possible sizes of the branchial chamber (Díaz & Rodríguez 1977).

Some terrestrial brachyuran crabs in TP:2b (*e.g.*, many Sesarmidae) have adapted to the challenge of aerial respiration in a slightly different way. These crabs rely on circulating branchial water around the outside of their bodies via a “rebreather system” of sorts to

adequately absorb oxygen from terrestrial environments (Felgenhauer & Abele 1983). The rebreather system depends on setal capillary action to passively conduct water from the moist substrate into the branchial chambers of the crabs, across the gills, and out of pores near the scaphognathites at the anterior end of the crab. Once the water is outside the body, sharply angled setae on the frontal exterior surface of the carapace then utilize the high surface tension of water to retain it near the carapace. This now-aerated water then either flows across the dorsal surface of the carapace, where it is cleansed by additional setae and recollected by plumose setae on the legs, or across the ridged exterior surface of the branchiostegites. Each ridge is associated with a pore in the cuticle and acts as a site of oxygen uptake and connects to the spongy, vascularized tissue in the branchial chamber that serves as a morphologically simple branchiostegal lung (Felgenhauer & Abele 1983).

The anomuran terrestrial crabs in TP:1a, represented entirely by the coenobitid hermit crabs (a group of about 16 species, including *Birgus latro*, see Greenaway 2003), exemplify yet another solution to the selective challenge of aerial respiration. Like the brachyurans discussed above, all coenobitid hermit crabs have fewer and smaller gills than their marine counterparts. While *B. latro* has also convergently evolved a large, complex, highly efficient branchiostegal lung, the remaining shell-bound coenobitids are limited by the available volume of their shells, and therefore possess comparatively small and inefficient branchiostegal lung structures. To overcome this limitation, they have additionally evolved a thin, highly vascularized membrane located on the dorsal surface of the abdomen for gas exchange. This membrane remains in contact with the air pocket present at the top of the shell, and the crabs use distinctive carapace movements to periodically renew the air in the internal apices of their shells (McMahon & Burggren 1979; Greenaway 2003; Farrelly & Greenaway, 2005).

To conclude, since there are phylogenetic and allometric constraints that determine which of the above strategies ultimately evolves in a given taxon, all of the described adaptations are alternative solutions to the aerial respiration functional challenge (Table 2). Furthermore, the diversity of branchiostegal lung types is itself evolutionarily intriguing; it suggests that this kind of accessory respiratory structure may have independently evolved multiple times across the land crabs.

5.4 Desiccation and thermoregulation

Because of their aquatic ancestry, desiccation and thermal stress are perhaps the greatest physiological challenges to the terrestrial crabs. Since nearly all terrestrial species can be found in the tropics and subtropics due to high ambient humidity, and all are ectothermic, they tend to consistently exist at or near their upper limit of heat tolerance (Bliss 1968). In these warm habitats, water evaporates rapidly from the moist respiratory membranes (which are constantly exposed to a stream of fresh air via the ventilating scaphognathites), and while this evaporative process can reduce their body temperature in the short term, over extended periods of time it can prove fatal for the crabs by drying out and compromising the gas exchange capabilities of their gills and accessory respiratory membranes. Thus, the fourth metric for determining a species' terrestrial grade relates to the various methods used by the crabs to maintain physiologically appropriate body temperatures while avoiding desiccation. Specifically, obligate fossoriality, diurnal or seasonal shelter-seeking behavior (*e.g.*, nocturnality and estivation), and the presence of water-wicking setae on the external carapace and appendages fundamentally determine a species' capacity for thermoregulation and its relative degree of terrestriality (Table 2).

The primary behavioral strategy that land crabs across all three TPs use to stay cool is

constructing burrows in the sand or soil, or inhabiting other suitable microhabitats (*e.g.*, in rocky coral rag crevices for *B. latro* and other coenobitid hermit crabs) when the surrounding air temperatures rise. Burrow construction is not unique to the brachyuran and anomuran land crabs. Many other aquatic, freshwater, and semi-terrestrial crustaceans, including those in the clades Stomatopoda, Isopoda, Amphipoda, Caridea, Astacidea, and Thalassinidea, burrow, and there is an extensive fossil record of crustacean burrows dating back to at least the lower Jurassic (Carmona et al. 2004; Atkinson & Eastman 2015). However, most terrestrial species are obligately fossorial, and some arboreal or vegetation-associated species (*e.g.*, some Sesarmidae) burrow as well, implying that the co-opting of this behavior in terrestrial environments may be a key innovation that permitted a more extensive colonization of land than may have otherwise been possible for these ancestrally aquatic animals. Aside from providing refuge from predators, burrows also provide safe locations for mating, molting, and stockpiling food; burrows also buffer crabs from environmental temperature extremes and provide access to either standing water or damp sand, which some crabs use for rehydration (Atkinson & Eastman 2015).

For highly terrestrial species across all transition pathways, temporal shifts in primary activity period during drier seasons mediate desiccation and thermal stress. For example, during the dry season, *Gecarcoidea natalis* (TP:1b) only appears to be active during the coolness of the early morning, taking refuge during the hotter periods of the day in burrows in the coastal forest (Orchard 2012). Other land crabs (*e.g.*, *Gecarcinus lateralis* (TP:1b) and most coenobitid terrestrial hermit crabs, except for *B. latro*, are exclusively nocturnal (*B. latro* is known to be cathemeral only on Christmas Island and is nocturnal elsewhere in its distributional range, see Amesbury 1980; Orchard 2012; and Caro & Morgan 2018) (Bliss 1968; Becchi et al. 2015). Additionally, *Austrothelphusa transversa* and *Parathelphusa guerini*, which are both TP:2b

crabs, periodically estivate during extreme droughts in the Australian desert and in India, respectively (McCann 1938; Hartnoll 1988; Waltham 2016). *A. transversa* in particular is extremely tolerant of dehydration, and is capable losing up to 45% of its total body water during the dry season. During these dry periods, *A. transversa* subsists solely on burrow condensation and metabolic water (Greenaway & MacMillen 1978; MacMillen & Greenaway, 1978).

Highly terrestrial land crabs are also quite setaceous. The densely distributed setae on their legs passively wick water up from moist substrates, allowing some species to rehydrate using only dew, mist, and moisture from damp sand. The ocypodid crabs (TP:1b) also have dense tufts of setae on the coxae of the third and fourth pairs of walking legs that have a comparable function (Becchi et al. 2015). Ocypodid crabs, for example, are able to "sit" on damp sand, drawing water into their branchial chambers via setal capillary action; some of this fluid is also passed to the mouth and swallowed (Bliss 1968; Wolcott 1984). Although in *G. natalis* this system of setae extends into the branchial chamber and lies densely across the posterior gills (Farrelly & Greenaway 1993), which may imply that the setae play some role in moistening the gills and the branchiostegal lung membranes, Wolcott (1984) found that a confamilial species was less capable of extracting water in bulk from the soil, suggesting that these species may instead rely on conservation to tackle the challenge of water balance in terrestrial environments instead of bulk water uptake from the substrate.

All crabs in the gecarcinid genera *Discoplax* and *Cardisoma* (TP:1b), as well as some *Sesarma* and *Geograpsus* species (McLay & Ryan 1990; TP:2b) also possess densely-packed patches of short setae on the pterygostomian region of the carapace. These setae may be used to absorb atmospheric moisture into the branchial chamber. As for the terrestrial anomuran TP:1a crabs, coenobitid hermit crabs also possess dense setal tufts on their claws that passively extract

water from damp sand, and they also carry reserves of water within their shells that they periodically refresh using their claws and setaceous maxillipeds (Becchi et al. 2015).

5.5 Ecdysis

As arthropods, all crabs must periodically shed and regrow their rigid exoskeleton in order to grow larger in a process known as ecdysis. As a new exoskeleton develops, the old one is partially enzymatically degraded before being cast away (Pechenik 2015). The crab then takes in a significant amount of water to inflate its body, while the new skeleton is still soft, and then waits for the new skeleton to harden. Once hardened, the crab releases this water and its tissues return to their pre-molt volumes. During the intermolt phase, the crab grows until this process begins anew. Thus, while biomass accumulation in arthropods is steady, their overall size increases discretely.

On land, crabs face two major challenges associated with exoskeletons and incremental growth. First, this mode of growth requires substantial amounts of water, and second, terrestrial environments lack readily available calcium and magnesium, which are essential to mineralizing the new skeleton after molting. Highly terrestrial crabs in TP:1b and TP:2b confront these challenges with two distinctive adaptations: (1) utilizing a pneumo-hydrostatic intermediary skeleton during ecdysis (Taylor & Kier 2006), and (2) storing large amounts of ions in either the pericardial sacs or as gastroliths in the gastric mill prior to molting. The presence of these two traits is therefore the fifth metric for determining a species' relative terrestrial grade. However, apart from what little is known about *Birgus latro*'s molting habits, even less is currently understood about molting in other coenobitid crabs in TP:1a. Most of the process appears to be completed within the shell, and immediately after molting, the exuviae are positioned at the

mouth of the shell while the remainder of the body is molded to fit the shell (Greenaway 2003) (Table 2).

In aquatic crabs (both marine and freshwater), molting is mediated by the presence of a hydrostatic skeleton: without the support of an external skeleton, this switch to a fluid-based skeletal support system is integral to the crabs' survival, as it allows them to maintain some degree of locomotory function until their new skeleton hardens (Taylor & Kier 2006). But land crabs instead rely on a pneumo-hydrostatic skeleton, since water is comparatively scarce in terrestrial habitats. This skeletal mode relies on both compressible air and incompressible liquid to provide skeletal support during molting (Taylor & Kier 2006). In preparation for molting, many terrestrial crabs store water in their gut (*Birgus latro*, see Richardson & Araujo 2015) or in their pericardial sacs (*G. lateralis*, see Bliss 1968). Once they molt, as in aquatic crabs, this water is utilized to partially inflate the new, soft exoskeleton. Since the freshly molted body is not compartmentalized, by then also inflating their flexible pyloric stomach with gas, the pressure of the hemolymph around the body rises, increasing their overall body turgidity (Bliss 1968; Taylor & Kier 2006).

During ecdysis, marine and freshwater crabs also lose the majority of their internal calcium and magnesium stores, but these lost ions are rapidly replenished from the surrounding water (Greenaway 1993). Since the concentrations of readily available ions drops off significantly between marine and terrestrial environments, before molting, land crabs stockpile excess calcium and magnesium from their herbivorous diet and also enzymatically resorb up to 65-75% of these ions from their old exoskeleton (Greenaway 1994; Greenaway & Linton 1995). These ions are stored in the pericardial sacs, the hepatopancreas (some *Sesarma* spp. and *Parathelphusa* spp.), in the hemolymph (as in *A. transversa*) or as gastroliths in the gastric mill

(a unique adaptation seen only in freshwater and terrestrial crabs (Bliss 1968; Greenaway & Farrelly 1991; Greenaway 1993; Richardson & Araujo 2015)). The ions are immediately mobilized after molting to strengthen the new skeleton. Some terrestrial crabs also consume their exuviae after molting to supplement their internal mineral stores. While some marine crabs can also store small amounts of calcium and magnesium between molts, since these minerals are readily available in seawater, the quantity stored is much smaller compared to what has been recorded from freshwater, semi-terrestrial, and terrestrial crabs (Bliss 1968).

5.6 Terrestrial early development

The sixth trait for determining a species' terrestrial grade is early developmental mode. Amphidromy, abbreviated or direct development, accelerated development, or extended development with associated migratory behaviors are all coupled to increasingly terrestrial lifestyles. Because there are significant phylogenetic constraints that determine which adaptation is possible for a given lineage, these variable strategies will be considered as alternative solutions to the terrestrial early development selective challenge (Table 2).

Ancestrally, about 80% of marine and estuarine crabs have a biphasic life cycle with an extended pelagic larval developmental phase and a benthic juvenile-adult phase (Anger et al. 2015). Gravid female crabs carry thousands of small eggs that hatch into free-swimming larvae. After hatching, the pelagic planktotrophic phase lasts between several weeks or months, and in this phase, the larvae pass through several morphologically distinct zoeal stages (four or five in most brachyurans, but up to eight in some Grapsidae (Anger et al. 2015)), and a single megalopa stage. Compared to the swimming zoeal stages, the megalopa stage has fully functional pleopods and pereiopods, which allows them to walk and swim. After the megalopa stage, the crab

undergoes a final molt into a juvenile crab: in males, the first two pairs of posterior swimming appendages are reduced and modified into gonopods, and the anterior swimming appendages become feeding appendages, the maxillipeds. Juvenile crabs then undergo several molts to grow until they reach sexual maturity (Anger et al. 2015). Unsurprisingly, this dependence on water for physical support, transport, and nourishment during development elicits a spectrum of modifications to early developmental mode in the terrestrial crabs.

Land crabs with freshwater origins in TP:2b show a marked shift towards abbreviated or direct development. Because larvae lack both the impermeable cuticle possessed by juvenile and adult crabs and have limited osmoregulatory abilities, they are particularly vulnerable to osmotic stress when exposed to freshwater. Some estuarine species respond to this challenge by exporting their larvae either via adult migration or larval behavior to more saline regions of an estuary (*i.e.,* amphidromy), where the larvae develop normally and return upstream to fresher waters as juvenile and adult crabs (McDowall 2007). Alternatively, the partial abbreviation of the larval stages apparent in freshwater species allows the embryo to develop entirely within the relatively impermeable egg while still remaining in freshwater. This particular adaptation is typical of the freshwater crabs (Ng et al. 2008; Yeo et al. 2008), and it may indeed be a critical preadaptation that allowed the freshwater-derived land crabs to invade land more completely than their saline-derived counterparts.

Some Sesarmidae spp. pass through three or fewer zoeal stages, as they develop in shallow pools of free-standing freshwater (Soh 1969). Most direct-developing *Geosesarma* spp., however, have suppressed the zoeal stages completely: while encapsulated in the egg, the embryo passes through several morphological stages that resemble the planktonic zoeal stages of other lineages, but it hatches as a fully independent juvenile crab (Ng & Tan 1995). Furthermore,

there is a marked shift in the development of feeding structures in many of these lineages: ancestrally, crab zoeae are planktotrophic and possess functional feeding appendages that they use to filter feed. Abbreviated or direct-developing larvae instead exhibit facultative or obligate lecithotrophy, which corresponds with changes in the relative sizes and quantities of the eggs carried by gravid female crabs (*i.e.,* lecithotrophic eggs tend to be yolkier, larger in size, and fewer in number). A few of these species also exhibit some degree of maternal care. For example, in the Jamaican sesarmid land crabs, mothers clean, ventilate, and feed their young while they develop in small pools of water in the axils of bromeliad plants (as in *Metopaulias depressus* (Diesel & Schubart 2007)), or in abandoned snail shells (as in *Sesarma jarvisi* (Diesel & Horst 1995)). In the Indo-Pacific species *Geosesarma notophorum*, mothers carry their newly hatched juvenile crabs on their backs and irrigate them with a steady stream of water for about a week after they hatch (Ng & Tan 1995).

In TP:1a and TP:1b crabs, there may instead be a shift towards accelerated development, with complete development to the megalopa stage taking a fraction of the time it does ancestrally, potentially over a period of a few days instead of over several weeks or months (as in *Uca subcylindrica* (Greenaway 2003; Anger et al. 2015)). Alternatively, the transition on to land may instead involve the evolution of a larval export strategy, where male and female adult crabs annually migrate to the sea to mate and release developing larvae into the ocean. These larvae then undergo extended pelagic larval development before returning to land as independent juvenile crabs (Greenaway 2003; Anger et al. 2015). Both the adult and juvenile migrations are conspicuous and dramatic, and they are in fact renowned as one of the wonders of the natural world.

6. The Six Grades of Terrestriality Framework

Grade I crabs inhabit the lower intertidal zone or estuaries and are only periodically emersed during low tides; the majority of their time is spent immersed. While out of water, they are typically inactive. In order to survive in these habitats, they must be able to tolerate variable environmental conditions. The key innovation at this stage was likely a shift in osmoregulatory status from the ancestral condition (stenohaline or euryhaline osmoconformation) to weak osmoregulation for TP:1a and TP:1b crabs, or facultative strong osmoregulation for TP:2b crabs. At this grade, TP:2b crabs may also display amphidromy, and export their larvae to more saline estuarine regions to develop.

Grade II land crabs inhabit either the upper intertidal zone (TP:1a and TP:1b) or full freshwater environments during adulthood (TP:2b). While exposed in the intertidal zone during low tides, TP:1a and TP:1b crabs are typically active, foraging, fleeing, and hiding from predators. To tolerate the high environmental variability of the upper intertidal zone or the dramatic change in salinity in freshwater habitats, the key innovations for crabs at this stage are further shifts in osmoregulatory status. For TP:1b crabs, the development of facultative strong osmoregulation facilitates this habitat shift; for TP:2b crabs, the development of obligate strong osmoregulatory capabilities plays a similar role. Of particular note is the evolution of lecithotrophic larvae with abbreviated development in the TP:2b freshwater crabs.

Grade III land crabs typically inhabit riverbanks or beaches. Two critical adaptive innovations occur at this grade in all three TPs, principally, an increase in air-breathing capabilities

associated with the evolution of some type of branchiostegal lung or rebreather system, and obligate fossoriality with wet burrows. For TP:1b and TP:2b crabs, “derived ammoniotely” (defined in 5.2 above) may also have been a key innovation. For TP:1b beach-dwelling crabs, the evolution of obligate strong osmoregulatory abilities and a dramatic increase in the density of water-wicking setae across the body may also have accompanied this grade.

Grade IV land crabs inhabit coastal forests and tropical jungles. Because of the lower availability of standing water and readily available critical ions in these habitats, the development of a pneumo-hydrostatic skeletal intermediate during ecdysis for TP:1b crabs and the ability to store large amounts of calcium and magnesium between molts for TP:1b and TP:2b crabs represent critical innovations. The acceleration of the larval developmental stages for some TP:1b and TP:1a crabs may have also been a key event. Finally, the evolution of annual migratory mating behaviors and nocturnality were also critical changes for crabs across all three TPs.

Grade V crabs from TP:1b and TP:1a also inhabit coastal forests and jungles, but they either dig dry burrows (Burggren & McMahon, 1988) or inhabit dry karst crevices (Greenaway 2003). The decreasing reliance on free-standing water in TP:1b crabs is facilitated by the further evolution of water-wicking tufts of setae on the posterior coxae or the pterygostomial region, which allows them to rehydrate themselves using dew, puddles, mist and other ephemeral sources of water. Extended periods of full immersion appear to be fatal for many of these species (Orchard 2012). Both TP:1b and TP:1a terminate at this stage: *Gecarcoidea natalis* and similar land crabs represent the most terrestrial marine-derived brachyuran land crab lineages, and *B. latro*

represents the most terrestrially adapted anomuran crab lineage (at this grade, the lineage leading to modern *B. latro* alone developed uricotely, which further decreased its reliance on free-standing water (Linton et al. 2017)). TP:2b crabs at this stage evolved lecithotrophic, direct-developing larvae, which further reduced their reliance on water. Some species of TP:2b land crabs at this grade also exhibit some degree of maternal care.

Grade VI: Very few land crabs exist at this stage. These crabs are capable of inhabiting arid environments, which is made possible only by their ability to estivate during long periods of drought. The freshwater derived land crabs at this stage, such as *Austrothelphusa transversa* and *Parathelphusa guerini*, arguably represent the most terrestrially adapted crabs on Earth (Greenaway & MacMillen 1978).

6. Discussion

There are apparently many different ways to be a "land crab", and there are several distinct ways that different crab lineages have responded to the key selective challenges they most likely encountered while colonizing terrestrial habitats. The many independent origins of terrestriality and the high degree of functional and phenotypic convergence in this system make it especially promising for exploring the nature of evolutionary predictability and its genomic underpinnings. With groundbreaking developments in high-throughput whole genome and transcriptome sequencing, we now have the tools to answer the deepest questions about the predictability of evolution and the genetic basis of major biological innovations and transitions, but reconceptualizing our fundamental understanding of the evolution of terrestriality and its associated adaptive traits is a critical first step in that process. Which adaptive traits are critical to

living on land? Do any of the traits appear to be associated with one other, and are they associated with particular terrestrial habitats or conditions? How many times have particular suites of traits evolved independently across the Tree of Life, and how do divergent terrestrial lineages solve similar selective challenges? What evolutionary rules govern how successful (in terms of species diversity) a particular lineage is when transitioning from marine to terrestrial habitats? Understanding the evolution of convergently evolved terrestriality in the land crabs can move us closer to more clearly defining these critical evolutionary rules.

Hartnoll's original classification of the grades of terrestriality in the land crabs (Hartnoll 1988) was fundamentally important, and his rankings are still widely used today to organize an evolutionarily complex system that seems to otherwise defy categorization. But by focusing primarily on one major trait (*i.e.,* reliance on standing water) and acknowledging critical phylogenetic distinctions and the ecological route-dependency of this system only in passing, the previous ranking left an opportunity to more rigorously describe the multivariate nature of terrestriality and the underlying phylogenetic constraints that determine which phenotypes may be possible for given lineages.

By designating four Transition Pathways (TP) based on deep evolutionary relationships between brachyurans and anomurans, and incorporating route-dependence in my classification of the key biological, physiological, behavioral, and ecological adaptive traits observed in this system, I have aimed to develop a more comprehensive trait-by-environment framework that can now be mapped phylogenetically to categorize the many land crab clades. It is important to reiterate that because of the many independent evolutionary origins of terrestriality in the land crabs, the TPs are not meant to represent monophyletic groupings. However, categorizing the lineages by the route most likely taken onto land is still a useful exercise as it provides some

context for understanding how suites of adaptive traits may have evolved in response to particular environment-specific selective forces, and whether specific lineages may have possessed key pre-adaptations that led to greater degrees of terrestrial adaptation.

The Six Grades of Terrestriality, which are ordered along a hypothetical environmental gradient, represent a coherent, but perhaps incomplete, framework for how these key traits may have evolved in this system across the four Transition Pathways. Future studies combining phylogenetic relationships and trait data from additional terrestrial lineages are essential for refining the sequence of trait evolution presented in the Six Grades of Terrestriality. Furthermore, since many of the traits I have identified appear to be convergently evolved across currently recognized lineages, in future studies, transcriptomic and genomic data should be used to determine if these traits display any degree of molecular parallelism, and if so, at which level of biological organization this convergence is observed.

Ultimately, understanding the evolutionary and genetic circumstances that may have made it possible for land crabs to repeatedly colonize terrestrial environments will help to reveal how these, and potentially other, marine animals might have done the same throughout the history of life. Comparable analyses in other marine-derived invertebrate lineages (*e.g.* terrestrial mollusks and crustaceans) will greatly extend our understanding about adaptive breakthroughs that made living on land possible for these species. Since sea-to-land transitions are rare and present major evolutionary challenges (Vermeij & Dudley 2000), understanding the successes of the land crabs may also provide critical insights into the phylogenetic, ecological, physical, and physiological constraints that explain why this transition is rare and challenging for other lineages across the Tree of Life.

Acknowledgements:

V.M.W.Z. wishes to thank Richard Grosberg, Ph.D., Rachael Bay, Ph.D., Timothy Caro, Ph.D., Geerat Vermeij, Ph.D., and Joanna Wolfe, Ph.D. for their detailed comments and suggestions for improvement, and Peter K.L. Ng, Ph.D. for his invaluable insights during the conceptual phase of this manuscript. She also wishes to thank Joshua Zink, M.A., for his unwavering support throughout the completion of this manuscript.

Funding: This work was supported by a National Science Foundation Graduate Research Fellowship, a National Science Foundation East Asia and Pacific Summer Institutes Fellowship, and National Science Foundation grants to R. K. Grosberg.

LITERATURE CITED

Amesbury, S.S., 1980. Biological studies on the coconut crab (*Birgus latro*) in the Mariana Islands. University of Guam Technical Report;17.

Anger, K., Queiroga, H., and Calado, R., 2015. Larval development and behavior strategies in Brachyura. The Crustacea, Vol. 9C, Pages 317 – 374.

Atkinson, R.J.A. and Eastman, L.B., 2015. Burrow dwelling in Crustacea. In: The Natural History of Crustacea, Vol. 2: Lifestyles and Feeding Biology, Oxford University Press, Oxford, pp. 78 – 117.

Atkinson, R.J.A., and Taylor, A.C., 1988. Aspects of decapod crustacean biology. Symposia of Zoological Society London, 59. Clarendon Press, Oxford.

Becchi, C., Innocenti, G., and Vannini, M., 2015. Water uptake in terrestrial hermit crabs: a morpho-functional analysis. Italian Journal of Zoology, Vol. 82:2, Pages 157-164.

Bliss, D.E., 1968. Transition from Water to Land in Decapod Crustaceans. American Zoologist, Volume 8, Issue 3, Pages 355–392.

Bond-Buckup, G., Jara, C.G., Pérez-Losada, M., Buckup, L., & Crandall, K.A., 2008. Global diversity of crabs (Aeglidae: Anomura: Decapoda) in freshwater. Hydrobiologia, 595(1), 267–273.

Broly, P. and Maillet, S., 2013. The origin of terrestrial isopods (Crustacea: Isopoda: Oniscidea). Evolutionary Ecology 27: 461-476.

Brookfield, M. E., Catlos, E. J., & Suarez, S. E., 2020. Myriapod divergence times differ between molecular clock and fossil evidence: U/Pb zircon ages of the earliest fossil millipede-bearing sediments and their significance. Historical Biology, 1–5.

Burggren, W.W. and McMahon, B.R., 1988. Biology of the Land Crabs. Cambridge University Press, Cambridge.

Burggren, W.W., 1992. Respiration and circulation in land crabs: novel variations on the marine design. American Zoologist, Vol. 32, Pages 417 – 427.

Carmona, N.B., Buatois, L.A., & Mángano, M.G., 2004. The trace fossil record of burrowing decapod crustaceans: evaluating evolutionary radiations and behavioural convergence. Fossils and Strata, 51, 141–153.

Chou, A., Lin, C., & Cronin, T.W., 2020. Visual metamorphoses in insects and malacostracans: Transitions between an aquatic and terrestrial life. Arthropod Structure & Development, Volume 59, 100974.

Copilaş-Ciocianu, D., Borko, S., & Fišer, C., 2020. The late blooming amphipods: Global change promoted post-Jurassic ecological radiation despite Palaeozoic origin. Molecular Phylogenetics and Evolution, 143.
Dehnel, P.A., 1962. Aspects of osmoregulation in two species of intertidal crabs. Biological Bulletin, Vol. 122, No. 2, pp. 208-227.

Díaz, H. and Rodríguez, G., 1977. The Branchial Chamber in Terrestrial Crabs: A Comparative Study. Biological Bulletin, 485 – 504.

Diesel, R.D., and Horst, 1995. Breeding in a snail shell: Ecology and biology of the Jamaican montane crab Sesarma jarvisi (Decapoda: Grapsidae). Journal of Crustacean Biology, Vol. 15 (1), 179 – 195.

Diesel, R.D., and Schubart, C.D., 2007. The Social Breeding System of the Jamaican Bromeliad Crab Metopaulias depressus. In: Evolutionary Ecology of Social and Sexual Systems. Oxford University Press, Oxford.

Engel, M.S. and Grimaldi, D.A., 2004. New light shed on the oldest insect. Nature, 427(6975): 627-30.

Farrelly, C.A., and Greenaway, P., 1992. Morphology and ultrastructure of the gills of terrestrial crabs (Crustacea, Gecarcinidae and Grapsidae): Adaptations for air-breathing. Zoomorphology, (1992) 112: 39-49.

Farrelly, C.A., & Greenaway, P., 1993. Land crabs with smooth lungs: Grapsidae, Gecarcinidae and Sundathelphusidae ultrastructure and vasculature. Journal of Morphology, 215:245-260.

Farrelly, C.A., & Greenaway, P.,1994. Gas exchange through the lungs and gills in air-breathing crabs. Journal of Experimental Biology, 187:113-130.

Farrelly, C.A., & Greenaway, P., 2005. The morphology and vasculature of the respiratory organs of terrestrial hermit crabs (Coenobita and Birgus): gills, branchiostegal lungs and abdominal lungs. Arthropod Structure and Development, 34:63-87.

Felgenhauer, B.E. and Abele, L.G., 1983. Branchial Water Movement in the Grapsid Crab Sesarma reticulatum, Say. Journal of Crustacean Biology, Vol. 3, No. 2, Pages 187-195.

Fernández, R., Edgecombe, G.D. & Giribet, G, 2018. Phylogenomics illuminates the backbone of the Myriapoda Tree of Life and reconciles morphological and molecular phylogenies. Scientific Reports, 8, 83.

Gratwicke, B., 2004. Migration of the freshwater crab Potamonautes unispinus in a seasonal stream, Zimbabwe. African Zoology, 39(1): 00–00.

Greenaway, P., 1993. Calcium and magnesium balance during molting in land crabs. Journal of Crustacean Biology, 13:191-197.

Greenaway, P., 1994. Salt and water balance in field populations of the terrestrial crab Gecarcoidea natalis. Journal of Crustacean Biology, Vol. 14 (3), Pages 438 – 453.

Greenaway, P., 1998. Physiological diversity and the colonization of land. Proceedings of the Fourth International Crustacean Congress, Pages 823 – 842.

Greenaway, P., 2003. Terrestrial adaptations in the Anomura (Crustacea: Decapoda). Memoirs of Museum Victoria, 60(1), 13–26.

Greenaway, P., Bonaventura, J., & Taylor, H.H., 1983. Aquatic Gas Exchange in the Freshwater/Land Crab, Holthuisana transversa. Journal of Experimental Biology 103:225-236.

Greenaway, P., & Farrelly, C.,1991. Trans-epidermal Transport and Storage of Calcium in Holthuisana transversa (Brachyura; Sundathelphusidae) During Premoult. Acta Zoologica, 72:29-40.

Greenaway, P., and Linton, S.M., 1995. Dietary Assimilation and food retention time in the herbivorous terrestrial crab, Gecarcoidea natalis. Physiological Zoology, 68 (6): 1006-1028.

Greenaway, P., & MacMillen, R.E., 1978. Salt and Water Balance in the Terrestrial Phase of the Inland Crab Holthuisana (Austrothelphusa) transversa Martens (Parathelphusoidea: Sundathelphusidae). Physiological Zoology, 51:217-229.

Greenaway, P., and Nakamura, T., 1991. Nitrogenous excretion in two terrestrial crabs Gecarcoidea natalis and Geograpsus grayi. Physiological Zoology, 64:767–786.

Greenaway, P., & Taylor, H., 1976. Aerial gas exchange in Australian arid-zone crab, Parathelphusa transversa Von Martens. Nature 262, 711-713.

Grosberg, R.K., Vermeij, G.J., and Wainwright, P.C., 2012. Biodiversity in water and on land. Current Biology, 22 (21): R900-R903.

Hartnoll, R.G., 1988. Evolution, Systematics, and Geographical Distribution. In: Burggren, W.W. and McMahon, B.R. (Eds). Biology of the Land Crabs. Cambridge University Press, Cambridge, pp. 6-53.

Krieger, J., Braun, P., Rivera, N.T., Schubart, C.D., Müller, C.H., & Harzsch, S., 2015. Comparative analyses of olfactory systems in terrestrial crabs (Brachyura): evidence for aerial olfaction? PeerJ, 3:e1433.

Lignot, J.H., and Charmantier, G., 2015. Osmoregulation and Excretion. In: The Natural History of Crustacea, Vol. 4: Physiology, Oxford University Press, Oxford, pp. 249 -284.

Linton, S. M., Wilde, J.E., and Greenaway, P., 2005. Excretory and Storage Purines in the Anomuran Land Crab *Birgus latro*; Guanine and Uric Acid. Journal of Crustacean Biology, Volume 25, Issue 1, Pages 100–104.

Linton, S.M., Wright, J.C., and Howe, C.G., 2017. Nitrogenous Waste Metabolism Within Terrestrial Crustacea, with Special Reference to Purine Deposits and Their Metabolism. In: Acid-Base Balance and Nitrogen Excretion in Invertebrates, Springer International Publishing, Switzerland, pp. 249 -284.

Ma, K.Y., Qin, J., Lin, C.W., Chan, T.Y., Ng, P.K.L., Chu, K.H., & Tsang, L.M. 2019. Phylogenomic analyses of brachyuran crabs support early divergence of primary freshwater crabs. Molecular Phylogenetics and Evolution, 135, Pgs. 62 – 66.

MacMillen, R.E., & Greenaway, P., 1978. Adjustments of Energy and Water Metabolism to Drought in an Australian Arid-Zone Crab. Physiological Zoology 51:230-240.

May, R.M. (1994) Biological diversity: differences between land and sea. Philosophical Transactions of the Royal Society of London B., 343: 105-111.

Mantel, L.H., and Farmer, L.L., 1983. Osmotic and Ionic Regulation. In: The Biology of Crustacea, Vol. 5. Academic Press, Inc., pp. 53 – 161.

McCann, C., 1938. Notes on the common land crab Parathelphusa (Barytelphusa) guerini (M. Edw.) of Salsette Island. Journal of the Bombay Natural History Society, Vol. 39: 531 – 542.

McDowall, R.M., 2007. On amphidromy, a distinct form of diadromy in aquatic organisms. Fish and Fisheries, Vol. 8, Pages 1 – 13.

McLaughlin, P.A., and Murray, T. 1990. Clibanarius Fonticola, New Species (Anomura: Paguridea: Diogenidae), from a Fresh-Water Pool on Espiritu Santo, Vanuatu. Journal of Crustacean Biology, Vol. 10, Issue 4, Pages 695–702.

McLay, C.L., and Ryan, P.A., 1990. The terrestrial crabs Sesarma (Sesarmops)impressum and Geograpsus crinipes (Brachyura,Grapsidae, Sesarminae) recorded from the Fiji Is. Journal of the Royal Society of New Zealand, Vol. 20 (1), Pages 107-118.

McMahon, B.R., and Burggren, W.W., 1979. Respiration and adaptation to the terrestrial habitat in the terrestrial hermit crab, Coenobita clypeatus. Journal of Experimental Biology, Vol. 79, Pages 265 – 281.

Morris, S., 2002. The ecophysiology of air-breathing in crabs with special reference to Gecarcoidea natalis. Comparative Biochemistry and Physiology B, 131:559-570.

Morris, S., and Ahern, M.D., 2003. Regulation of urine reprocessing in the maintenance of sodium and water balance in the terrestrial Christmas Island red crab Gecarcoidea natalis investigated under field conditions. Journal of Experimental Biology, 206, 2869 – 2881.

Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H., and Donoghue, P.C.J., 2018. Timescale of early land plant evolution.

Proceedings of the National Academy of Sciences, 115 (10) E2274-E2283; DOI: 10.1073/pnas.1719588115.

Ng, P.K.L., and Davie, P., 2012. The Blue Crab of Christmas Island, Discoplax celeste sp. nov. (Crustacea: Decapoda: Brachyura: Gecarcinidae). The Raffles Bulletin of Zoology, 60(1):89-100.

Ng, P.K.L., Guinot, D., and Davie, P.J.F., 2008. Systema Brachyuroum: Part 1. An annotated checklist of extant brachyuran crabs of the world. Raffles Bulletin of Zoology, Vol. 17, Pages 1 – 286.

Ng, P.K.L., and Tan, C.G.S., 1995. Geosesarma notophorum sp. nov. (Decapoda, Brachyura, Grapsidae, Sesarminae), A terrestrial crab from Sumatra, with novel brooding behaviour. Crustaceana, 68 (3).

O'Mahoney, P.M., and Full, R. J., 1984. Respiration of crabs in air and water. Comparative Biochemistry and Physiology, Vol. 79A, No. 2, 275-282.

Orchard, M. 2012. Crabs of Christmas Island. 1st ed. Christmas Island, Indian Ocean: Christmas Island Natural History Association.

Pechenik, J.A., 2015. Biology of the Invertebrates, Seventh Edition. McGraw Hill Education.

Powers, L.W. and Bliss, D.E., 1983. Terrestrial Adaptations. In: The Biology of Crustacea, Vol. 8. Pages 271 – 333. Academic Press, Inc.

Rebach, S. and Dunham, D., 1983. Studies in adaptation: The behavior of higher Crustacea. New York: Wiley Interscience.

Richardson, A. and Araujo, P.B., 2015. Lifestyles of Terrestrial Crustaceans. In: The Natural History of Crustacea, Vol. 2: Lifestyles and Feeding Biology. Oxford University Press, Oxford, pp. 299 – 336.

Shubin, N.H., Daeschler, E.B., and Jenkins, F.A., 2013. Pelvic girdle and fin of Tiktaalik roseae. Proceedings of the National Academy of Sciences, 111(3):893-9.

Soh, L.C., 1969. Abbreviated development of non-marine crab, Sesarma (Geosesarma) perracae (Brachyura: Grapsidae), from Singapore. Journal of the Zoological Society of London, Vol. 158, 357 – 370.

Speer, B.R.,1995. The Devonian Period (http://www.ucmp.berkeley.edu/devonian/devonian.php).

Stensmyr M.C., Erland, S., Hallberg, E., Wallen, R., Greenaway, P. and Hansson, B.S., 2005. Insect-like olfactory adaptations in the terrestrial giant robber crab. Current Biology, Vol. 15, 116–121.

Taylor, J.R.A., and Kier, W.M., 2006. A pneumo-hydrostatic skeleton in land crabs. Nature, Vol. 440, 1005.

Taylor, J. R.A., 2018. Aquatic versus terrestrial crab skeletal support: morphology, mechanics, molting and scaling. Journal of Experimental Biology, 221.

Tsang, L.M, Schubart, C.D., Ahyong, S.T., Lai, J.C.Y., Au, E.Y.C., Chan, T.Y., Ng, P.K.L., and Chu, K.H., 2014. Evolutionary History of True Crabs (Crustacea: Decapoda: Brachyura) and the Origin of Freshwater Crabs. Molecular Biology and Evolution, 31 (5), 1173–1187.

Vermeij, G.J. and Dudley, R., 2000. Why are there so few evolutionary transitions between aquatic and terrestrial ecosystems? Biological Journal of the Linnean Society, 70: 541 -554.

Waltham, N.J., 2016. Unravelling life history of the Inland Freshwater Crab Austrothelphusa transversa in seasonal tropical river catchments. Australian Zoologist, Vol. 38:2, 217 – 222.

Warner, G. F., 1977. The biology of crabs. London: HarperCollins Distribution Services.

Weihrauch, D., Morris, S, and Towle, D.W., 2004. Ammonia excretion in aquatic and terrestrial crabs. Journal of Experimental Biology, 207, 4491-4504.

Wolcott, T.G., 1984. Uptake of Interstitial Water from Soil: Mechanisms and Ecological Significance in the Ghost Crab Ocypode quadrata and Two Gecarcinid Land Crabs. Physiological Zoology, 57:161-184.

Wolcott, T.G., and Wolcott, D.L., 1990. Ion conservation by reprocessing of urine in the land crab Gecarcinus lateralis (Freminville). Physiological Zoology, Vol. 64, No. 1, Pages 344 – 361.

Wolfe, J.M., Breinholt, J.W., Crandall, K.A., Lemmon, A.R., Lemmon, E.M., Timm, L.E., Siddall, M.E., & Bracken-Grissom, H.D., 2019. A phylogenomic framework, evolutionary timeline and genomic resources for comparative studies of decapod crustaceans. Proceedings of the Royal Society B., 286 (1901), 20190079–.

Yeo, D.C.J., Ng, P.K.L., Cumberlidge, N., Magalhães, C., Daniels, S.R., and Campos, M.R., 2008. Global diversity of crabs (Crustacea: Decapoda: Brachyura) in fresh water. Hydrobiologia, Vol. 595, 275 – 286.

TABLES AND FIGURES

Infraorder	Section	Family	Transition Pathways
Anomura	Paguroidea	Coenobitidae	1a
Anomura	Paguroidea	Diogenidae	2a
Anomura	Aegloidea	Aeglidae	2a
Brachyura	Thoracotremata	Ocypodidae	1b/2b
Brachyura	Thoracotremata	Grapsidae	1b
Brachyura	Thoracotremata	Gecarcinidae	1b
Brachyura	Thoracotremata	Mictyridae	1b
Brachyura	Thoracotremata	Varunidae	1b/2b
Brachyura	Thoracotremata	Sesarmidae	1b/2b
Brachyura	Heterotremata	Pseudothelphusidae	2b
Brachyura	Heterotremata	Gecarcinucidae	2b
Brachyura	Heterotremata	Potamidae	2b

Table 1.1 Taxonomy and transition pathway in 12 families of decapod crabs with terrestrial representatives. Key: TP:1a – Anomuran crabs that transitioned into terrestrial habitats via marine environments. TP: 1b – Brachyuran crabs that transitioned into terrestrial habitats via marine environments. TP:2a – Anomuran crabs that transitioned into terrestrial habitats via freshwater environments. TP:2b – Brachyuran crabs that transitioned into terrestrial habitats via freshwater environments.

Selective Challenge	(TP:1a) Anomuran crabs that colonized land via marine environments	(TP:1b) Brachyuran crabs that colonized land via marine environments	(TP:2b) Brachyuran crabs that colonized land via freshwater environments
Osmotic and Ionic Regulation	Weak osmoregulation; Obligate strong osmoregulation (Greenaway 2003)	Weak osmoregulation (Lignot & Charmantier 2015); Facultative strong osmoregulation (Dehnel 1962); Obligate strong osmoregulation (Lignot & Charmantier 2015)	Facultative strong osmoregulation; Obligate strong osmoregulation (Lignot & Charmantier 2015)
Nitrogenous Waste Excretion	Ancestral ammoniotely (Weihrauch et al. 2004); Unknown (Greenaway 2003); Uricotely (Linton et al. 2017)	Ancestral ammoniotely; Derived ammoniotely (Weihrauch et al. 2004)	Ancestral ammoniotely; Derived ammoniotely (Linton et al. 2017)
Aerial Respiration	Ancestral gills; Dorsal respiratory membrane; Enlarged branchial chamber and complex branchiostegal lung (Greenaway 2003)	Ancestral gills (Farrelly & Greenaway 1992); Slightly enlarged branchial chamber and complex branchiostegal lung (Diaz & Rodriguez 1977); Greatly enlarged branchial chamber and simple branchiostegal lung (Diaz & Rodriguez 1977; Burggren 1992)	Ancestral gills (Farrelly & Greenaway 1992); Complex branchiostegal lung without enlarged branchial chamber; Greatly enlarged branchial chamber and simple branchiostegal lung (Diaz & Rodriguez 1977; Burggren 1992); Rebreather system (Felgenhauer & Abele 1983)
Desiccation and Thermoregulation	Obligate fossoriality (Atkinson & Eastman 2015); Nocturnality (Greenaway 2003); Water-wicking setae (Becchi et al. 2015; McLay & Ryan 1990)	Obligate fossoriality (Atkinson & Eastman 2015); Nocturnality; Water-wicking setae (Bliss 1968; Becchi et al. 2015)	Obligate fossoriality (Atkinson & Eastman 2015); Nocturnality (Bliss 1968; Becchi et al. 2015); Water-wicking setae (McLay & Ryan 1990); Estivation (Hartnoll 1988; McCann 1938; Waltham 2016; MacMillen & Greenaway 1978)

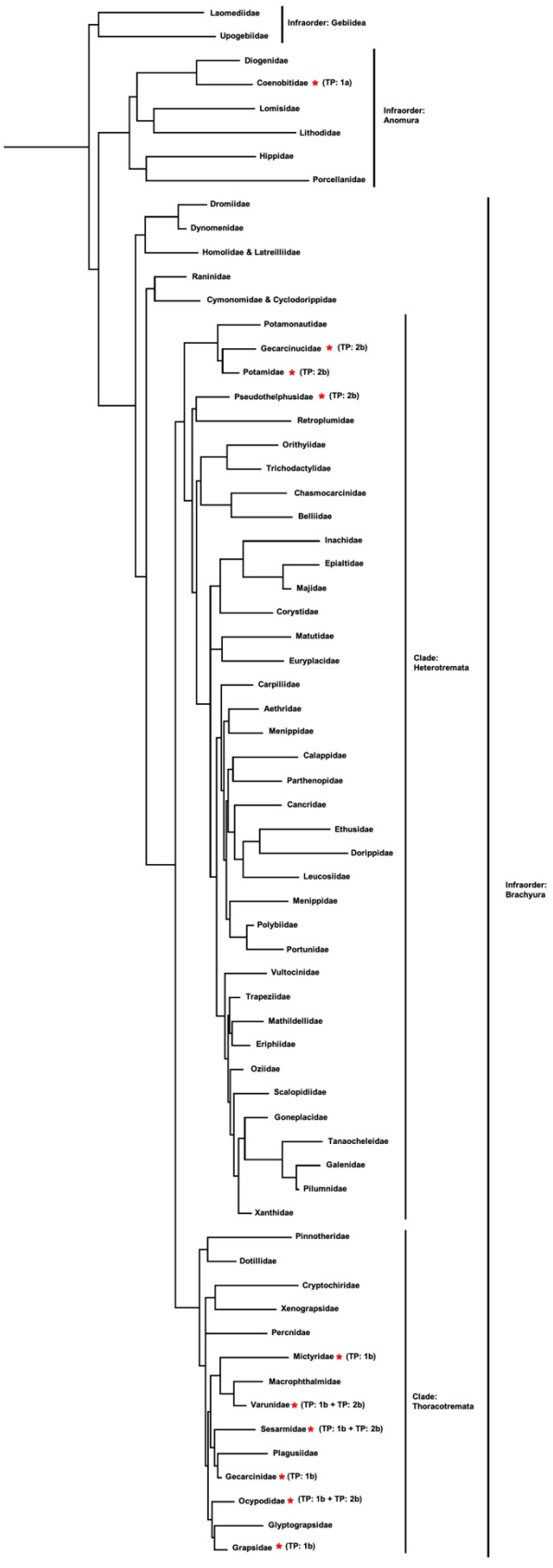
Ecdysis	Hydrostatic skeleton (Pechenik 2015); Unknown (Greenaway 2003); Exuviae consumption (Greenaway 2003)	Hydrostatic skeleton (Pechenik 2015); Pneumo-hydrostatic skeleton (Taylor & Kier 2006); Ion storage during intermolt phase (Bliss 1968; Richardson & Araujo 2015); Exuviae consumption (Bliss 1968)	Hydrostatic skeleton (Pechenik 2015); Ion storage during intermolt phase (Bliss 1968; Richardson & Araujo 2015); Exuviae consumption (Bliss 1968)
Terrestrial Early Development	Extended planktonic development; Accelerated development; Migratory larval export strategy (Greenaway 2003)	Extended planktonic development; Accelerated development; Migratory larval export strategy (Anger et al. 2015)	Amphidromy (McDowall 2007); Migratory larval export strategy (Gratwicke 2004); Abbreviated development (Soh 1969); Direct development (Ng & Tan 1995)

Table 1.2 Adaptive responses to the six primary selective challenges across land crabs classified according to the three transition pathways (TP) in brachyuran and anomuran decapods.

Six Grades of Terrestriality Framework						
	Grade I	**Grade II**	**Grade III**	**Grade IV**	**Grade V**	**Grade VI**
Environmental gradient of increasing desiccation stress	**Lower Intertidal/Estuaries**	**Upper Intertidal/ Freshwater**	**Beaches/ Riverbanks**	**Coastal Forests/Tropical Jungles**		**Arid Zone**
(TP:1a) Anomuran crabs that colonized land via marine environments	Weak osmoregulation		Alternative respiratory structures; Obligate fossoriality with wet burrows	Accelerated planktonic larval development (some species); Migratory larval export strategy; Nocturnality	Dry burrows; Uricotely	Phenotype not observed
(TP:1b) Brachyuran crabs that colonized land via marine environments	Weak osmoregulation	Facultative strong osmoregulation	Alternative respiratory structures; Obligate fossoriality with wet burrows; Derived ammoniotely; Obligate strong osmoregulation; Water-wicking setae	Pneumo-hydrostatic skeletal intermediate during ecdysis; Intermolt ion storage; Accelerated planktonic larval development (some species); Migratory larval export strategy; Nocturnality	Fossoriality with dry burrows; posterior coxal setal tufts or pterygostomial setae; Extended full-immersion fatal	Phenotype not observed
(TP:2a) Anomuran crabs that colonized land via freshwater	Trait analyses not conducted for this phenotype.					
(TP:2b) Brachyuran crabs that colonized land via freshwater environments	Facultative strong osmoregulation; Amphidromy	Obligate strong osmoregulation; Lecithotrophic larvae with abbreviated development	Alternative respiratory structures; Obligate fossoriality with wet burrows; Derived ammoniotely	Intermolt ion storage; Migratory larval export strategy; Nocturnality	Lecithotrophic larvae with direct development; Some maternal care (some species)	Estivation

Table 1.3 Six Grades of Terrestriality Model displaying trait-by-habitat associations across the extant lineages of land crabs from all observed Transition Pathways (TP).

Fig. 1.1 Evolutionary relationships between major decapod families. Starred (*) families contain terrestrial representatives, and each family's respective transition pathway (TP) is indicated in parenthesis. *Key:* TP:1a – Anomuran crabs that transitioned into terrestrial habitats via marine environments. TP:1b – Brachyuran crabs that transitioned into terrestrial habitats via marine environments. TP:2b – Brachyuran crabs that transitioned into terrestrial habitats via freshwater environments. (Modified with permission from Tsang et al. 2014, *MBE*).


Laomediidae
Upogebiidae
Infraorder: Gebiidea
Diogenidae
Coenobitidae (TP: 1a)
Lomisidae
Lithodidae
Hippidae
Porcellanidae
Infraorder: Anomura
Dromiidae
Dynomenidae
Homolidae & Latreilliidae
Raninidae
Cymonomidae & Cyclodorippidae
Potamonautidae
Gecarcinucidae (TP: 2b)
Potamidae (TP: 2b)
Pseudothelphusidae (TP: 2b)
Retroplumidae
Orithyiidae
Trichodactylidae
Chasmocarcinidae
Belliidae
Inachidae
Epialtidae
Majidae
Corystidae
Matutidae
Euryplacidae
Carpiliidae
Aethridae
Menippidae
Calappidae
Parthenopidae
Cancridae
Ethusidae
Dorippidae
Leucosiidae
Menippidae
Polybiidae
Portunidae
Vultocinidae
Trapeziidae
Mathildellidae
Eriphiidae
Oziidae
Scalopidiidae
Goneplacidae
Tanaocheleidae
Galenidae
Pilumnidae
Xanthidae
Clade: Heterotremata
Pinnotheridae
Dotillidae
Cryptochiridae
Xenograpsidae
Percnidae
Mictyridae (TP: 1b)
Macrophthalmidae
Varunidae (TP: 1b + TP: 2b)
Sesarmidae (TP: 1b + TP: 2b)
Plagusiidae
Gecarcinidae (TP: 1b)
Ocypodidae (TP: 1b + TP: 2b)
Glyptograpsidae
Grapsidae (TP: 1b)
Clade: Thoracotremata
Infraorder: Brachyura

Unearthing the evolutionary transcriptomics of sea-to-land transitions in decapod crabs

Victoria M. Watson-Zink[1,2†], Richard K. Grosberg[1], Joelle C.Y. Lai[3], & Rachael A. Bay[1]

[1] Department of Evolution and Ecology, University of California, Davis, United States

[2] Stanford University, Department of Biology, United States

[3] National University of Singapore, Department of Biological Sciences, Republic of Singapore

[†]Corresponding Author: Victoria M. Watson-Zink

vmmorgan@ucdavis.edu

1. Introduction

All life began in the oceans, but much of Earth's current macroscopic biodiversity exists on land and in the air (Grosberg et al. 2012). Virtually all of this terrestrial biodiversity can be traced to a few ancestrally marine lineages (*e.g.*, plants, arthropods, vertebrates, and mollusks) that have successfully colonized land and diversified there (Vermeij and Dudley 2000; Vermeij and Watson-Zink 2022). This rarity has been decisive in determining patterns of biodiversity on our planet, and at the same time suggests that the transition from sea to land presents nearly insurmountable physical challenges. The evolutionary pathways to overcoming these challenges are largely unexplored, however, as the genomic tools to answer these foundational questions were not developed until relatively recently. Just as importantly, a suitable model system for exploring these questions was not readily apparent.

Decapod land crabs are an exceptional example of an ancestrally marine lineage that has repeatedly transitioned into terrestrial habitats. At least ten lineages, to varying degrees, have independently colonized land from the sea, either directly via the littoral zone or after having first adapted to freshwater environments, and each lineage has independently evolved unique solutions to the demands of terrestrial life (Watson-Zink 2021). Due to the vast differences between air and water as media, land crabs have evolved a stunning suite of physiological, physical, and behavioral strategies to address several selective pressures they encountered while colonizing terrestrial environments, among these being challenges related to osmotic and ionic regulation (Weihrauch et al. 2004), nitrogenous waste excretion (Bliss et al. 1968), aerial respiration (Morris 2002), desiccation and thermoregulation (Wood et al. 1986), ecdysis (Taylor and Kier, 2006), terrestrial larval development (Anger et al. 2015), visual and olfactory signal

transmission and reception (Stensmyr et al. 2005; Krieger et al. 2015), locomotion (Dunham and Gilchrist, 1988; Herreid and Full, 1988; Taylor 2018), and nutrient acquisition (Greenaway and Linton 1995).

Over the last four decades, many attempts have been made to categorize and classify the stunning degree of trait diversity displayed by terrestrial crabs in this system. Most recently, Watson-Zink (2021) proposed four distinct transition pathways (TPs) onto land that crabs have taken over the past 66 million years that account for deep phylogenetic differences between Brachyura and Anomura, while also incorporating ecological constraints that each lineage would experience if transitioning to terrestrial habitats via marine environments (*e.g.*, lower intertidal zone, mudflats, sandflats, mangrove forests), or via freshwater environments and estuaries. The framework then posited six Grades of Terrestriality that represent observable trait-by-habitat associations that serve as a hypothetical sequence of key trait evolution for crabs in this system. The promise of this new framework lies in the testability of its hypotheses, and how investigating them might inform our understanding of how shared terrestrial traits may have convergently evolved in response to common conditions in this evolutionarily labile system (Table 2.1).

The high degree of functional and ecological convergence displayed by land crabs in this system sets the stage to investigate, step-by-step, the kinds of genomic and transcriptomic changes that have made such a functionally challenging transition possible for the land crabs. For example, when land crabs from different terrestrial grades are exposed to a similar selective stressor, are similarities in the mechanisms underlying acclimation responses more likely to be observed between close relatives, or between crabs from similar terrestrial grades? While the null hypothesis would propose greater similarity between close relatives due to shared phylogenetic history (*i.e.*, evolutionary conservatism and phylogenetic constraint), many studies have found

that distantly related species often evolve similar adaptive solutions to common selective challenges. Examples of adaptive phenotypic convergence have been found across several levels of biological organization across the Tree of Life, such as for camera lenses in the eyes of cephalopods and vertebrates (Ogura et al. 2004), coat pigmentation in beach mice (Steiner et al. 2009), gill raker numbers in marine-to-freshwater sticklebacks (Glazer et al. 2014), transcriptomically similar symbiotic bioluminescent organs in squid (Pankey et al. 2014), and cardiac glycoside insensitivity in milkweed-consuming insect specialists (Groen and Whiteman, 2021).

In a similar fashion to the above studies, we sought to characterize the transcriptomic changes associated with the transition from marine-to-terrestrial habitats in gecarcinid crabs that exhibit varying grades of terrestrial adaptation. We aimed to (1) compare gene expression responses in three land crab species from different terrestrial grades through time and across two tissues that have been found to be critical in how the crabs tolerate desiccation stress and the resulting downstream osmoregulatory stress that results from hydration limitation, and (2) ascertain the evolutionary history and putative functions of differentially expressed genes (DEGs). Are certain genes repeatedly associated with more terrestrial lifestyles, and if so, what are their evolutionary histories? Are they pre-existing genes that have been co-opted to play novel roles in terrestrial habitats, or are they novel genes that are not shared between species?

To address these questions, we measured tissue-specific differential gene expression in two congeneric land crab sister species from different terrestrial grades, *Tuerkayana celeste* (a freshwater-dependent, air breathing terrestrial crab; Grade III), and *T. magna* (an air-breathing terrestrial crab that digs dry burrows in coastal forests; Grade IV), and a confamilial species, *Gecarcoidea natalis* (an air-breathing terrestrial crab that digs dry burrows in coastal forests that

perishes after extended periods of partial immersion in water; Grade V) after exposing the crabs to a time series of increasing desiccation stress (Table 2.1; Fig. 2.1A). These three species are otherwise physiologically similar, but *T. celeste* requires frequent gill immersion to release its nitrogenous wastes (Dela-Cruz et al. 1997), whereas *T. magna* and *G. natalis* can release this waste via their gills directly into the air (Weihrauch et al. 2004). *T. celeste* also reprocesses its urine to extract vital salts via its antennal gland, which is an ancestral trait, while the posterior gills mediate urine reprocessing in *T. magna* and *G. natalis* (Fig. 2.1A).

Congeneric phylogenetic contrasts between land crabs with differing degrees of terrestriality (*i.e*., between *T. celeste* and *T. magna*) might shed light on how much of their response may be constrained by shared ancestry, and how much their response can vary depending on their terrestrial niche. On the other hand, confamilial phylogenetic contrasts between more distantly related land crabs with similar degrees of terrestriality (*i.e*., between *T. magna* and *G. natalis*) may inform our understanding of how similarities in terrestrial niches may or may not drive similarities in gene expression responses between species. By then ascertaining the putative functions of genes we find to be critical in their respective gene expression responses, we can shed light on how land crabs from different terrestrial grades adapt to the unique selective challenges posed by a transition to a terrestrial life.

2. Materials and Methods

2.1 Sample collection and experimental methods

We collected 21 intermolt adult males of *T. celeste* from the stream leading to Anderson's Dale in Christmas Island National Park, Australia (10°28.7127' S, 105°33.5138'E) in August

2017, and 21 intermolt adult male *G. natalis* from the dry coastal forest near The Blowholes in Christmas Island National Park (10°51.4458'S, 105°62.7927'E) in September 2017. In December 2017, we obtained 21 intermolt adult males of *T. magna* from the pet trade in Singapore (originally collected in Java, Indonesia). We used 20 individuals from each species for individual expression analyses and one crab per species for the *de novo* transcriptome assembly. We collected all Christmas Island specimens by hand and transported them to the laboratory in individual 10L plastic buckets; we also stabilized all *T. magna* specimens in 10L buckets after their arrival from the pet traders (Fig. 2.1B). We kept all animals at local ambient air temperature (which ranged from 28°C - 41°C) and humidity (which ranged from 60% - 82%) throughout the duration of the acclimation and desiccation treatments. After selecting only crabs of similar physical condition (*i.e.* no missing limbs or obvious external physical deformities), we measured each crab's carapace width (CW) and carapace length (CL) immediately after collection, and weighed each crab after blotting off excess water before acclimation and immediately before sampling to record changes in wet weight throughout the experimental period. On average, *T. celeste* males measured CW: 77.07mm (+/- 7.46mm), CL: 63.55mm (+/- 6.34mm), with initial mean weights of 212g (+/- 57.45g) and ending mean weights of 209g (+/- 56.58g). *G. natalis* males measured CW: 77.36mm (+/- 5.11mm), CL: 57.72mm (+/- 4.21mm), and had initial mean weights of 163.14g (+/- 28.33g) and ending mean weights of 154g (+/- 28.97g). *T. magna* males measured CW: 60.03mm (+/- 4.96mm), CL: 50.13mm (+/- 3.49mm), and had initial mean weights of 104g (+/- 17.93g) and ending mean weights of 102.5g (+/- 16.70g). Since there was very little variation among individual crabs in CW and CL, we assumed that all sampled crabs within each species were approximately at the same developmental stage and could be considered biological replicates. While all crabs were intermolt, variation in wet weights for

similar carapace dimensions could reflect different degrees of intermolt biomass accumulation or hydration status in the wild.

We acclimated all crabs for five days in buckets filled with 1.5L Christmas Island freshwater (CIFW), collected daily from a small stream at Ross Hill Gardens, Christmas Island, Australia (10°29.0923'S, 105°40.8031'E). For laboratory studies in Singapore, where natural CIFW was not available, we followed the methods of Wood et al. (1986) for imitating freshwater found naturally at the base of crab burrows on Christmas Island. We refreshed the water in each bucket every 12h. After the 5d acclimation period, we reduced the volume of CIFW in each bucket to 50mL, which was sufficiently high for drinking but not for gill submersion, and maintained the crabs in this desiccation condition for 7 additional days. At the conclusion of the desiccation stress trial, we returned the crabs to acclimation conditions (*i.e.*, 1.5L CIFW in each bucket) to measure their recovery gene expression responses.

Throughout the experiment, we sacrificed four crabs per species at each of five time-points (*i.e.*, immediately after the acclimation period ended (0h; "Baseline"), 24h ("Acute"), 96h, and 168h after the start of the desiccation stress trial ("Extreme"), and finally 24h after returning all crabs to acclimation conditions ("Recovery")), and dissected each individually on ice. Due to unanticipated mortality during the acclimation period for *G. natalis* and restricted catch limits for this protected species within Christmas Island National Park, we did not sample *G. natalis* for the recovery timepoint (Fig. 2.1C). From the crabs sampled for individual gene expression analyses, we took tissue samples from all seven gills from the right side of each crab and its antennal gland; from the one crab per species that we sampled for *de novo* transcriptome preparation, we sampled all seven right-sided gills, the antennal gland, pereiopod muscle tissue, the branchiostegal lung, and the hepatopancreas. We stored all collected samples in 0.5mL

cryovials containing Trizol reagent (Invitrogen) and immediately froze them in LN_2. We transported all samples back to the United States on dry ice held in a 20L dewar storage canister.

2.2 Sequencing and transcriptome assembly

We homogenized each tissue sample for one minute in 100µL Trizol reagent (Invitrogen) using a Qiagen Tissue Lyser and beads at 50Hz, and then added an additional 500µL of Trizol to each tube before centrifuging at 16,000G for 30 seconds. We removed the supernatant and placed it into a clean centrifuge tube, and used the Zymo Direct-zol RNA Miniprep extraction kits to extract and purify whole RNA from each sample, following the manufacturer's instructions. We measured sample concentration and purity using the Agilent 2100 Bioanalyzer (Agilent RNA Eukaryote Total RNA Nano Kit), and the Invitrogen Qubit RNA Broad Range Assay Kit. We then sent all samples to the University of California, Davis Genome and Biomedical Sciences Facility for library preparation and Illumina HiSeq 4000 sequencing. Samples destined for transcriptome assembly were paired-end sequenced at the length of 150bp, while samples slated for differential expression analyses were prepared using a RNA 3' Tag-Seq protocol and sequenced as single-end 90bp reads.

We checked and visualized read quality using the default parameters in FastQC (Andrews 2010) and eliminated low quality reads with Trimmomatic version 0.36 (Bolger et al. 2014) using parameters for paired reads. We removed leading and trailing low-quality base pairs (LEADING:2, TRAILING:2) using a sliding window of 4 base pairs, removing base pairs when quality dropped below 2 (SLIDINGWINDOW:4:2), and kept reads of a minimum length of 25bp (MINLEN:25). In the absence of reference genomes for all three species, we used Trinity (Grabherr et al. 2011) to perform *de novo* transcriptome assemblies for each species using

default parameters, and matched all transcripts to NCBI's BLAST database of protein and translated DNA sequences using DIAMOND (Buchfink et al. 2021) for functional annotation of candidate genes. We matched the top hit per transcript based on e-value and overall percent sequence identity. We then assessed the quality of each assembly with BUSCO's Arthropoda database (Manni et al. 2021).

2.3 Differential gene expression

We estimated transcript abundance using Salmon (Patro et al. 2017), and used DESeq2 (Love et al. 2014) to identify differentially expressed transcripts across three pairwise time-point intervals: (1) Baseline - Acute, (2) Acute - Extreme, and (3) Extreme - Recovery for all possible species and tissue combinations (a total of six comparisons each for *T. celeste* and *T. magna*, and four comparisons for *G. natalis*). Each time-point consisted of four biological replicates. After filtering out genes with low expression (raw read count ≤ 10), we considered any genes with absolute log2FoldChange expression > 1 and an adjusted p-value < 0.05 as significantly differentially expressed. For significantly differentially expressed genes, we then ran one-way ANOVA to determine if there were differences in terms of log2FoldChange expression values across the three pairwise time intervals for each tissue type for all species (total number of ANOVA = 6). When ANOVA results were significant, we used Tukey's HSD Test for multiple comparisons to determine whether there were statistically significant differences between time intervals for each tissue type. To determine the putative functions of significantly differentially expressed genes, we used a custom Python script to match all DEGs from each species to their respective DIAMOND databases, and then selected the top 10 genes with the greatest magnitude of log2FoldChange for each time interval for closer examination.

2.4 Co-expression analysis

To examine shared and divergent evolutionary patterns of gene expression, we conducted a weighted gene correlation network analysis (WGCNA; Langfelder and Horvath 2008). First, we used TransDecoder (Hass et al. 2013) to translate transcripts into protein sequences and identify putative coding regions, and then used OrthoFinder (Emms and Kelly 2019) to sort the filtered transcripts into orthogroups based on orthologous regions. After matching single-copy (1:1:1) orthologs shared by all three species with their respective normalized transcript abundance data from Salmon, we then used WGCNA to identify genes that showed correlated expression patterns across all three species, all four time-points, and both tissue types. We assessed the scale-free topology of our data to choose an appropriate soft-thresholding power (soft thresholding power = 8) by which to raise pairwise correlations of expression, and then transformed ortholog co-expression similarity into a signed adjacency matrix based on the soft thresholding power, which clustered highly correlated genes into modules (minimum module size = 30). We then merged modules with eigengene expression values that were highly correlated (dissimilarity threshold = 0.25), and separated eigengene expression values by tissue type for downstream analysis.

Using a custom Python script, we then matched each ortholog with its respective Trinity transcript identifiers, and used each sequence's corresponding DIAMOND output to link each ortholog with an NCBI accession ID based on percent sequence similarity and e-value. From each module, we then selected the top ten genes with the highest absolute module membership (|MM|) to assign putative functions to each module.

2.5 Evolutionary history of differentially expressed genes

To determine if differentially expressed genes (DEGs) were more likely to be assigned to shared orthogroups, and therefore conserved across lineages, we used a custom Python script to match transcript names from DESeq2's list of significantly differentially expressed genes with transcripts assigned to orthogroups by OrthoFinder. Since OrthoFinder also outputs copy number for each orthogroup, we also determined if assigned genes were more likely to be single-copy or part of expanded gene families (*i.e.,* represented by more than one transcript, although we did not differentiate between isoforms for this analysis). We used Fisher's exact test to determine if associations between DEGs and one of three gene types (i.e., "unassigned", "single-copy", and "expanded") were more likely for all species, tissue type, and time-point combinations (total number of Fisher tests = 32), and then ran an ANOVA to determine if there were differences in terms of log2FoldChange expression between gene type categories for all species, tissue type, and time-point combinations (total number of ANOVA = 16). We then used Tukey's HSD Test for multiple comparisons to determine whether there were statistically significant differences between the log2FoldChange expression values of pairwise gene type groupings for each time interval if prior ANOVA results were significant.

3. Results

3.1 *de novo* reference transcriptome assembly

We obtained 89 – 150 million raw reads per transcriptome, which were assembled into about 113 – 221 thousand transcripts per transcriptome (Table 2.2). Mean contig N50 was 1,482 base pairs, and the mean GC content for each transcriptome was 46%. Using the BUSCO Arthropoda

database to assess transcriptome completeness, we found that the *T. magna* transcriptome was most complete (88% complete), followed by the *T. celeste* transcriptome (82% complete), and then by the *G. natalis* transcriptome (79% complete). Since the BUSCO Arthropoda database was compiled using mostly terrestrial arthropods (i.e., insects and arachnids), and our transcriptomes would not have included any developmentally expressed genes or tissue-specific genes from tissues we did not sample, it is perhaps not surprising that a substantial fraction of transcripts were not represented in any of the decapod crustacean transcriptomes compiled in this study.

3.2 Tissue-specific differential gene expression

3.2.1 *Tuerkayana celeste*

T. celeste responds to sudden desiccation stress primarily by dramatically upregulating genes in its posterior gills, and also in its antennal gland to a much smaller degree. The majority of DEGs were expressed during the Baseline-Acute time interval in both the posterior gill (2,039 transcripts) and the antennal gland (143 transcripts), and most of these genes were upregulated (Fig. 2.2A). The dramatically upregulated gene expression response displayed by the posterior gill during the Baseline - Acute time interval, where we recorded several orders of magnitude more DEGs than at any other time interval for this species, was especially notable (Fig 2.3A; Fig 2.3B).

In the posterior gill, there were statistically significant differences in log2FoldChange expression across all three time intervals, indicating that throughout the time series trial there were large

shifts in overall gene expression patterns in this tissue in response to acute and prolonged desiccation stress and eventual re-immersion (Fig. 2.2B).

3.2.2 *Tuerkayana magna*

T. magna primarily responds to acute and prolonged desiccation stress by modifying gene expression in its antennal gland through time. The majority of DEGs for *T. magna* occurred in antennal gland tissues during the Baseline - Acute interval, where most of the DEGs were downregulated (132 transcripts total), and in the antennal gland tissues during the Acute - Extreme interval, where most of the DEGs were upregulated (96 transcripts total) (Fig. 2.2A; Fig. 2.2C). The signal in the posterior gill during the Extreme – Recovery time interval was elevated but comparatively moderate, where there was an almost equal number of DEGs upregulated as there was downregulated (82 transcripts total) (Fig 2.2A).

3.2.3 *Gecarcoidea natalis*

As was the case for *T. celeste*, *G. natalis* may also respond to acute desiccation stress by upregulating genes in its posterior gill. Most DEGs for *G. natalis* were expressed in the posterior gill during the Baseline - Acute time interval (1,160 transcripts); overall, the gene expression signal during this interval was one of significant upregulation (Fig 2.2A; Fig 2.3E-F). In comparison, during the subsequent time interval (Acute - Extreme), there was only one gene that was significantly differentially expressed in the posterior gill (Fig 2.2A).

3.3 Orthology assignment and co-expression analysis

TransDecoder identified a total of 131,781 coding sequences across all three transcriptomes (52,781 coding sequences in *T. celeste*, 45,084 coding sequences in *T. magna*, and 33,684 coding sequences in *G. natalis*). OrthoFinder assigned 85,186 of these genes (64.8% of total) into 17,661 orthogroups with a mean orthogroup size of 4.8 genes, and a median orthogroup size of 4 genes (although the largest orthogroup consisted of 91 genes). Fifty-eight orthogroups were species-specific (17 were specific to *T. celeste*, 14 were specific to *T. magna*, and 27 were specific to *G. natalis*, while 9,516 orthogroups were present in all three species (Fig. 2.4). OrthoFinder also recovered 2,185 1:1:1 single-copy orthogroups, all of which were used for weighted gene correlation network analyses.

After merging modules with highly correlated eigengene values (dissimilarity threshold = 0.25), WGCNA clustered 2,185 1:1:1 orthologs into six modules of highly correlated genes; unclustered genes were assigned to Module 0. Eigengene expression is defined as the first principal component of a given module, and represents the overall gene expression profile of that particular module. We therefore used sample eigengene values to split each module's eigengene expression into relative contributions by tissue type (*i.e.*, posterior gill and the antennal gland) to determine how module expression varied between species through time in each tissue type.

Overall, eigengene expression reflected species divergence rather than experimental treatments. Five modules showed statistically significant species associations, while none of the modules were associated with any of the timepoints (Fig 2.5A). Across species, some modules showed a phylogenetic signal (*i.e.*, congeners showed more similarity in magnitude and directionality of their gene expression responses, as seen for Modules 1 and 3), while other modules displayed a "grade effect", where expression profiles were more similar across crab species from higher terrestrial grades (as seen for Modules 5 and 6). Finally, Module 4 was very

strongly positively associated with *T. magna* alone; expression of genes in this module were constitutively lower in the remaining two species (Fig. 2.5B). This *T. magna* signal was also seen to a lesser degree in Module 2, but these patterns were not statistically significant.

3.4 Evolutionary context of DEGs

Differentially expressed genes within species tended to be conserved across all three species (i.e. unassigned to any orthogroup), and also tended to be single-copy genes (Supplemental Table 2.1), suggesting that gecarcinid land crabs respond to desiccation stress primarily by using single-copy conserved genes. In nearly every case, except for the Baseline - Acute time interval in the posterior gill of *T. celeste* ($F(2,1) = [83.5962]$, $p = 1.34 \times 10^{-35}$) (Fig 2.6), there were no statistically significant differences between gene types in terms of log2FoldChange expression. For the Baseline - Acute time interval in posterior gill of *T. celeste*, Tukey's HSD Test for multiple comparisons found that the mean value of log2FoldChange expression significantly differed between unassigned genes and single-copy genes ($p < 0.001$, 95% C.I. = [2.3888, 3.7095]), between unassigned genes and genes from expanded gene families ($p < 0.001$, 95% C.I. = [1.474, 2.7216]), and between single-copy genes and genes from expanded gene families ($p = 0.029$, 95% C.I. = [-1.8236, -0.0791]) (Fig 2.6). This result suggests that while some component of the desiccation stress response in *T. celeste* can be traced back to single-copy conserved genes, a significant portion is driven by both genes unique to this species and genes that have undergone gene family expansion. Furthermore, the directionality of expression shifts across these categories varied: unassigned genes at this timepoint were overwhelmingly upregulated, while single-copy genes tended to be downregulated. Genes from expanded gene

families appeared to be modestly bimodal – some were upregulated, while others were downregulated (Fig 2.6).

3.5 Candidate gene functionalization by species

For *T. celeste*, the majority of the top 10 genes with the greatest magnitude of log2FoldChange expression between time intervals for each species-by-tissue-by-time interval combination had no corresponding matches in the NCBI BLAST database. Of the genes that successfully matched publicly available BLAST records, none were single-copy conserved genes. Furthermore, none of the top 10 genes from the gill tissue matched functionally informative records. In the antennal gland during the Baseline-Acute time interval, protein kinase C, which is responsible for signal transduction pathways in the cell, was upregulated. During the Acute-Extreme time interval, there was significant downregulation of genes related to ATP synthesis (the F_1F_0-ATP synthase beta subunit) and upregulation of genes related to megakaryocyte differentiation and protein phosphorylation. Gene copy number for the F_1F_0-ATP synthase beta subunit also varied across species: OrthoFinder revealed 11 copies of this gene in *T. celeste*, only one in *T. magna*, and none in *G. natalis* (which could reflect sequencing error). Finally, during the Extreme – Recovery time interval, there was significant downregulation in genes related to megakaryocyte differentiation and protein phosphorylation (Supplemental Materials, Table 2.2).

For *T. magna*, many more of the top 10 log2FoldChange genes matched to BLAST entries than for *T. celeste*. As was the case for *T. celeste*, none of these genes were single-copy conserved genes. Overall, mitochondrial proteins appeared to be significantly involved in *T. magna*'s desiccation response through time. There were also significant expression changes in

genes involved in vesicular transport, ATP generation, DNA repair, and in those responsible for protecting the mitochondria from oxidative damage (Supplemental Materials, Table 2.3).

For *G. natalis*, more of the top 10 log2FoldChange genes had corresponding BLAST hits than we observed for *T. celeste*, but there were fewer matches than in *T. magna*. As was the case for both of the other species, none of these genes were single-copy conserved genes. Overall, genes encoding mitochondrial enzymes throughout *G. natalis*' desiccative response, including NADH dehydrogenase (subunits 4 and 5) and Cytochrome C oxidase, subunit 1, were significantly upregulated in the posterior gill during the Baseline-Acute time interval. During the Acute-Extreme time interval, the single DEG expressed in the posterior gill matched to a signal recognition particle (SRP) protein, considered responsible for localizing proteins to their proper cellular destinations. Finally, in the antennal gland, genes related to alternative splicing and cyanide detoxification were significantly upregulated (Supplemental Materials, Table 2.4).

3.6 Putative functions of single-copy conserved genes

After selecting the top ten 1:1:1 orthologs with the highest absolute module membership for each module and matching these genes to their respective BLAST hits, we found that nearly all of the selected genes encoded specific transcription factors and chaperone proteins that assist in protein folding. There were also several genes that matched to enzymes responsible for DNA repair (Supplemental Materials, Table 2.5).

4. Discussion

Terrestrial crabs have independently colonized land at least ten times in the past 66 million years. They display many diverse terrestrial ecotypes and inhabit a variety of terrestrial

landscapes while also showing a high degree of functional and ecological convergence in response to the extraordinary adaptive challenges that accompany a terrestrial life. Watson-Zink (2021) developed a novel framework for categorizing degrees of terrestrial adaptation in terms of habitat-by-trait associations after considering deep phylogenetic and ecological differences between anomuran and brachyuran crabs, and also between crabs that colonized land via marine or freshwater transition pathways. This study uses this foundational framework to explore the genomic underpinnings of several physiological terrestrial traits in land crabs, and our findings may address fundamental evolutionary questions about convergent evolution and the nature of sea-to-land life-history transitions across the Tree of Life.

We used tissue-specific RNAseq to assess how three gecarcinid land crabs from different terrestrial grades (*Tuerkayana celeste*, Grade III; *T. magna*, Grade IV; and *Gecarcoidea natalis*, Grade V) (Table 2.1) responded to their most pervasive terrestrial stressor: desiccation. We identified which genes were differentially expressed after exposure to desiccation stress across each species and between two tissue types (the posterior gill and the antennal gland), and then we inferred the evolutionary histories (*i.e.*, species-specific genes, single-copy conserved genes, and genes that have undergone gene expansion) and putative functions of these genes.

In terms of overall tissue-specific patterns of gene expression, our DESeq2 differential expression analyses showed that *T. celeste* and *G. natalis* responded very strongly in their posterior gills, whereas *T. magna* responded to a lesser degree in its antennal gland. Since posterior gills in land crabs play more of an osmoregulatory than a respiratory role (Péqueux 1995), this finding suggests that both *T. celeste* and *G. natalis* were under significant osmoregulatory stress 24h after the start of the desiccation stress trial. In fact, the greatest gene expression response for both *T. celeste* and *G. natalis*, both in terms of quantities of DEGs and

magnitude of expression change, occurred in the posterior gill during the Baseline - Acute time interval. In contrast, *T. magna* displayed significantly fewer DEGs than the other two species during the same time interval in its gills (Fig. 2.3C and Fig 2.3D).

In interpreting these patterns, it may be important to consider that these two species may have been in fact responding to different kinds of osmoregulatory pressure. *T. celeste* relies heavily on freshwater to release its nitrogenous waste, and cannot release this waste unless its gills are submerged (Dela-Cruz and Morris 1997), so the observed changes in posterior gill gene expression may truly represent a response to the acute desiccation stress experienced at the start of the time trial for this species. *G. natalis*, however, is highly terrestrial, inhabiting dry burrows that do not contact the water table, and it uses thick setal tufts on the coxae of its posterior walking legs to passively wick moisture from dew drops and puddles (Watson-Zink 2021). Since these tufts were constantly submerged during the 5d acclimation period and four otherwise healthy *G. natalis* crabs unexpectedly perished during acclimation, the acclimation conditions themselves may actually have been stressful for this species, to the point that the crabs succumbed to overhydration and extreme hypo-osmotic stress. In this instance, the gene expression signal from its gills 24h after the start of the desiccation trial may reflect extreme stress response during the so-called acclimation period.

Due to its obligatory reliance on submersion in freshwater for urination (Weihrauch et al. 2004; Dela-Cruz and Morris 1997), we also predicted a dramatic response in *T. celeste*'s antennal gland during the Extreme - Recovery time interval. Compared to its response during the Baseline - Acute time interval, however, its response during the Extreme - Recovery time interval was relatively muted (2,039 vs 35 DEGs, respectively). Dela-Cruz and Morris (1997) found that *T. celeste* rapidly releases its nitrogenous wastes within 20 minutes to an hour after

reimmersion, leaving open the possibility that the sampling 24h post re-immersion failed to capture the majority of this recovery response.

The genes with the greatest change in gene expression in response to acute desiccation stress encoded mitochondrial enzymes responsible for ATP generation in the cell and enzymes involved in vesicular transport using endosomes (Supplemental Tables 2.2 – 2.4), suggesting that land crabs employ energy-intensive processes to combat desiccation. The extreme degree of apparent upregulation (e.g., a log2FoldChange of 27.3 corresponds to a 165 million-fold change) across the sampling intervals for many of these differentially expressed genes could be related to gene copy number. In particular, in many of the instances where gene expression differences exceeded 20-fold, the respective gene was part of an expanded gene family (Supplemental Tables 2.2 – 2.4). A deeper analysis of gene family expansion is warranted to more deeply explore these trends in land crabs.

After grouping differentially expressed genes into modules of similar expression, differences between species best explained the prevailing signal. This may simply reflect the estimated 700,000 years of independent evolution between *T. celeste* and *T. magna* (Ng and Shih 2014), and presumably the even greater phylogenetic distance between the *Tuerkayana* crabs and their confamilial, *G. natalis*. We found that, for some of our examined modules (Modules 1 and 3), their respective responses reflected shared ancestry (*i.e.*, we found more similarities in gene expression responses between *T. celeste* and *T. magna*, which are sister species), implying that these particular modules may be under greater degrees of phylogenetic constraint. In other cases (for Modules 5 and 6), land crabs from similar terrestrial grades (*i.e.*, *T. magna* and *G. natalis*, which are confamilial species) responded more similarly, suggesting that genes in these modules may either be related to genes that allow these crabs to exist at higher terrestrial grades (as may

be the case for Module 6), or alternatively, related to genes that allow *T. celeste* to survive in its unique terrestrial niche (*i.e*., restricted to the shores of freshwater streams for urinary purposes (Dela-Cruz and Morris, 1997)), as may be the case for Module 5. Interestingly, for Module 5, *T. celeste* showed a U-shaped response curve in both its antennal gland and its posterior gill, reflecting the overall degree of protein translation for genes in this module through time. Expression levels dropped significantly after the start of the desiccation trial for this species and began to increase again post-recovery, suggesting perhaps that desiccation stress causes *T. celeste* to shut down protein translation for genes in this module until its hydration status improves. In the posterior gill in particular, during the Acute – Extreme time interval, its modular gene expression closely resembled that seen for *T. magna* and *G. natalis*, suggesting adaptive plasticity in response to hydration status in this species.

The single-copy conserved genes in Modules 5 and 6 primarily encode transcription factors for cellular protein complexes and chaperones. Species-level differences in gene expression for transcription factors and chaperones potentially reflect constitutive differences in protein translation for specific protein complexes between species. While an in-depth analysis of these key complexes and the involved biochemical pathways is outside the scope of this study, future work should further investigate how these particular pathways may be critical to allowing land crabs to mitigate the osmoregulatory challenges associated with their terrestrial lifestyles.

Finally, while DEGs tended to be single-copy conserved genes, the DEGs with the greatest log2FoldChange expression for all three species were exclusively genes that were part of expanded gene families or novel genes that were not assigned to any orthogroup. In fact, for *T. celeste*, 85% of its gill-derived transcriptomic response to acute desiccation stress was driven by these novel, species-specific genes. Furthermore, in *T. celeste*, novel genes were overwhelmingly

upregulated in response to acute desiccation stress, whereas single-copy, conserved genes were significantly downregulated, which further supports the hypothesis that this species primarily uses novel genes to combat desiccation stress while on land. Nevertheless, the putative functions of these novel genes and their corresponding proteins remain to be identified.

LITERATURE CITED

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Anger, K., Queiroga, H., and Calado, R., (2015). Larval development and behavior strategies in Brachyura. *The Crustacea,* Vol. 9C, Pages 317 – 374.

Bliss, D.E., (1968). Transition from Water to Land in Decapod Crustaceans. *American Zoologist*, Volume 8, Issue 3, Pages 355–392.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

Buchfink, B., Reuter, K., & Drost, H.G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND, *Nature Methods* 18, 366–368.

Dela-Cruz, J., & Morris, S. (1997). Water and ion balance and nitrogenous excretion as limitations to terrestrial excursion in the Christmas Island blue crab, *Cardisoma hirtipes* (Dana). *Journal of Experimental Zoology*, 279(6), 537–548.

Dunham, D.W. and Gilchrist, S.L. (1988) Behavior. In Burggren, W.W. and McMahon, B.R. (1988) *Biology of the Land Crabs.* Pg. 97-137. Cambridge University Press.

Emms, D.M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20, 238.

Glazer, A.M., Cleves, P.A., & Erickson, P.A. Parallel developmental genetic features underlie stickleback gill raker evolution. *EvoDevo* 5, 19 (2014). https://doi.org/10.1186/2041-9139-5-19

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology, 29 (7), 644–652.

Greenaway, P., & Linton, S.M., (1995). Dietary assimilation and food retention time in the herbivorous terrestrial crab, *Gecarcoidea natalis*. *Physiological Zoology* 68 (6): 1006-1028.

Groen, S.C. & Whiteman, N.K. (2021) Convergent evolution of cardiac-glycoside resistance in predators and parasites of milkweed herbivores. *Current Biology*; 31(22):R1465-R1466.

Grosberg, R.K., Vermeij, G.J., & Wainwright, P.C., (2012). Biodiversity in water and on land. *Current Biology* 22 (21): R900-R903.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, & Regev A. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*; 8(8):1494-512.

Herreid, C.F. & Full, R.J (1988) Energetics and Locomotion. In Burggren, W.W. and McMahon, B.R. (1988) *Biology of the Land Crabs.* Pg. 333-376. Cambridge University Press.

Krieger, J., Braun, P., Rivera, N.T., Schubart, C.D., Müller, C.H., & Harzsch, S., (2015). Comparative analyses of olfactory systems in terrestrial crabs (Brachyura): evidence for aerial olfaction? *PeerJ*. 3:e1433.

Langfelder, P., & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559 (2008).

Love MI, Huber W, Anders S (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology*, 15, 550.

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., & Zdobnov, E.M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, Volume 38, Issue 10, Pages 4647–4654.

Morris, S., (2002). The ecophysiology of air-breathing in crabs with special reference to *Gecarcoidea natalis. Comparative Biochemistry and Physiology,* B 131:559-570.

Ng, P.K.L. & Shih, H.T. (2014) The systematics of the land crabs of the Discoplax hirtipes (Dana, 1851) species-group (Crustacea: Decapoda: Brachyura: Gecarcinidae), with description of a new species from the eastern Indian Ocean. *Raffles Bulletin of Zoology* Supplement No. 30: 109–135.

Ogura, A., Ikeo, K., & Gojobori, T. (2004) Comparative Analysis of Gene Expression for Convergent Evolution of Camera Eye Between Octopus and Human. *Genome Research*, 14: 1555-1561.

Pankey, M. S., Minin, V. N., Imholte, G. C., Suchard, M. A., & Oakley, T. H. (2014). Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. *Proceedings of the National Academy of Sciences,* 111(44), E4736–E4742.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14(4), 417–419.

Péqueux, A. (1995). Osmotic regulation in crustaceans. *Journal of Crustacean Biology*, Vol. 15, No. 1, Pages 1-60.

Steiner, C.C., Römpler, H., Boettger, L.M., Schöneberg, T., & Hoekstra, H.E., (2009) The Genetic Basis of Phenotypic Convergence in Beach Mice: Similar Pigment Patterns but Different Genes, *Molecular Biology and Evolution*, Volume 26, Issue 1, Pages 35–45.

Stensmyr M.C., Erland, S., Hallberg, E., Wallen, R., Greenaway, P. & Hansson, B.S., (2005). Insect-like olfactory adaptations in the terrestrial giant robber crab. *Current Biology*, Vol. 15, 116–121.

Taylor, J. R.A., (2018). Aquatic versus terrestrial crab skeletal support: morphology, mechanics, molting and scaling. *Journal of Experimental Biology,* 221, jeb185421.

Taylor, J.R.A., & Kier, W.M., (2006). A pneumo-hydrostatic skeleton in land crabs. *Nature*, Vol. 440, Page 1005.

Vermeij, G.J. & Dudley, R., (2000). Why are there so few evolutionary transitions between aquatic and terrestrial ecosystems? *Biological Journal of the Linnean Society,* 70: 541 -554.

Vermeij, G.J., & Watson-Zink, V.M. (2022). Terrestrialization in gastropods: lineages, ecological constraints and comparisons with other animals, *Biological Journal of the Linnean Society*, blac053.

Watson-Zink, V.M. (2021). Making the grade: Physiological adaptations to terrestrial environments in decapod crabs. *Arthropod Structure and Development* 64: 101089.

Weihrauch, D., Morris, S, & Towle, D.W., (2004). Ammonia excretion in aquatic and terrestrial crabs. *Journal of Experimental Biology,* 207:4491-4504.

Wood, C.M., Boutilier, R.G., & Randall, D.J. (1986) The physiology of dehydration stress in the land crab, *Cardisoma carnifex*: respiration, ionoregulation, acid-base balance and nitrogenous waste excretion. *Journal of Experimental Biology*, 126:271-296.

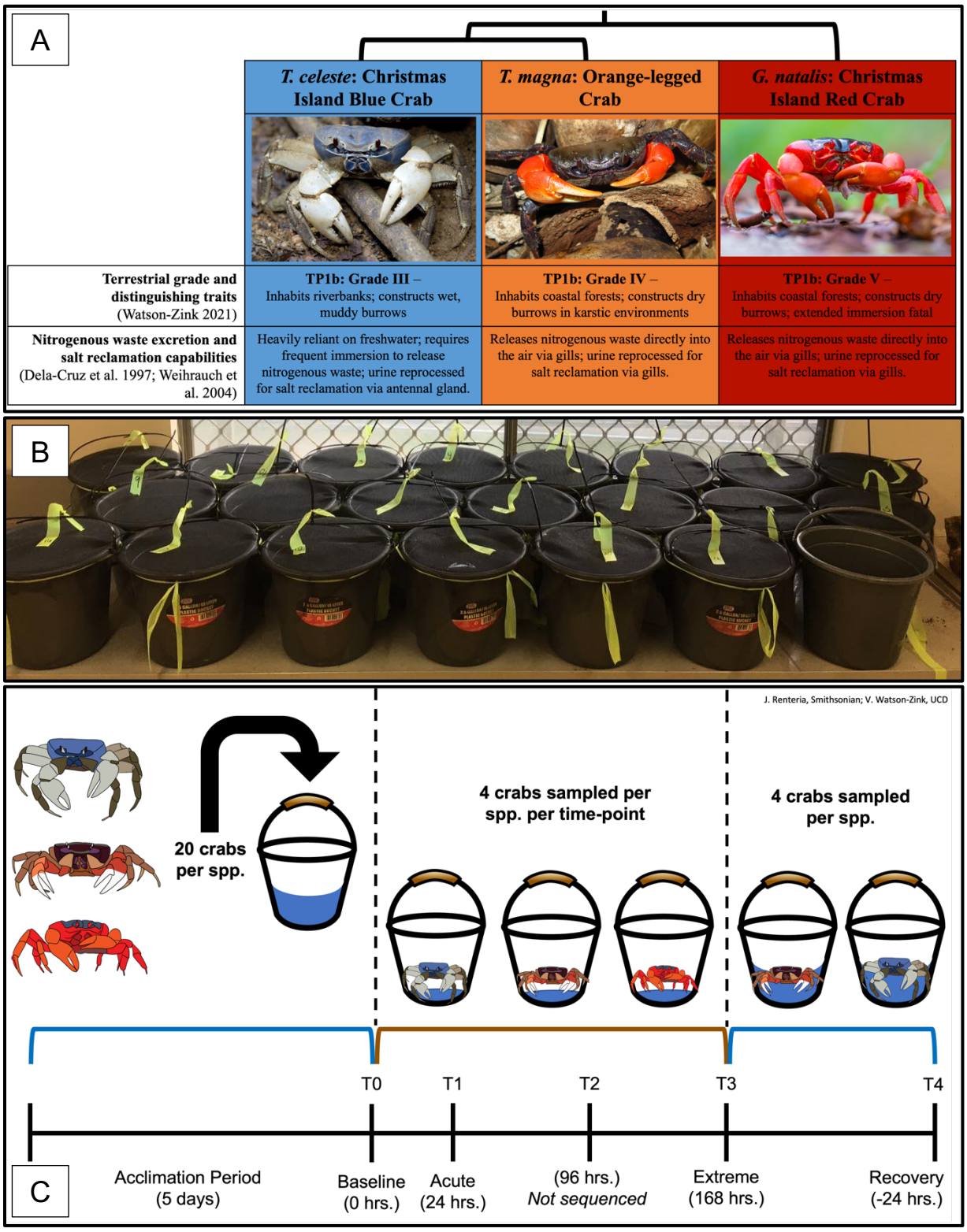


Fig. 2.1: Experimental design and sampling schedule for *T. celeste*, *T. magna*, and *G. natalis*. (A) Cladogram depicting relatedness between three experimental species, their respective

terrestrial grades (Watson-Zink 2021), and their distinguishing physiological traits in relation to nitrogenous waste excretion and salt reclamation; (B) Crabs were held in covered 10L buckets throughout the acclimation, desiccation, and recovery periods; (C) Antennal gland and posterior gill 7 tissues were sampled from each crab at four timepoints to measure gene expression responses in each tissue type through time and between species. (Crab pictograph credits: J. Renteria (Smithsonian) and V. Watson-Zink (UCD)).

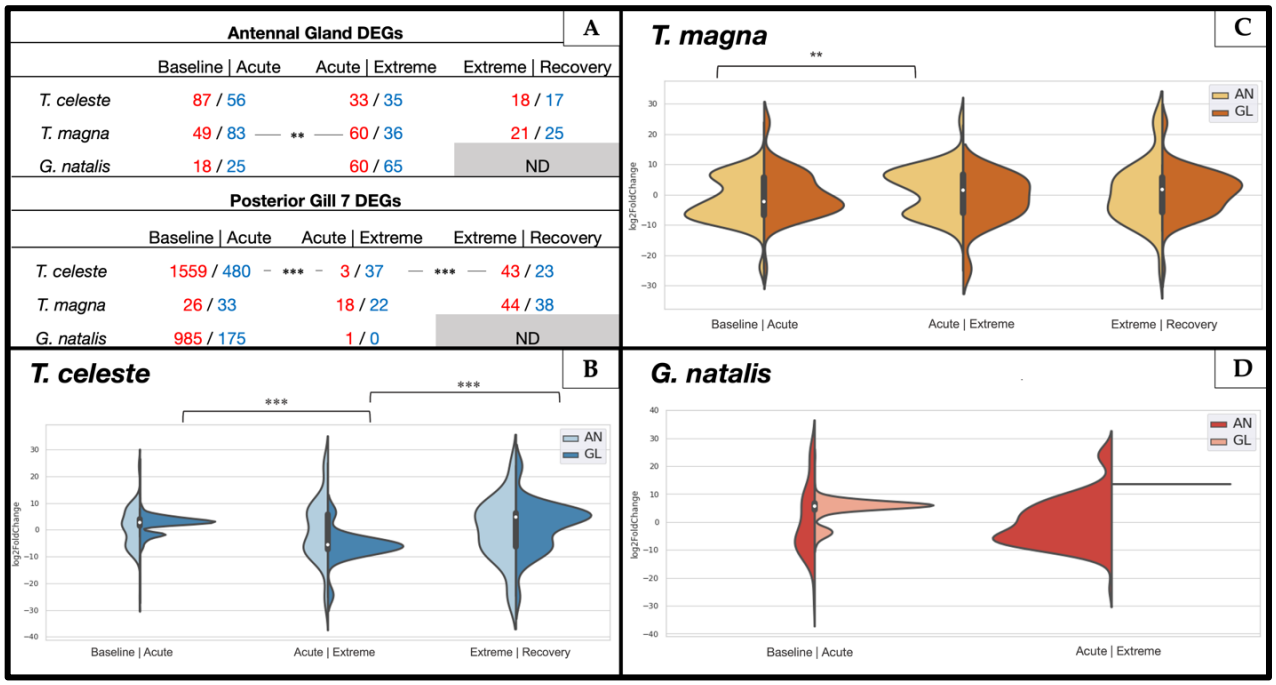


Fig. 2.2: Tissue-specific differentially expressed genes for all three species through time. (A) Table depicting quantities of upregulated (red text) and downregulated (blue text) genes during each time interval for both tissue types and across all three species; (B-D) Violin plots depicting distribution of log2FoldChange expression values and summary statistics during each time interval for both tissue types for *T. celeste* (B), *T. magna* (C), and *G. natalis* (D). Asterisks reflect degree of statistical significance between time interval pairings: $p < 0.05$ (*); $p < 0.01$ (**); $p < 0.001$ (***).

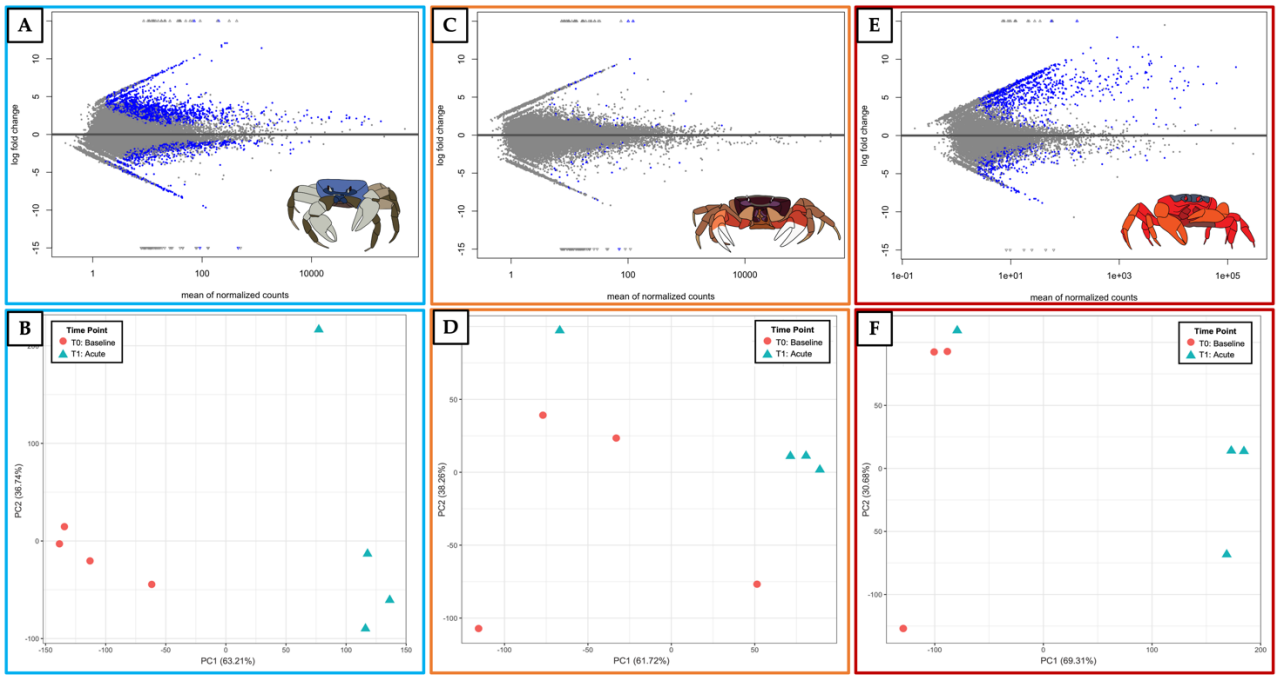


Fig. 2.3: (A, C, E): Volcano plots showing mean of normalized counts by log2FoldChange expression for *T. celeste* (A), *T. magna* (C), and *G. natalis* (E) during the Baseline – Acute time interval. (B, D, F): Principal component analyses between four biological replicates each for *T. celeste* (B), *T. magna* (D), and *G. natalis* (F) during the Baseline – Acute time interval. (Crab pictograph credits: J. Renteria (Smithsonian) and V. Watson-Zink (UCD)).

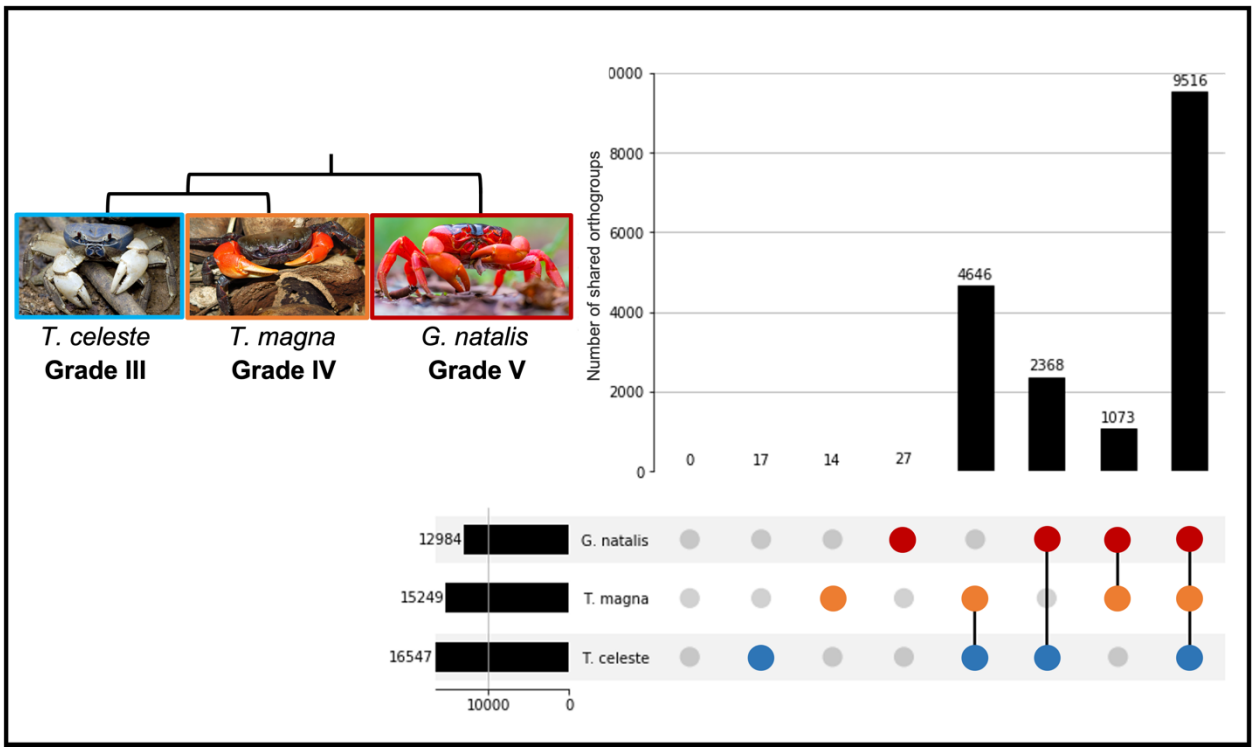


Fig. 2.4: Upset diagram of OrthoFinder (CITATION) results, showing the relative contributions of each species (*T. celeste*, *T. magna*, and *G. natalis*) to the orthogroups, and intersections between species. Each number represents the number of orthogroups unique to that species or grouping of species. A graphical representation of total orthogroups per species is also plotted along the x-axis. Inset: Pictorial cladogram displaying phylogenetic relationships between species and relative terrestrial grades (Watson-Zink 2021).

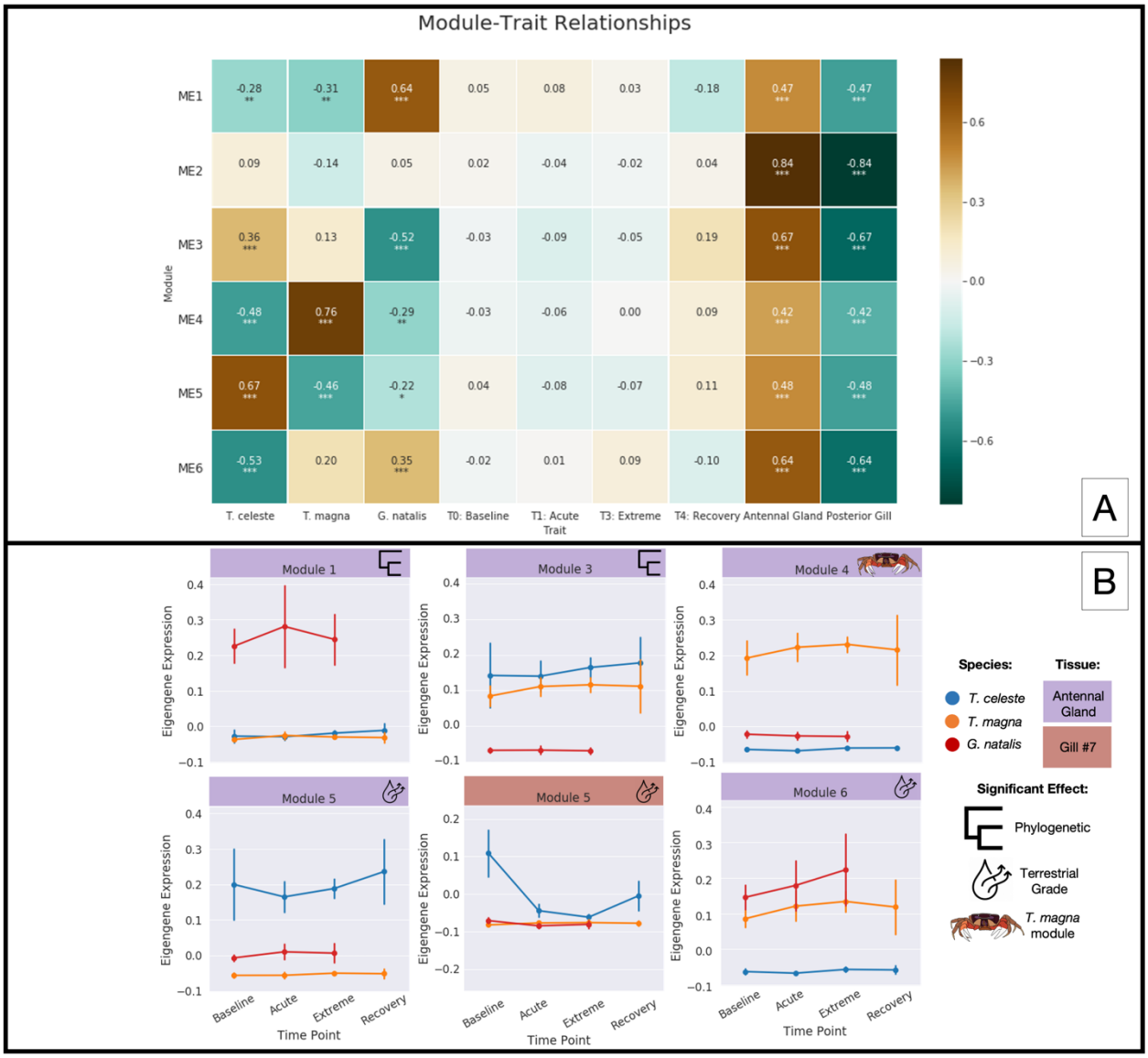


Figure 2.5: (A) Module-trait relationships of co-expression modules designated by WGCNA (Citation). Warmer colors represent more positive correlations between particular modules and traits (*i.e.*, timepoints, species, tissue types), while cooler colors represent more negative correlations between the aforementioned factors. Asterisks reflect degree of statistical significance between relationships: $p < 0.05$ (*); $p < 0.01$ (**); $p < 0.001$ (***). (B) Tissue-specific eigengene expression values for key co-expression modules for all three species. Icons depict significant effects for each module (*i.e.*, phylogenetic effects, grade effects, etc.).

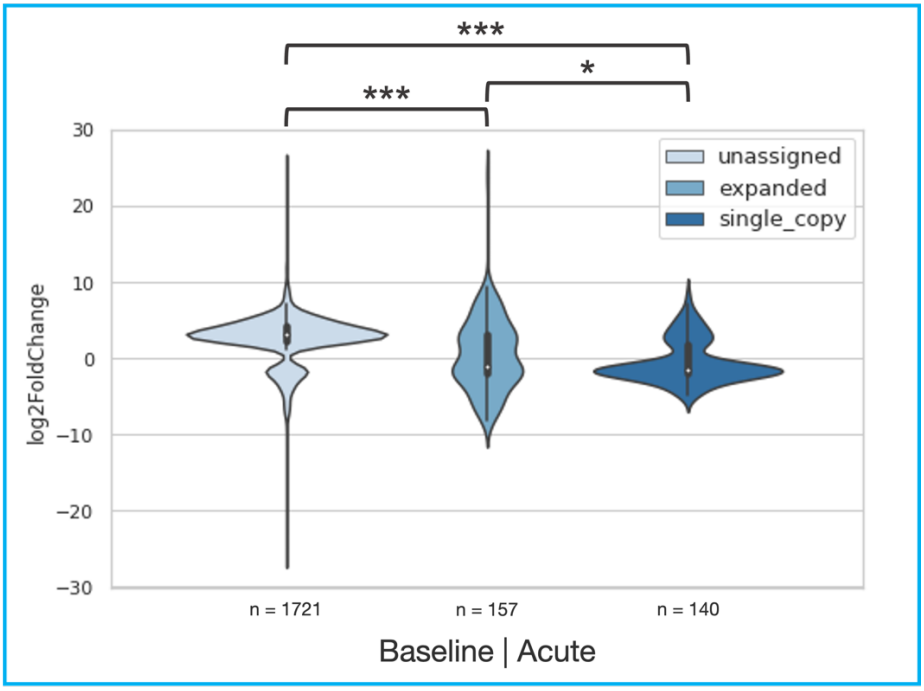


Figure 2.6: Violin plots representing quantities and log2FoldChange expression values and distributions for different gene types (*i.e.*, "unassigned": genes not clustered in any shared orthogroups by WGCNA; "expanded": genes having more than one copy in an orthogroup for a single species; "single-copy": genes that have only one copy in an orthogroup for that species) during the Baseline – Acute time interval in the posterior gill tissue of *T. celeste*. Asterisks reflect degree of statistical significance between log2FoldChange expression values of gene type comparisons: $p < 0.05$ (*); $p < 0.01$ (**); $p < 0.001$ (***).

	Grade I	Grade II	Grade III	Grade IV	Grade V	Grade VI
	Lower intertidal/mudflats/ sandflats/mangrove forests/estuaries	Upper intertidal/ Freshwater	Beaches/ Riverbanks	Coastal forests / Tropical jungles		Arid zone
(TP:1a) Anomuran crabs that colonized land via marine environments	Weak osmoregulation		Alternative respiratory structures; Obligate fossoriality with wet burrows	Accelerated planktonic larval development (some species); Migratory larval export strategy; Nocturnality	Dry burrows; Uricotely	Ecotype not observed
(TP:1b) Brachyuran crabs that colonized land via marine environments	Weak osmoregulation	Facultative strong osmoregulation	Alternative respiratory structures; Obligate fossoriality with wet burrows; Derived ammoniotely; Obligate strong osmoregulation; Water-wicking hairs	Pneumo-hydrostatic skeletal intermediate during ecdysis; Intermolt ion storage; Accelerated planktonic larval development (some species); Migratory larval export strategy; Nocturnality	Fossoriality with dry burrows; posterior coxal setal tufts or pterygostomial setae; Extended full-immersion fatal	Ecotype not observed
(TP:2a) Anomuran crabs that colonized land via freshwater	*Trait analyses not conducted for this ecotype.*					
(TP:2b) Brachyuran crabs that colonized land via freshwater environments	Facultative strong osmoregulation; Amphidromy	Obligate strong osmoregulation; Lecithotrophic larvae with abbreviated development	Alternative respiratory structures; Obligate fossoriality with wet burrows; Derived ammoniotely	Intermolt ion storage; Migratory larval export strategy; Nocturnality	Lecithotrophic larvae with direct development; Some maternal care (some species)	Estivation

Table 2.1: The Six Grades of Terrestriality framework (Watson-Zink 2021) that proposed a hypothetical sequence of terrestrial trait evolution across four Transition Pathways (TP): anomuran crabs that colonized land via marine environments, brachyuran crabs that colonized land via marine environments, anomuran crabs that colonized land via freshwater habitats, and brachyuran crabs that colonized land via freshwater environments. The Terrestrial Grades represent observable trait-by-habitat associations for the four proposed TPs, and can be understood to represent varying degrees of terrestrial adaptation for the land crabs.

T. celeste		***T. magna***		***G. natalis***	
Total transcripts	221,111	Total transcripts	143,688	Total transcripts	113,191
Total bases	150,915,012	Total bases	133,242,897	Total bases	89,956,155
GC%	47%	GC%	46%	GC%	46%
BUSCO (Arthropoda)	82%	BUSCO (Arthropoda)	88%	BUSCO (Arthropoda)	79%
All transcript contigs		**All transcript contigs**		**All transcript contigs**	
N50 transcript length (bp)	1,114	N50 transcript length (bp)	1,769	N50 transcript length (bp)	1,562
Smallest contig length	201	Smallest contig length	201	Smallest contig length	201
Mean contig length	682.53	Mean contig length	927.31	Mean contig length	794.73
Largest contig length	17,698	Largest contig length	19,701	Largest contig length	21,899

Table 2.2: *de novo* transcriptome summary statistics for the four species represented in this study.

Species	Target accession	Sequence identity	e-value	Length	log2FoldChange	Protein name	Description	Gene type	Copy number (TC)	Copy number (TM)	Copy number (GN)
T. celeste	*No BLAST hits for sequences from this timepoint for this species*										
T. magna	XP_042203314.1	93.1	2.50E-103	204	23.71436481	ADP-ribosylation factor-like protein 4C	Regulates the microtubule-dependent intracellular vesicular transport from early endosome to recycling endosome process	EX	1	15	0
T. magna	MPC10256.1	93.5	1.00E-32	77	10.02549835	ATP synthase subunit alpha, mitochondrial	ATP generation	U			
T. magna	KAG0722361.1	76.5	2.20E-44	119	-8.371220164	RNA-directed DNA polymerase from mobile element jockey	internal promoter described for RNA polymerase II	U			
T. magna	CAR82606.1	96.8	5.80E-122	218	8.356303211	mitochondrial manganese superoxide dismutase	key enzyme that protects the energy-generating mitochondria from oxidative damage	EX	3	8	1
G. natalis	EGZ74219.1	97.2	6.20E-09	36	24.30399782	hypothetical protein		U			
G. natalis	YP_009107263.1	77.9	5.50E-27	95	12.86534008	NADH dehydrogenase subunit 5 (mitochondrial)	ATP generation	U			
G. natalis	YP_009535655.1	84.1	8.80E-86	208	11.62756527	Cytochrome C oxidase subunit 1 (mitochondrial)	Electron transport chain	U			
G. natalis	KPV71458.1	90	1.60E-12	40	11.61158546	hypothetical protein		U			
G. natalis	OFQ97201.1	94	2.20E-15	50	11.51054721	hypothetical protein		U			
G. natalis	QRK27379.1	84.9	2.10E-21	73	10.75707609	NADH dehydrogenase subunit 4 (mitochondrial)	ATP generation	U			
G. natalis	ORD93174.1	96.5	1.20E-41	85	10.29791325	hypothetical protein		U			

Table 2.3: Gene names and descriptions for top 10 DEGs in terms of log2FoldChange expression in the posterior gill for the Baseline-Acute time interval for all three species. Only genes with BLAST hits are shown. Gene types (EX: expanded gene family; U: unassigned to any orthogroup) and copy numbers per species (TC: *T. celeste*; TM: *T. magna*; GN: *G. natalis*) are also shown.

	T. celeste									
	Antennal Gland					**Posterior Gill**				
Baseline - Acute		Expanded	Single-copy		Unassigned	Assigned		Expanded	Single-copy		Unassigned	Assigned
	DEG	13	0	DEG	119	13	DEG	155	137	DEG	1720	292
	nonDEG	166	23	nonDEG	15639	189	nonDEG	281	114	nonDEG	15443	395
	p-value		0.369412	p-value		**3.34E-08**	p-value		**3.28E-06**	p-value		**7.55E-101**
Acute - Extreme		Expanded	Single-copy		Unassigned	Assigned		Expanded	Single-copy		Unassigned	Assigned
	DEG	10	2	DEG	49	12	DEG	4	4	DEG	27	8
	nonDEG	217	39	nonDEG	20841	256	nonDEG	61	5	nonDEG	9352	66
	p-value		1	p-value		**4.29E-11**	p-value		**1.14E-02**	p-value		**5.53E-10**
Extreme - Recovery		Expanded	Single-copy		Unassigned	Assigned		Expanded	Single-copy		Unassigned	Assigned
	DEG	7	0	DEG	24	7	DEG	2	0	DEG	61	2
	nonDEG	368	137	nonDEG	22790	505	nonDEG	79	6	nonDEG	10388	85
	p-value		0.197849	p-value		**8.37E-06**	p-value		1	p-value		0.095321971

	T. magna									
	Antennal Gland					**Posterior Gill**				
Baseline - Acute		Expanded	Single-copy		Unassigned	Assigned		Expanded	Single-copy		Unassigned	Assigned
	DEG	23	7	DEG	96	30	DEG	10	3	DEG	45	13
	nonDEG	469	150	nonDEG	21289	619	nonDEG	125	7	nonDEG	10026	132
	p-value		1	p-value		**2.50E-18**	p-value		0.04707683	p-value		**4.02E-12**
Acute - Extreme		Expanded	Single-copy		Unassigned	Assigned		Expanded	Single-copy		Unassigned	Assigned
	DEG	16	2	DEG	64	18	DEG	10	1	DEG	26	11
	nonDEG	271	21	nonDEG	19539	292	nonDEG	181	13	nonDEG	12112	194
	p-value		0.6326168	p-value		**1.92E-15**	p-value		0.54999316	p-value		**4.23E-11**
Extreme - Recovery		Expanded	Single-copy		Unassigned	Assigned		Expanded	Single-copy		Unassigned	Assigned
	DEG	9	0	DEG	28	9	DEG	19	3	DEG	57	22
	nonDEG	308	63	nonDEG	19676	371	nonDEG	95	3	nonDEG	9036	98
	p-value		0.3660359	p-value		**5.42E-08**	p-value		0.0743691	p-value		**5.15E-23**

	G. natalis									
	Antennal Gland					**Posterior Gill**				
Baseline - Acute		Expanded	Single-copy		Unassigned	Assigned		Expanded	Single-copy		Unassigned	Assigned
	DEG	6	2	DEG	33	8	DEG	130	173	DEG	833	303
	nonDEG	354	103	nonDEG	17364	457	nonDEG	100	7	nonDEG	5503	107
	p-value		1	p-value		**1.78E-05**	p-value		**1.80E-21**	p-value		**1.54E-155**
Acute - Extreme		Expanded	Single-copy		Unassigned	Assigned		Expanded	Single-copy		Unassigned	Assigned
	DEG	19	8	DEG	95	27	DEG	1	0	DEG	0	1
	nonDEG	419	130	nonDEG	18566	549	nonDEG	262	58	nonDEG	10475	320
	p-value		0.4905793	p-value		**1.31E-15**	p-value		1	p-value		0.029733235

Supplemental Table 2.1: Fisher tests were performed on both tissue types for all three species at all three time intervals for (1) associations between differential expression status and whether genes were part of expanded gene families or were single-copy, and (2) associations between differential expression status and assignment status (*i.e.*, assigned to orthogroup or unique to that particular species). The p-values that were statistically significant are shown in bold.

Tissue	Time interval	Target accession	Sequence identity	e-value	Length	log2FoldChange	Protein name	Description	Gene type	Copy number (TC)	Copy number (TM)	Copy Number (GN)
G7	T0 - T1	*No BLAST hits from any sequences at this timepoint*										
G7	T1 - T3	KAF7655963.1	38.9	2.40E-13	126	7.306859143	hypothetical protein		U			
G7	T3 - T4	KAG0713759.1	26.1	1.40E-06	176	-24.49554774	hypothetical protein		U			
G7	T3 - T4	KAG0730152.1	64.2	2.50E-16	81	-8.815294651	hypothetical protein		U			
AN	T0 - T1	XP_042224072.1	92.5	7.00E-119	227	22.90468694	protein GDAP2 homolog isoform X1	Response to retinoic acid	U			
AN	T0 - T1	MPC39525.1	89.3	2.70E-116	225	10.98368639	Protein kinase C (PKC)	Signal transduction	EX	5	2	1
AN	T1 - T3	ACU31053.1	97.8	3.60E-67	139	-24.14261223	F_1F_0-ATP synthase beta subunit	ATP synthesis	EX	11	1	0
AN	T1 - T3	KAG0725500.1	89	1.50E-192	356	23.77767112	Fatty acid hydroxylase domain-containing protein 2	Regulation of megakaryocyte differentiation and protein phosphorylation	U			
AN	T3 - T4	KAG0725500.1	89	1.50E-192	356	-24.82250696	Fatty acid hydroxylase domain-containing protein 2	Regulation of megakaryocyte differentiation and protein phosphorylation	U			
AN	T3 - T4	MPC51976.1	90.2	2.60E-19	61	23.6496862	hypothetical protein		EX	5	1	0
AN	T3 - T4	MPC51976.1	90.2	2.60E-19	61	23.30334081	hypothetical protein		U			

Supplemental Table 2.2: Gene names and descriptions for top 10 DEGs in terms of log2FoldChange expression for *T. celeste* for both tissues (G7: Posterior gill; AN: Antennal gland) and all time intervals (T0 – T1: Baseline – Acute; T1 – T3: Acute – Extreme; T3 – T4: Extreme – Recovery). Only genes with BLAST hits are shown. Gene types (EX: expanded gene family; U: unassigned to any orthogroup) and copy numbers per species (TC: *T. celeste*; TM: *T. magna*; GN: *G. natalis*) are also shown.

Tissue	Time interval	Target accession	Sequence identity	e-value	Length	log2FoldChange	Protein name	Description	Gene type	Copy number (TC)	Copy number (TM)	Copy number (GN)
G7	T0 - T1	XP_042203314.1	93.1	2.50E-103	204	23.71436481	ADP-ribosylation factor-like protein 4C	Regulates the microtubule-dependent intracellular vesicular transport from early endosome to recycling endosome process	EX	1	15	0
G7	T0 - T1	MPC10256.1	93.5	1.00E-32	77	10.02549835	ATP synthase subunit alpha, mitochondrial	ATP generation	U			
G7	T0 - T1	KAG0722361.1	76.5	2.20E-44	119	-8.371220164	RNA-directed DNA polymerase from mobile element jockey	internal promoter described for RNA polymerase II	U			
G7	T0 - T1	CAR82606.1	96.8	5.80E-122	218	8.356303211	mitochondrial manganese superoxide dismutase	key enzyme that protects the energy-generating mitochondria from oxidative damage	EX	3	8	1
G7	T1 - T3	XP_042218110.1	97.4	2.20E-111	228	-9.688395924	POU domain, class 6, transcription factor 2-like	involved in early steps in the differentiation of amacrine and ganglion cells	U			
G7	T1 - T3	ROT85526.1	86.5	0.00018	37	8.893788483	putative ribosome-binding protein 1-like isoform X11	mediates interaction between the ribosome and the endoplasmic reticulum membrane	EX	2	14	3
G7	T1 - T3	MPC17943.1	94.8	0	659	8.520698922	Stress-70 protein, mitochondrial	Chaperone protein which plays an important role in mitochondrial iron-sulfur cluster (ISC) biogenesis	EX	1	3	1
G7	T1 - T3	KAG0729092.1	95.1	1.90E-171	385	-8.32228754	Hippocampus abundant transcript 1 protein	transmembrane transporter activity	EX	2	2	2
G7	T1 - T3	MPC29685.1	91	1.10E-179	368	-8.178637824	hypothetical protein		EX	2	5	1
G7	T3 - T4	XP_027214065.1	87.4	2.10E-83	175	9.056749228	remodeling and spacing factor 1-like	Regulatory subunit of the ATP-dependent RSF-1 and RSF-5 ISWI chromatin-remodeling complexes	EX	0	2	1
G7	T3 - T4	KAG0695240.1	93.2	1.90E-173	340	-8.557259444	DNA repair protein RAD51 1	assists in repair of DNA double strand breaks	EX	1	3	1
AN	T0 - T1	MPC27037.1	74.3	0.009	35	-23.80067488	hypothetical protein		EX	13	5	27
AN	T0 - T1	KAG0713905.1	61	8.10E-23	118	-10.05655494	hypothetical protein		U			
AN	T0 - T1	MPC48966.1	92.1	9.70E-61	126	9.772745131	Multiple coagulation factor deficiency protein 2	Plays a role in the secretion of coagulation factors	EX	2	8	1
AN	T1 - T3	KAG0713905.1	61	8.10E-23	118	10.36134167	hypothetical protein		U			
AN	T1 - T3	MPC60099.1	94.3	2.20E-18	53	9.439730982	hypothetical protein		U			
AN	T3 - T4	ADZ46233.1	77.9	1.80E-52	122	23.91750434	anti-lipopolysaccharide factor 3	antimicrobial peptide (AMP) which show broad-spectrum antimicrobial activity against Gram-positive bacteria, Gram-negative bacteria, fungi and viruses.	EX	1	5	0
AN	T3 - T4	MPC60099.1	94.3	2.20E-18	53	-9.205252756	hypothetical protein		U			
AN	T3 - T4	KAG0703349.1	90.9	2.90E-188	363	-8.929850906	Fructose-bisphosphate aldolase	Aldolase physically associates with vacuolar H-ATPase/glycolysis pathway	EX	1	6	0

Supplemental Table 2.3: Gene names and descriptions for top 10 DEGs in terms of log2FoldChange expression for *T. magna* for both tissues (G7: Posterior gill; AN: Antennal gland) and all time intervals (T0 – T1: Baseline – Acute; T1 – T3: Acute – Extreme; T3 – T4: Extreme – Recovery). Only genes with BLAST hits are shown. Gene types (EX: expanded gene family; U: unassigned to any orthogroup) and copy numbers per species (TC: *T. celeste*; TM: *T. magna*; GN: *G. natalis*) are also shown.

Tissue	Time interval	Target accession	Sequence identity	e-value	Length	log2FoldChange	Protein name	Description	Gene type	Copy number (TC)	Copy number (TM)	Copy number (GN)
G7	T0 - T1	EGZ74219.1	97.2	6.20E-09	36	24.30399782	hypothetical protein		U			
G7	T0 - T1	YP_009107263.1	77.9	5.50E-27	95	12.86534008	NADH dehydrogenase subunit 5 (mitochondrial)	ATP generation	U			
G7	T0 - T1	YP_009535655.1	84.1	8.80E-86	208	11.62756527	Cytochrome C oxidase subunit 1 (mitochondrial)	Electron transport chain	U			
G7	T0 - T1	KPV71458.1	90	1.60E-12	40	11.61158546	hypothetical protein		U			
G7	T0 - T1	OFQ97201.1	94	2.20E-15	50	11.51054721	hypothetical protein		U			
G7	T0 - T1	QRK27379.1	84.9	2.10E-21	73	10.75707609	NADH dehydrogenase subunit 4 (mitochondrial)	ATP generation	U			
G7	T0 - T1	ORD93174.1	96.5	1.20E-41	85	10.29791325	hypothetical protein		U			
G7	T1 - T3	MPC73101.1	86.6	2.50E-32	82	13.58716419	Signal recognition particle (SRP) protein	Protein localization to proper cellular destinations	EX	1	1	2
AN	T0 - T1	KAG8170603.1	52.6	1.70E-17	133	24.07390478	hypothetical protein		U			
AN	T0 - T1	MPC50955.1	80.4	2.80E-39	97	23.24827991	putative thiosulfate sulfurtransferase (TST) (mitochondrial)	Formation of iron-sulfur complexes, cyanide detoxification or modification of sulfur-containing enzymes.	EX	2	1	3
AN	T0 - T1	KAG0722448.1	90.7	2.90E-220	408	22.67300886	tubulin polyglutamylase TTLL4	catalyzes the posttranslational formation of polyglutamyl side chains	EX	2	2	4
AN	T0 - T1	MPC13266.1	66	4.80E-92	1059	11.23040395	hypothetical protein		EX	3	2	3
AN	T1 - T3	XP_042204822.1	94.5	5.30E-67	201	24.86388684	zinc finger matrin-type protein 2-like	Involved in pre-mRNA splicing as a component of the spliceosome.	EX	1	4	4
AN	T1 - T3	KAG0716941.1	73.1	1.40E-127	327	23.48642385	hypothetical protein		EX	3	1	5
AN	T1 - T3	KAG7172599.1	45	0.00096	60	23.07693708	hypothetical protein		U			
AN	T1 - T3	WP_208094023.1	72.9	5.40E-84	310	13.81598152	hypothetical protein		EX	6	5	2

Supplemental Table 2.4: Gene names and descriptions for top 10 DEGs in terms of log2FoldChange expression for *G. natalis* for both tissues (G7: Posterior gill; AN: Antennal gland) and all time intervals (T0 – T1: Baseline – Acute; T1 – T3: Acute – Extreme). Only genes with BLAST hits are shown. Gene types (EX: expanded gene family; U: unassigned to any orthogroup) and copy numbers per species (TC: *T. celeste*; TM: *T. magna*; GN: *G. natalis*) are also shown.

Supplemental Table 2.5: Gene names and descriptions for top 10 genes in terms of Module Membership for all modules with significant correlations to species. Only genes with BLAST hits are shown.

Module	Target accession	Protein name	Description	Length	Sequence identity	e-value	|Module Membership|
ME1	XP_042216344.1	fumarylacetoacetase-like	phenylalanine and tyrosine degradation	416	80	3.90E-206	0.977858078
	KAG0725090.1	Elongin-C	subunit of the transcription factor B (SIII) complex	117	99.1	3.20E-60	0.971496157
	MPC37806.1	hypothetical protein		148	81.1	1.00E-56	0.971187492
	ROT79510.1	RMD5-like protein A	negative regulation of gluconeogenesis; ubiquitin protein ligase activity	263	93.9	1.20E-139	0.956820355
	KAG0724140.1	Endoribonuclease LACTB2	RNA phosphodiester bond hydrolysis	302	68.2	2.80E-114	0.954221121
	KAG0704322.1	Ras-like GTP-binding protein Rho1	actin cytoskeletal organization	194	99	1.90E-104	0.952294347
	XP_042219201.1	short/branched chain specific acyl-CoA dehydrogenase, mitochondrial-like isoform X1	fatty acid metabolic process; isoleucine catabolic process	418	83	3.00E-193	0.950659183
	AIS24843.1	T-complex protein 1 subunit alpha	TRiC complex mediates the folding of WRAP53/TCAB1, thereby regulating telomere maintenance	555	95	1.40E-292	0.949287268
	MPC23252.1	RPII140-upstream gene protein-like	RNA polymerase	278	64.7	1.30E-94	0.947865926
	XP_037797882.1	phytanoyl-CoA dioxygenase, peroxisomal-like isoform X2	fatty acid metabolism	296	79.7	8.40E-145	0.945516155
ME3	KAG0724012.1	COP9 signalosome complex subunit 6	protein deneddylation	293	96.6	1.70E-156	0.95228519
	ALP46200.1	nascent polypeptide-associated complex alpha	negative regulation of protein localization to endoplasmic reticulum	209	100	1.00E-81	0.94535869
	ROT63357.1	grpE protein homolog 1, mitochondrial-like	controls the nucleotide-dependent binding of mitochondrial HSP70 to substrate proteins	143	81.1	1.60E-58	0.93943885
	XP_042231848.1	cleavage and polyadenylation specificity factor subunit 5-like	activator of the pre-mRNA 3'-end cleavage and polyadenylation processing required for the maturation of pre-mRNA into functional mRNAs	228	99.1	2.00E-129	0.93924806
	ROT69872.1	hypothetical protein		284	87.3	1.20E-139	0.93869091
	KAG0717700.1	U4/U6 small nuclear ribonucleoprotein Prp31	Involved in pre-mRNA splicing as component of the spliceosome	494	91.1	1.60E-243	0.93431374
	KAG0723646.1	Hydroxysteroid dehydrogenase-like protein 2	oxidoreductase activity	420	85.7	2.00E-204	0.92884357
	XP_042210450.1	trigger factor-like	Chaperone activity	71	40.8	0.00028	0.9279747
	XP_042215298.1	60S ribosomal protein L31-like	cytoplasmic translation	125	94.4	1.70E-58	0.92657039
	KAG0710068.1	Short-chain specific acyl-CoA dehydrogenase, mitochondrial	aerobic process breaking down fatty acids into acetyl-CoA and allowing the production of energy from fats	434	86.2	9.80E-206	0.92628632

ME4	EGW03135.1	Elongation factor 2	essential factor for protein synthesis	395	100	1.60E-232	0.979095787
	XP_042236946.1	L-2-hydroxyglutarate dehydrogenase, mitochondrial-like isoform X3	oxidoreductase activity	402	82.8	3.50E-202	0.967639558
	XP_042226004.1	transitional endoplasmic reticulum ATPase	Autophagy, DNA damage, DNA repair, Transport, Ubl conjugation pathway	794	98.4	0	0.966802196
	MPC32893.1	Protein ROP	Transcription regulation	356	94.7	5.00E-192	0.965047157
	MPC14080.1	Histone deacetylase complex subunit SAP18	mRNA processing, mRNA splicing, Transcription, Transcription regulation	150	97.3	3.40E-79	0.955379515
	QDE54938.1	eukaryotic translation initiation factor 3 subunit B, partial	Protein biosynthesis, Initiation factor, RNA-binding	562	93.1	0	0.954118167
	KAG0710311.1	AP-2 complex subunit mu	Endocytosis, Protein transport, Transport, adaptor protein complex	293	100	1.90E-163	0.952188669
	KAG0713716.1	CDGSH iron-sulfur domain-containing protein 1	regulating maximal capacity for electron transport and oxidative phosphorylation	110	84.5	6.30E-50	0.948170144
	ACY66390.1	FK506-binding protein 1A	immunoregulation and cellular processes involving protein folding and trafficking	110	87.3	7.40E-49	0.947954257
	MPC68061.1	hypothetical protein		216	88	3.90E-98	0.944352237
ME5	KAG0714100.1	Tyrosyl-DNA phosphodiesterase 2	DNA repair	363	72.2	6.10E-145	0.940638747
	KAG0725244.1	T-complex protein 1 subunit delta	The TRiC complex mediates the folding of WRAP53/TCAB1, thereby regulating telomere maintenance	534	95.1	1.50E-279	0.937633129
	QHD64853.1	heat shock protein 60	heat inducible and act as a molecular chaperones, assisting in protein folding	556	97.3	6.90E-294	0.932209271
	XP_042230358.1	reticulon-4-interacting protein 1 homolog, mitochondrial-like isoform X1	Plays a role in the regulation of retinal ganglion cell (RGC) maturation and neurite outgrowth	370	68.6	9.50E-145	0.918332399
	KAG0726339.1	Methylthioribose-1-phosphate isomerase	L-methionine biosynthesis via salvage pathway	348	81.9	1.80E-158	0.91789729
	ANN46488.1	calnexin	Chaperone activity	442	95.9	4.30E-168	0.914040218
	MPC18378.1	Centromere protein X	Cell cycle, Cell division, DNA damage, DNA repair, Mitosis	65	76.9	7.10E-18	0.911495751
	KAG0729193.1	Growth arrest and DNA damage-inducible proteins-interacting protein 1	Acts as a negative regulator of G1 to S cell cycle phase progression by inhibiting cyclin-dependent kinases	264	62.9	7.90E-63	0.90744937
	MPC17875.1	Oligosaccharyltransferase complex subunit ostc-B	involved in protein glycosylation and protein modification	150	94	1.10E-73	0.905814954
	KAG0726480.1	Mitochondrial import inner membrane translocase subunit TIM14	Chaperone activity	84	91.7	1.80E-34	0.903171201
ME6	XP_042209693.1	40S ribosomal protein SA-like	Cell adhesion, cytoplasmic translation, ribosomal small subunit assembly	212	96.2	1.70E-112	0.969824031
	ACG60901.1	eukaryotic translation initiation factor 6	stimulatory translation initiation factor downstream insulin/growth factors	245	86.5	4.50E-116	0.961991376
	XP_042222113.1	60S ribosomal protein L24-like	encodes a ribosomal protein that is a component of the 60S subunit	105	94.3	7.10E-49	0.954621492
	XP_037784230.1	eukaryotic translation initiation factor 3 subunit E-like	Component of the eukaryotic translation initiation factor 3 (eIF-3) complex	424	95.3	4.00E-232	0.95200711
	XP_042233103.1	transcription elongation factor S-II-like isoform X1	part of complex required to increase the catalytic rate of RNA polymerase II transcription	301	83.4	1.30E-110	0.943234722
	KAG7162228.1	Mediator of RNA polymerase II transcription subunit 21-like	a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes.	147	98	6.60E-70	0.939832243
	MPC31537.1	hypothetical protein		95	98.9	2.90E-40	0.931888824
	MPC56548.1	hypothetical protein		381	57.7	7.80E-85	0.929501031
	ADE60733.1	myosin essential light chain	regulating the actin-myosin interaction of smooth muscle	147	98	1.90E-73	0.925459286
	KAG0714221.1	Mitochondrial-processing peptidase subunit alpha	Substrate recognition and binding subunit of the essential mitochondrial processing protease	324	90.1	6.60E-171	0.916987419

Genomics of red-blue color polymorphisms in coconut crabs (*Birgus latro*)

Victoria M. Watson-Zink[1,2†], Tim Caro[3], Richard K. Grosberg[1], Shin-ichiro Oka[4], Rashid S. Rashid[5], Ulrike Kloiber[6], & Rachael A. Bay[1]

[1] Department of Evolution and Ecology, University of California, Davis, United States

[2] Stanford University, Department of Biology, United States

[3] School of Biological Sciences, University of Bristol, United Kingdom

[4] Okinawa Churashima Foundation Research Center, Okinawa, Japan

[5] Department of Forestry, Pemba, Zanzibar, Tanzania

[6] Chumbe Island Coral Park, Stonetown, Zanzibar, Tanzania

[†]Corresponding Author: Victoria M. Watson-Zink

vmmorgan@ucdavis.edu

1. Introduction:

Phenotypic color polymorphisms are widespread in nature (White and Kemp 2016), occurring in a diversity of taxa including hares (Giska et al. 2019), lizards (Brock et al. 2022), birds (Aguillon et al. 2021), grasshoppers (Karpestam et al. 2016), tunicates (Sabbadin and Graziani 1967), mollusks (Williams 2017), isopods (Merilaita 2001), and crabs (Krause-Nehring et al. 2010). The relative frequencies of different morphs within and among populations reflect the impacts of fundamental evolutionary processes including natural and sexual selection, gene flow, mutation, and genetic drift (Cain and Sheppard 1954; Svensson 2017). In principle, color variation can affect individual fitness through various mechanisms underlying protective coloration or through mate choice, but how spatially and temporally variable selection maintains genetically based color variation remains an elusive and classic question in ecological genetics (e.g., Cain and Sheppard 1954; Losey et al. 1997; Karpestam et al. 2016).

The coconut crab (*Birgus latro*; Decapoda: Anomura: Paguridea: Coenobitidae) is the world's largest terrestrial arthropod with up to a 1m leg span, a maximum weight of 4kg, and a lifespan of up to 60 years (Drew et al 2010; Laidre 2018). Coconut crabs live on tropical islands across the Indian Ocean and the Western Pacific (Fig. 3.1A), principally existing as either red or blue morphs across their range. "Red" morphs are so-called because their dorsal cephalothorax varies from deep red to bright red-orange, and "blue" morphs have dark blue to near black dorsal cephalothoraxes. Their ventral surfaces show substantial variation too: the ventral color of red coconut crabs is usually white but may incorporate patches ultramarine blue, while the ventral color of blue coconut crabs is always a shade of blue varying from light blue to deep ultramarine blue to shades of violet (Fig. 3.2). Previous research tested several classical adaptive hypotheses proposed to explain color polymorphisms in this system, including differences between sexes

(where color would function as a signal for mate choice), differences between body sizes (where color would reflect ontogenetic changes), differences in aggressive behavior and pinch force (where color would function as an intraspecific signal for relative competitive ability), and differences between background matching ability and the occupation of different island microhabitats (Caro and Morgan 2018). We found no significant associations between color morphs across any of the ecological and life-history metrics they examined, and could not identify selective forces maintaining the red-blue color polymorphism in this system (Nokelainen et al. 2018; Caro and Morgan 2018; Caro et al. 2019; Caro 2021).

Studies on other crustaceans suggest that color variation in coconut crabs may reflect differences in diet. For example, normally tiger shrimp (*Penaeus monodon*) possess a red and black-banded exoskeleton, but when fed commercially available diets low in carotenoids (specifically astaxanthin), the shrimp turned sky blue (*i.e.*, "blue disease"), reflecting their nutritional deficiency (Howell and Matthews 1991). In the case for *Birgus*, adults of both color morphs typically live sympatrically and both feed on fallen fruits, seeds, and nuts (including coconuts), and occasionally on carrion (Cumberlidge et al. 2022), making a dietary explanation untenable. Furthermore, the body color patterning of coconut crabs is retained through successive molts, indicating that color is not a trait that fluctuates throughout the lifespan of an individual crab (Oka et al. 2015; Hamasaki et al. 2014).

Frequency-dependent selection has also been proposed as a potential mechanism maintaining the red:blue color polymorphism in this system; however, adult coconut crabs have few natural predators (except for humans), and juvenile coconut crabs of both colors begin their lives as tan and cream-colored shell bound hermit crabs, a color that camouflages them in coral rubble, their

typical juvenile habitat (Hamasaki et al. 2014). They only begin acquiring their adult coloration after they abandon their protective gastropod shells.

Based on the global distribution of color morphs and the existence of populations on islands at the periphery of *Birgus*' distributional range that contain only one of the two color morphs (*e.g.*, Okinawa Island, all blue (pers. obs.), and Papua New Guinea, all blue (Takdir, M., pers. comm.)), and the absence of any obvious selective explanation for the distribution and frequencies of *Birgus* color morphs across their range, genetic drift and founder effects may also be playing some role in this system. Although adult coconut crabs are fully terrestrial and air-breathing, larval coconut crabs are planktonic, passing through multiple developmental stages in the surface waters of the ocean before migrating back ashore as juvenile crabs (Watson-Zink 2021). The duration of their planktonic phase, which lasts for about four weeks, suggest extensive gene flow among populations, and little opportunity for genetic drift to operate. Nevertheless, Laverty (1995) found clear evidence for geographically structured genetic variation in this species across its Indian and Pacific Ocean distributional range. Thus, on islands at the periphery of *Birgus*' range, limited gene flow might explain the absence of one or the other morph in such populations. In comparison, Caro and Morgan (2018) reported an approximate 3:1 ratio of red:blue crabs on all the islands they sampled (*i.e.*, Zanzibar, Tanzania, and Christmas Island, Australia), and Appoo et al. (2021) found a 4:1:1 ratio of red: red-blue: blue coconut crabs on Aldabra Atoll, Seychelles, suggesting that in nearly all documented cases where the color morphs co-occur, populations tend to have more red crabs than blue. Therefore, the numerical constancy of the red:blue polymorphism across geographic space might reflect higher overall gene flow between island populations.

In this study, we sought to identify the genomic basis of the red-blue color polymorphism in this system as a means of understanding, at a genetic and functional level, how this polymorphism might be maintained. Understanding the genetic architecture underlying this color polymorphism may allow us to move closer to elucidating the adaptive significance (or lack thereof, Caro 2021) of coloration in coconut crabs.

Prior work on crustacean coloration has repeatedly singled out a protein that appears to play a critical and conserved role in determining shell color in crustaceans – crustacyanin (CRCN). Wade et al. (2009) described how the two subunits of CRCN (A and C) and its carotenoid ligand, astaxanthin, combine to form a multimeric protein complex that is critical in determining the array of external shell colors in lobsters (*Homarus gammarus*), and that CRCN was an evolutionary innovation unique to the malacostracan crustaceans. Furthermore, the spatial regulation of the expression of CRCN genes in the hypodermis of the crustacean exoskeleton is a major contributor to the diversity of color patterns observed in many, if not all, malacostracans (Wade et al. 2009). CRCN proteins are also linked to the production and modification of carapace color, and correlations between the presence of CRCN genes, astaxanthin, and the resulting crustacean shell colors confirm this link between phenotype and genotype in lobsters (Ferrari et al. (2012). Most recently, Budd et al. (2017) isolated 35 new CRCN orthologs from 12 species of penaeid shrimp and found that while this gene family has undergone significant expansion and sequence divergence, the structure and functional role of CRCN proteins in the production of shrimp shell coloration has been strictly conserved across species.

In this study, we aimed to extend previous work on the red-blue color polymorphism in *Birgus latro* by using genomic tools to (1) characterize genetic structure and the overall role of drift, founder effects, and gene flow in producing the distribution of color variation, (2)

determine if there is a simple underlying genetic basis to color variation, and (3) specifically ascertain whether the crustacyanin gene plays a comparable role in coloration as it does in lobsters, shrimp, and other crustaceans. We examined genomic divergence between red and blue coconut crabs across their genome to determine if there are signals of population structure (i.e. reduced gene flow) between nearby and distant islands. Such a genome-wide scan of sequence divergence between the color morphs within a population would also allow us to more directly determine if color is a signal for mate choice -- because matings are difficult to observe directly due to the cryptic nature of this species during its breeding season (Helfman 1977), if red and blue coconut crabs were differentiated across their genomes, it would support the hypothesis that coconut crabs mate assortatively by color.

In addition, if single nucleotide polymorphisms (SNPs) differ more between red and blue morphs in crustacyanin-containing regions than randomly selected SNPs across the *Birgus* genome, then this would support the hypothesis that CRCN sequence variation may functionally explain color differences between the morphs.

2. Materials and Methods

2.1 Sample collection and sequencing

We assembled a *de novo* transcriptome and genome from a single coconut crab individual for this study. This individual was a "sunset" (*i.e.,* red dorsum, blue ventrum) intermolt adult male collected by hand from Christmas Island National Park, Australia in September 2017 (Fig. 3.1C). After sacrificing the crab, we dissected it on ice and sampled tissue from its walking legs, hepatopancreas, hindgut, and posterior gills. We stored all samples individually in 0.5mL

cryovials containing Trizol Reagent (Invitrogen), and flash froze them in liquid N_2. We then transported all samples back to the USA on dry ice held in a 20L dewar storage canister.

To sequence the transcriptome, we homogenized muscle samples for one minute in 100µL Trizol reagent (Invitrogen) using a Qiagen Tissue Lyser and beads at 50Hz, and then added an additional 500µL of Trizol to each tube before centrifuging at 16,000G for 30 seconds. We removed the supernatant and placed it into a clean centrifuge tube, and used the NEBNext Ultra Directional II kit (New England BioLabs, Ipswich, MA) to prepare cDNA libraries, following the manufacturer's instructions. We measured sample concentration and purity using the Agilent 2100 Bioanalyzer (Agilent RNA Eukaryote Total RNA Nano Kit), and the Invitrogen Qubit RNA Broad Range Assay Kit. We then sent all muscle tissue samples to Novogene for Illumina HiSeq 4000 sequencing, and they were paired-end sequenced at the length of 150bp. We aimed to sequence 40M raw reads (12Gb) per sample.

To sequence the genome, we sent frozen whole muscle tissue samples to the University of California, Davis Genome and Biomedical Sciences Facility for HMW DNA extraction, library preparation, and three flow cells of Pacific Biosciences HiFi ultra long-read sequencing. At the time of sequencing, the total size of the *Birgus* genome had yet to be published, so we used a kmer-based approach to estimate coverage and targeted 10X to 15X coverage.

For the resequencing study, we sampled tissue from toe clippings from live adult coconut crabs across the global range of this species (N = 121). To obtain toe clippings, we used garden shears to remove approximately 3cm from one of the posterior walking legs, and stored each clipping individually in 1.5mL microcentrifuge tubes containing 100% EtOH and held at ambient temperatures. Coconut crab color morphs were determined by eye based on coarse dorsal coloration of the cephalothorax (Fig. 3.2). For the purposes of this study, we considered

any coconut crab with an orange, red, or dark red dorsum to be a “red” crab, irrespective of its ventral color, while we considered all blue or dark blue dorsum crabs to be “blue” crabs, irrespective of their ventral color.

In May 2018, we sampled 21 blue coconut crabs near Churaumi Aquarium in Okinawa, Japan (Fig. 3.1D), and in July 2018, we sampled 17 crabs on Chumbe Island, Tanzania (8 blue; 9 red), 5 crabs on Misali Island, Tanzania (1 blue; 4 red), 35 crabs on Kisiwa Panza, Tanzania (9 blue; 26 red), 11 crabs on Pemba Island, Tanzania (1 blue; 10 red), and 32 crabs on Bongoyo and Mbudya Islands near Dar es Salaam, Tanzania (8 blue; 24 red) (Fig. 3.1B).

After extracting all 121 samples, we then selected 96 samples (41 blue; 55 red) that were of sufficient quality for whole genome resequencing based on Thermo Scientific Nanodrop 260/230 ratios (between 1.5 – 2.27), and DNA concentration results (>20ng/μL, determined using the Invitrogen Qubit DNA Broad Range quantification assay, following the manufacturer’s instructions). We sent samples to Texas A&M Genomics and Bioinformatics Service (TxGen) for Illumina NovaSeq paired-end sequencing. Sequence cluster identification, quality prefiltering, base calling and uncertainty assessment were done in real time using Illumina's NCS 1.0.2 and RFV 1.0.2 software with default parameter settings. Sequencer .cbcl basecall files were demultiplexed and formatted into .fastq files using script configureBclToFastq.pl from bcl2fastq 2, version 2.19.0.

2.1 Transcriptome and genome assembly and mapping

We used the default parameters in FastQC (Andrews 2010) to visualize read quality, and eliminated low quality reads with Trimmomatic version 0.36 (Bolger et al. 2014) using parameters for paired reads. with Trimmomatic version 0.36 (Bolger et al. 2014) using

parameters for paired reads. We removed leading and trailing low-quality base pairs (LEADING:2, TRAILING:2) using a sliding window of 4 base pairs, removing base pairs when quality dropped below 2 (SLIDINGWINDOW:4:2), and kept reads of a minimum length of 25bp (MINLEN:25). We used default parameters in Trinity (Grabherr et al. 2011) to assemble a *de novo* muscle tissue transcriptome, and then assessed the quality of the assembly using the BUSCO Arthropoda database (Manni et al. 2021). In the absence of a long-read contiguous reference genome for *Birgus*, we created the first *de novo* long-read genome assembly for this species using pbipa.

2.2 SNP calling and analysis of genetic divergence at *CRCN* loci and across random sampling of genome

Due to the published size of the *Birgus* genome (6.22Gbp; Veldsman et al. 2021), and the large size of our own assembly (3.3Gbp), we focused on a subset of the genome for the purposes of this study. Since previous work has described the role of crustacyanin (CRCN) in crustacean exoskeleton coloration in lobsters (Ferrari et al. 2012), penaeid shrimp (Budd et al. 2017), and a diversity of malacostracan crustaceans (Wade et al. 2009), we focused our analysis on all 40 contigs that possessed CRCN-like sequences, and on 25 randomly selected non-CRCN-containing contigs for global genomic comparisons (N = 65 contigs).

To ascertain which contigs contained CRCN-like reads, we first downloaded the publicly available CRCN protein sequences from UniProt for the blue lobster, *Homarus gammarus* (Subunit A2: P80007; Subunit C1: P80029), and reverse transcribed these sequences to obtain cDNA sequences for both CRCN subunits. Using BLAST/2.11.0, we used this sequence as a search term across the *Birgus* transcriptome to find similar reads. Once we found matching reads

in the *Birgus* transcriptome (N = 13 transcripts; 7 transcripts for subunit A2, and 6 transcripts for subunit C1), we used BLAST to use the *Birgus*-specific CRCN sequences to identify regions in the genome that contained CRCN-like sequences.

We used the default settings in bwa (Li and Durbin 2009) to map the reads from the 96 resequenced individuals back to the genome assembly, and then sorted the reads from each sample using samtools (Li et al. 2009). We then used ANGSD (Korneliussen et al. 2014) to estimate the site-frequency spectra across all 65 selected scaffolds after filtering out reads that were present in less than half of the individuals ("minInd": 26 for red morphs and 21 for blue morphs), reads that mapped to multiple locations ("remove_bads" and "uniqueOnly"), and reads of quality less than 20 ("minMapQ" and "minQ" of 20). The site-frequency spectrum is a metric that summarizes allele frequencies within a sample of DNA sequences, therefore, we calculated it separately for our pooled red crabs and pooled blue crabs. We then used ANGSD to calculate the value of the 2D site-frequency spectrum prior, and to calculate per-site F_{st} at each SNP.

To get a general idea of genetic divergence and gene flow across our samples, we selected SNPs across a random 2,000 contigs to plot using a principal component analysis (PCA).

To identify if genomic divergence at CRCN loci is associated with differences in exoskeleton coloration between red and blue coconut crabs, we used a custom Python script to sort the SNPs into three categories: (1) SNPs located within the bounds of a CRCN sequence, (2) SNPs on a CRCN-containing scaffold but not within the bounds of a CRCN gene, and (3) SNPs on randomly selected scaffolds that do not contain a CRCN-like sequence. Since the fewest number of SNPs across these categories was found in the first grouping (i.e., within the bounds of a CRCN sequence; N = 9,381 SNPs), we randomly selected 9,381 SNPs from the remaining

two categories and performed a one-way ANOVA to determine if there were statistically significant differences among mean per-site F_{st} values across all three groupings.

3. Results

3.1 *de novo* genome and transcriptome assembly

Our genome assembly was 3.3 billion base pairs that were assembled into approximately 22 thousand scaffolds with a contig N50 of 189Kbp (Table 3.1). For the transcriptome, we assembled 216,708 transcripts with a contig N50 of 1,053bp. We found that the transcriptome was 80% complete when compared to the BUSCO Arthropoda database.

3.2 Gene flow between color morphs

For the resequencing portion of this study, we obtained 3.7Gb of sequence data per sample, which averaged out to approximately 1X coverage per individual. Across the 2,000 random contigs, all Tanzanian samples (*i.e.*, Chumbe, Misali, Kisiwa Panza, Pemba, Dar es Salaam) displayed substantial gene flow, and there were no signals of clustering by Tanzanian island or by color morph (Fig. 3.3a). There was, however, significant differentiation between all the Tanzanian coconut crabs and the Okinawa, Japanese coconut crabs (Fig. 3.3b).

3.3 Genetic divergence at *CRCN* loci

We found no significant differences between mean per-site F_{st} values for SNPs across the three examined categories: (1) SNPs located within the bounds of a CRCN sequence, (2) SNPs on a CRCN-containing scaffold but not within the bounds of a CRCN gene, and (3) SNPs on randomly selected contigs that do not contain a CRCN-like sequence ($F(1, 2) = [2.5912]$, $p = 0.0749$) (Fig. 3.4).

Discussion

Color polymorphisms provide an excellent opportunity to investigate fundamental evolutionary processes such as natural and sexual selection, gene flow, and genetic drift (Cain and Sheppard 1954; Svensson 2017; Estévez et al. 2020; Dyson et al. 2020). Although they are widespread in nature, the adaptive value and ecological consequences of color polymorphisms are often difficult to ascertain. Furthermore, the evolutionary and genetic mechanisms maintaining color polymorphisms across variable natural environments is a long-standing question in evolutionary biology (Hedrick 2006).

In this study, we set out to understand whether evolutionary processes such as genetic drift and limited gene flow between island populations plays a significant role in the global distribution of color morphs across the Indian and West Pacific Oceans. Based on our PCA results from 2,000 random genomic contigs, we inferred substantial differences in gene flow between geographically distant regions that appear to correspond with reduced color morph diversity on islands at the periphery of *Birgus latro*'s global range, notably between the far westerly populations in the western Indian Ocean and the northerly population in the Sea of Japan. Consistent with Lavery et al. (1995), at large geographical distances, the *Birgus* population displays isolation by distance, but at much smaller geographic scales (*i.e*., up to 200 km) populations display panmixia (as in Tanzanian populations, where we found no substantial genomic differentiation between islands even as far away as 200km). Therefore, founder effects and limited gene flow between populations might be useful for understanding some of the differences that have been reported in color morph frequencies across isolated island complexes in *Birgus*' range.

We used a genome-wide comparative approach to determine if there was population structure between blue and red coconut crabs. Our results showing lack of genomic differentiation between color morphs suggests that there is substantial gene flow between sympatric color morphs. This finding suggests that red crabs and blue crabs interbreed, contrary to what would be expected with strong positive assortative mating with respect to color (but see Nosil (2005) for an example of assortative mating that occurs in the face of substantial gene flow), and in line with what might be expected in a species that is nocturnal across much of its range.

This study only considered 2,000 out of the total 21,909 scaffolds in our long-read assembly, which itself represents only half of the total size of the coconut crab genome (estimated to be 6.22Gbp (Veldsman et al. 2021)). Consequently, it is critical to assemble a more contiguous genome for this species, making it possible to achieve a more complete scan of genomic scaffolds that could uncover stable genomic differences between the color morphs and reveal the potential adaptive significance of this trait in coconut crabs.

We also used a candidate gene approach to search for sequence divergence between red and blue coconut crabs at putative crustacyanin (CRCN) loci. CRCN has been implicated in exoskeleton coloration in most, if not all, malacostracan crustaceans, and may represent a key innovation underlying the evolutionary success of this clade due to the critical role shell color and pattern play on survival (*e.g.*, Nokelainen et al. 2019), reproduction (e.g., Christy 1988), and communication (e.g., Todd et al. 2011) in crustaceans (Wade et al. 2009; Caro 2018). Surprisingly, we found that SNPs within the bounds of nominal CRCN genes in coconut crabs were no more likely to be significantly divergent between color morphs than SNPs randomly selected from across the genome, which implies that red/blue coloration in coconut crabs does

not appear to result from sequence variation at CRCN loci across the genome. Nevertheless, CRCN may play an important role in regulating differences in coloration in *Birgus*, albeit at a level of biological organization downstream of DNA sequence variation (i.e., regulatory elements, post-translational modification, gene expression, etc.). For example, Budd et al. (2017) found that CRCN isoforms were differentially expressed in *Penaeus monodon*, and that downregulation of specific isoforms produced subtle phenotypic differences, suggesting that multiple copies of CRCN genes within a species may be critical to color production. In a similar fashion, we found 40 unique copies of CRCN across the *Birgus* genome, likely potentially reflecting different paralogs of this gene. Future studies should use a comparative differential gene expression approach to determine how expression levels of different CRCN paralogs may be related to differences in coloration in this species.

Finally, we have produced novel genomic resources for a species that has recently been listed as "Vulnerable" on the IUCN Red List of Threatened Species principally due to the effects of overharvesting and habitat fragmentation (Cumberlidge et al. 2022). These resources include the first long-read draft genome assembly for the coconut crab and an 80% complete transcriptome from the same individual that was sequenced for the genome, and will be critical to future studies that seek to understand the incredible terrestrial adaptations displayed by these charismatic crustaceans (Watson-Zink 2021).

LITERATURE CITED

Aguillon, S.M., Walsh, J., & Lovette, I.J., (2021). Extensive hybridization reveals multiple coloration genes underlying a complex plumage phenotype. *Proceedings of the Royal Society B* 288(1943): 20201805.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." *Journal of Molecular Biology*, 215:403-410.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Appoo, J., Cagua, E., van de Crommenacker, J., Burt, A., Pistorius, P., Taylor, F., Richards, H., Haupt, P., Doak, N., Fleischer-Dogley, F., & Bunbury, N. (2021). Population dynamics of the coconut crab *Birgus latro* on Aldabra Atoll, Seychelles. *Marine Ecology Progress Series*, 665, 103-113.

Budd, A.M., Hinton, T.M., Tonks, M., Cheers, S., & Wade, N.M. (2017). Rapid expansion of pigmentation genes in penaeid shrimp with absolute preservation of function. *Journal of Experimental Biology*, 220 (22): 4109–4118.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

Brock, K.M., McTavish, E.J., & Edwards, D.L. (2022). Color Polymorphism is a Driver of Diversification in the Lizard Family Lacertidae, *Systematic Biology*, Volume 71, Issue 1, January 2022, Pages 24–39.

Cain, A.J., & Sheppard, P.M. Natural selection in *Cepaea*, *Genetics*, Volume 39, Issue 1, 1 January 1954, Pages 89–116.

Caro, T., 2018. The functional significance of coloration in crabs. *Biological Journal of the Linnean Society*, *124*(1), pp. 1-10.

Caro, T. & Morgan, V.M. (2018). Correlates of color polymorphism in coconut crabs, *Birgus latro*. *Zoology* 129, Pgs. 1-8.

Caro, T., Cluff, E., & Morgan, V.M. (2019). Colour polymorphism and protective coloration in coconut crabs. *Ethology Ecology & Evolution* 31 (6): Pgs. 1-12.

Caro, T. (2021). When animal coloration is a poor match. *Evolutionary Ecology*, 35: 1-13.

Christy, J.H., 1988. Pillar function in the fiddler crab *Uca beebei* (II): competitive courtship signaling. *Ethology*, *78*(2), pp.113-128.

Cumberlidge, N., Caro, T., Watson-Zink, V.M., Naruse, T., Ng, P.K.L., Orchard, M., Rahayu, D.L., Wowor, D., Yeo, D.C.J., & White, T. (2022). Troubled giants: The updated conservation status of the coconut crab (*Birgus latro*). *Raffles Bulletin of Zoology* 70: 1-21.

Drew, M.M., Harzsch, S., Stensmyr, M., Erland, S., & Hansson, B.S. (2010). A review of the biology and ecology of the Robber Crab, *Birgus latro* (Linnaeus, 1767) (Anomura*: Coenobitidae*), *Zoologischer Anzeiger*, 249(1), 45–67.

Dyson, M.L., Perez, D.M., Curran, T., McCullough, E.L., & Backwell, P.R.Y. (2020). The role of claw color in species recognition and mate choice in a fiddler crab. *Behavioral Ecology and Sociobiology* 74, 116.

Estévez, D., Kozminsky, E., Carvajal-Rodríguez, A., Caballero, A., Faria, R., Galindo, J. & Rolán-Alvarez, E. (2020) Mate Choice Contributes to the Maintenance of Shell Color Polymorphism in a Marine Snail via Frequency-Dependent Sexual Selection. *Frontiers in Marine Science,* 7:614237.

Ferrari, M., Folli, C., Pincolini, E., McClintock, T. S., Rössle, M., Berni, R., & Cianci, M. (2012). Structural characterization of recombinant crustacyanin subunits from the lobster Homarus americanus. *Acta crystallographica. Section F, Structural biology and crystallization communications*, *68* (Pt 8), 846–853.

Giska, I., Farelo, L., Pimenta, J., Seixas, F. A., Ferreira, M. S., Marques, J. P., Miranda, I., Letty, J., Jenny, H., Hackländer, K., Magnussen, E., & Melo-Ferreira, J. (2019). Introgression drives repeated evolution of winter coat color polymorphism in hares. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(48), 24150–24156.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, *29* (7), 644–652.

Hamasaki, K., Ishiyama, N., Yamashita, S., & Kitada, S. Survival and Growth of Juveniles of the Coconut Crab *Birgus latro* Under Laboratory Conditions: Implications for Mass Production of Juveniles, *Journal of Crustacean Biology*, Volume 34, Issue 3, 1 May 2014, Pages 309–318.

Hedrick, P.W., 2006. Genetic polymorphism in heterogeneous environments: the age of genomics. *Annual Review of Ecology Evolution and Systematics*, 37, pp.67-93.

Helfman, G.S., (1977). Copulatory behavior of the coconut or robber crab *Birgus latro* (L.)(Decapoda Anomura, Paguridea, Coenobitidae). *Crustaceana*, pp.198-202.

Howell, B.K., & Matthews, A.D. (1991). The carotenoids of wild and blue disease affected farmed tiger shrimp (*Penaeus Monodon*, Fabricus), *Comparative Biochemistry and Physiology,* 98(2-3), 375–379. doi:10.1016/0305-0491(91)90193-h .

Karpestam, E., Merilaita, S. & Forsman, A, (2016). Colour Polymorphism Protects Prey Individuals and Populations Against Predation. *Scientific Reports,* 6, 22122 (2016).

Korneliussen, T.S., Albrechtsen, A. & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15, 356.

Krause-Nehring, J.J., Starck, M.A., & Palmer, R. (2010). Juvenile colour polymorphism in the red rock crab, Cancer productus: patterns, causes, and possible adaptive significance, *Zoology*, Volume 113, Issue 3, 2010, Pages 131-139.

Laidre, M.E., (2018). Coconut crabs, *Current Biology* 28, R51 – R65.

Lavery, S., Moritz, C., & Fielder, D.R. (1995). Changing patterns of population structure and gene flow at different spatial scales in *Birgus latro* (the coconut crab), *Heredity* 74, 531-541.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 25:1754-60.

Li, H., Handsaker, B., Wysoker, A., Fennell, T. Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools, *Bioinformatics*, Volume 25, Issue 16, Pages 2078–2079.

Losey, J., Harmon, J., Ballantyne, F. A polymorphism maintained by opposite patterns of parasitism and predation. *Nature* 388, 269–272 (1997).

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., & Zdobnov, E.M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, Volume 38, Issue 10, Pages 4647–4654.

Merilaita, S. (2001). Habitat heterogeneity, predation and gene flow: colour polymorphism in the isopod, *Idotea baltica* . *Evolutionary Ecology* 15, 103–116.

Nokelainen, O., Stevens, M., & Caro, T. (2018). Colour polymorphism in the coconut crab (*Birgus latro*). *Evolutionary Ecology* 32, 75–88.

Nokelainen, O, Maynes, R, Mynott, S, Price, N, & Stevens, M., (2019). Improved camouflage through ontogenetic colour change confers reduced detection risk in shore crabs. *Functional Ecology* 33: 654– 669.

Nosil, P. (2005). The role of selection and gene flow in the evolution of sexual isolation in *Timema* walking sticks and other Orthopteroids. *Journal of Orthoptera Research* 14(2): 247-253.

Oka, S., Miyamoto, K., Matsuzaki, S., & Sato, T. (2015). Growth of the coconut crab, *Birgus latro*, at its northernmost range estimated from mark-recapture using individual identification based on carapace grooving patterns. *Zoological Science*, 32: 260-265.

Sabbadin, A., & Graziani, G. (1967). Microgeographical and ecological distribution of colour morphs of *Botryllus schlosseri* (Ascidiacea). *Nature*, 213, 815–816.

Svensson, E.I. (2017), Back to basics: using colour polymorphisms to study evolutionary processes. *Molecular Ecology*, 26: 2204-2211.

Todd, P.A., Wang, W.Y., Huang, H., Belle, C.C., Lim, M.L. and Yeo, D.C., 2011. The function of colourful facial bands in mangrove crab (*Perisesarma*) communication. *Journal of Experimental Marine Biology and Ecology*, *407*(1), pp. 26-33.

Veldsman, W.P., Ma, K.Y., Hui, J.H.L., Chan, T.F., Baeza, J.A., Qin, J., & Chu, K.H. (2021). Comparative genomics of the coconut crab and other decapod crustaceans: exploring the molecular basis of terrestrial adaptation. *BMC Genomics*. 2021 Apr 30;22(1):313.

Wade, N.M., Tollenaere, A., Hall, M.R., & Degnan, B.M. (2009). Evolution of a Novel Carotenoid-Binding Protein Responsible for Crustacean Shell Color, *Molecular Biology and Evolution*, Volume 26, Issue 8, Pages 1851–1864.

Watson-Zink, V.M. (2021). Making the grade: Physiological adaptations to terrestrial environments in decapod crabs. *Arthropod Structure and Development* 64: 101089.

White, T.E., & Kemp, D.J. (2016), Color polymorphism, *Current Biology* 26, R515-R522.

Williams, S.T. (2017). Molluscan shell colour. *Biological Reviews*, 92, 1039–1058.

FIGURES AND TABLES

Fig. 3.1: (A) Global distribution of *Birgus latro*. Each point represents a distinct locality where coconut crabs have been reported (N = 210 localities). Data from Cumberlidge et al. 2022; (B) Map inset showing Tanzanian sampling sites. From north to south: Pemba Island, Misali Island, Kisiwa Panza Island, Chumbe Island, Mbudya Island, and Bongoyo Island; (C) Sampling location (Pink House Research Station) on Christmas Island, Australia where individual was collected for whole genome ultra long-read sequencing; (D) Sampling location (Ocean Expo Park, Churaumi Aquarium) on Okinawa Island, Japan.

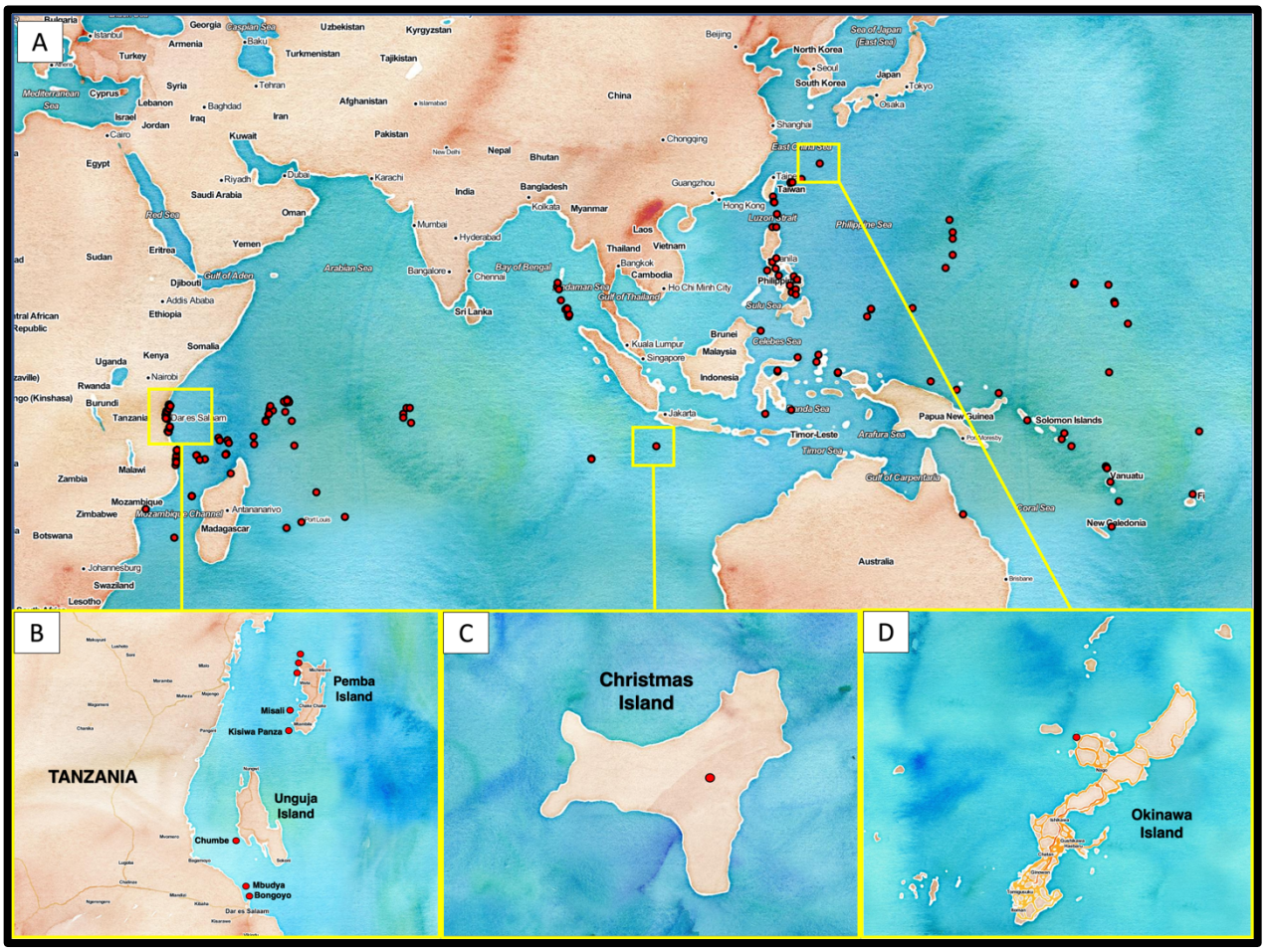


Fig. 3.2: Diversity of dorsum and ventrum colors of coconut crab morphs. For the purposes of this study, the first three crabs from the left are "red" crabs, while the last two are "blue" crabs. The third crab from the left is considered a "sunset" morph. Photo credits: V. Watson-Zink.

Fig. 3.3: Principal component analysis of 2,000 random contigs from across the *B. latro* genome assembly. Island Codes: CHU: Chumbe Island; DSM: Islands near Dar es Salaam, Tanzania (Bongoyo and Mbudya); KIP: Kisiwa Panza; MIS: Misali Island; PEM: Pemba. (A) PCA of samples from Tanzanian islands only; (B) PCA of samples from all islands, key as before but with OKIN: Okinawa, Japan included.

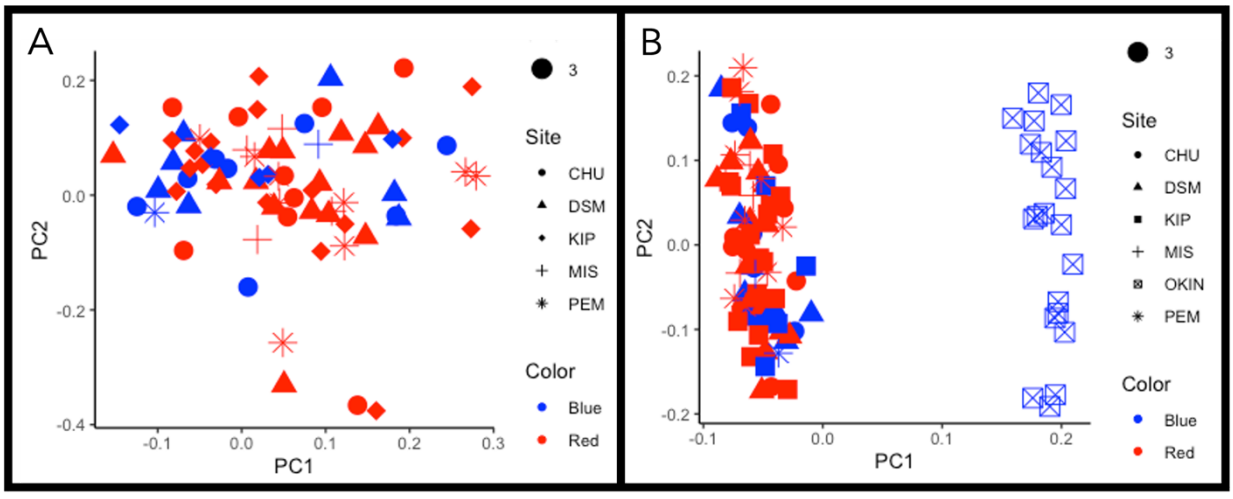


Fig. 3.4: Violin plots of per-site F_{st} values for three examined SNP categories – "random": SNPs randomly selected from contigs not found to possess CRCN sequences; "not in CRCN": SNPs randomly selected from contigs possessing CRCN sequences, but not within the bounds of the suspected CRCN sequences; "in CRCN": SNPs found within the bounds of suspected CRCN genes. Each distribution is plotted for 9,381 SNPs from that respective category.

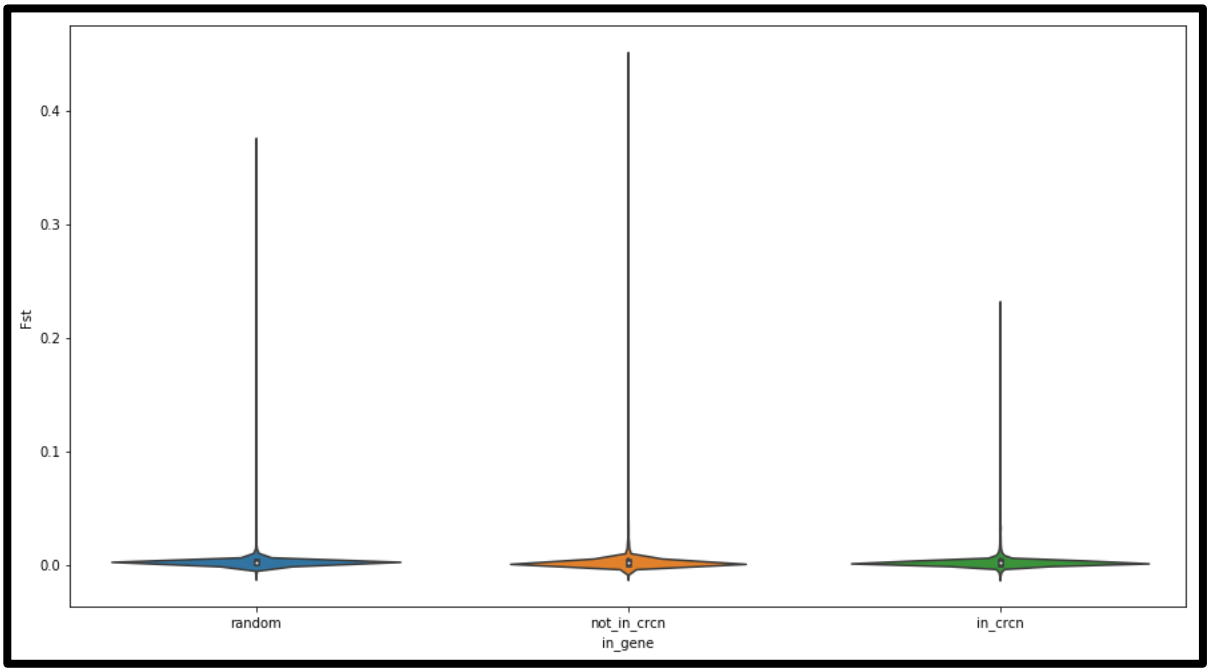


Table 3.1: *de novo B. latro* genome summary statistics.

***B. latro* genome assembly statistics**	
Raw data (Gb)	50.5
Total bases	3,341,376,526
N50 contig length (bp)	188,857
L50 contig length (bp)	4,842
Number of contigs	21,909
Estimated assembly size (Gbp)	3.34