THE EVOLUTION OF BROOD PARASITISM IN BEES AND OTHER ANIMALS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Trevor Joshua Law Sless

May 2022

THE EVOLUTION OF BROOD PARASITISM IN BEES AND OTHER ANIMALS

Trevor Joshua Law Sless, Ph. D.
Cornell University 2022

Parasitism is one of the most widespread strategies among animals, with up to half of all species and hundreds of independent groups exploiting host species in some way. Brood parasitism in particular is unique among other forms of parasitism for its exploitation of the parental care behaviors of host species, though it has been poorly studied in many respects.

In my first chapter, I used literature searches to conduct a review of taxa exhibiting brood parasitism across the animal tree of life and investigated the macroevolutionary patterns shared among these diverse organisms. This research ultimately resulted in the identification of nearly 60 independent origins of brood parasitism, as well as some evidence for reduced diversification in such species compared to their non-parasitic sister groups, though differing evolutionary backgrounds also impact this pattern.

My second chapter investigated the evolutionary relationships and history of the largest and oldest clade of brood parasitic bees, the subfamily Nomadinae. Ancestral reconstructions of host-parasite associations and parasitic strategies indicate that the early ancestors of this group used closely related bees as hosts, but that later

behavioral innovations likely facilitated the expansion of their host range by allowing the exploitation of a wider variety of distantly related species.

Finally, the third chapter of my dissertation focused on the evolutionary signatures of brood parasitism detectable within the genome of a single bee species, *Holcopasites calliopsidis*. In agreement with previous findings on the genomic evolution of other parasitic organisms, this species shows reduced genome size compared to non-parasitic relatives, but no net loss of either repetitive or genic content. The genome assembly also shows signatures of dynamic evolution in comparison with other bee species, including the inferred presence of novel and rapidly evolving transposable elements.

BIOGRAPHICAL SKETCH

Trevor Sless was born in Toronto, Canada, where he graduated from Northern Secondary School in 2011. He proceeded to attend the University of Toronto from 2011-2015, where he completed his Honours Bachelor of Science with double majors in Ecology and Evolutionary Biology as well as Molecular Genetics and Microbiology. He then entered the Ecology and Evolutionary Biology department at Cornell as a graduate student in August 2015.

Dedicated to my friends and family, near and far, for being good company through highs and lows, and for making me the person I am

ACKNOWLEDGMENTS

I'd like to thank the various funding sources which supported me over the course of my doctoral degree, including Cornell's Presidential Life Sciences Fellowship and an NSERC Postgraduate Scholarship-Doctoral. My research was also supported by a variety of small grants including EEB departmental funding, an Andrew W. Mellon research grant, and multiple grants in aid of research through the Cornell chapter of Sigma Xi. Some of my research was also funded through a U.S. National Science Foundation grant (DEB-1555905) awarded to B.N. Danforth, S.G. Brady, J.P. Pitts, and R. Ross.

I am also thankful to various collaborators I worked with over the course of my PhD. Michael Branstetter of the USDA Pollinator Research Group in particular was instrumental in generating and analyzing the phylogenomic data which ultimately became my second chapter. I also thank my other co-authors, including Jessica Gillung, Erin Krichilsky, Kerrigan Tobin, Jakub Straka, Jerry Rozen Jr., Felipe Freitas, Aline Martins, and Silas Bossert.

There are many members of the Cornell community who I greatly appreciate for their help or inspiration in my journey over the past six and a half years. Firstly though, I acknowledge that Cornell University is located on the unceded traditional lands of the Gayogo̱hó꞉nǫ̀ʾ people within the Haudenosaunee Confederacy. I'm grateful to the EEB department's administrative staff who have worked to keep things running smoothly, and of course far too many graduate students to list individually who have been wonderful friends and

colleagues. I also want to thank faculty members who I worked with or got advice from in various capacities, particularly Mariana Wolfner, Dan Barbash, Nina Therkildsen, Corrie Moreau, and Michelle Smith. My committee members, Matt Hare and Amy McCune, have both been incredibly helpful in bringing their respective knowledge to my work and providing feedback on the many research ideas I've explored.

I owe huge thanks to my co-advisors, Bryan Danforth and Jeremy Searle, for supporting me throughout my time in their labs. Bryan's endless enthusiasm and knowledge on all things bee-related has been absolutely invaluable in identifying and steering potential projects, and I've learned so much from him and other members of the Danforth lab. And Jeremy's willingness to help me navigate interesting (though sometimes esoteric) evolutionary questions while constantly being a supportive, patient, and understanding advisor has also been instrumental in completing my degree.

I also wish to express my gratitude to my initial PhD advisor at Cornell, Rick Harrison, who passed away suddenly in my first year here. Though my time with Rick felt far too short, he had an important formative influence on me during my early time as a graduate student, and I hope that I have been able to keep his academic spirit as a part of my work even as I've moved on to other research topics.

Finally, I want to express my deepest thanks to my family, particularly my mother and sister, as well as the many friends outside of Cornell who have been endless sources of comfort and support.

April 2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

EVOLUTIONARY ORIGINS AND PATTERNS OF DIVERSIFICATION IN ANIMAL BROOD PARASITISM

Abstract

Brood parasitism involves the exploitation of host parental care rather than the extraction of resources directly from hosts. Although often excluded, we show that brood parasitism fits within existing frameworks of parasitic strategies, and formulate defining criteria, including that the parasites be obligate and the hosts non-eusocial (thereby distinguishing them from social parasites). A systematic literature survey revealed 59 independently derived brood parasitic lineages with most origins (49) in insects, particularly among bees and wasps, and other origins in birds (7) and fish (3). Insects account for over 98% of brood parasitic species, with much of that diversity reflecting ancient (100+ million-year-old) brood parasitic lineages. Brood parasites usually, but not always, evolve from forms that show parental care. In insects, brood parasitism often first evolves through exploitation of a closely related species, following Emery's rule, but this is less typical in birds, which we discuss. We conducted lineage-level comparisons between brood parasitic clades and their sister groups, finding mixed results but an overall neutral to negative effect of brood parasitism on species richness and diversification. Our review of brood parasites reveals many unanswered questions requiring new research including further modelling of the co-evolutionary dynamics of brood parasites and their hosts.

Introduction

The parasitic exploitation of one animal species by another is a pervasive strategy - it has been suggested that parasites make up as many as half of all animal species globally (Windsor, 1998). Parasitism has evolved on numerous occasions in almost every major group of animals, and so it should not be surprising that the array of modes and strategies employed by parasites is just as diverse as those exhibited in non-parasitic animals. From the minuscule, sometimes even intracellular Myxozoa to thirty-meter tapeworms in the guts of whales, and from the occasional usurper of a mutualistic relationship to species that inevitably kill their host, parasites run the gamut of ways to live in, on, and around other animals. Though such a diverse set of organisms defies simple categorization, several attempts have been made to review the phylogenetic distribution of parasitic animals and to divide parasites into overarching categories based on their methods of action and consequences for host organisms.

Brood parasitism, the strategy of exploiting another species' parental care resources to raise a parasite's own offspring, has been a subject of interest studied across a wide variety of taxa, and yet it has rarely been included within the broader literature of parasitism as a whole. For example, brood parasitism is excluded from Weinstein and Kuris' (2016) report of over 200 independent origins of parasitism across 15 animal phyla and is not mentioned at all in Poulin and Randhawa's (2015) review of parasitic strategies. The most salient difference between brood parasitism and other forms, which may

account for its frequent exclusion from such discussions, is the nature of the exploited resources. While nearly all other forms of parasitism work to extract nutrients, energy, and sometimes protection directly from the body of the host (termed “trophic parasites” by Pollock et al. 2021), brood parasitism instead usurps efforts made by the host - most commonly food provisioning intended for their young. In effect, brood parasitism acts upon a host’s extended phenotype (sensu Dawkins, 1982).

Though it has evolved on multiple occasions in animals, brood parasitism cannot truly be described as equally common across all branches of the tree of life. There are clear taxonomic biases as to which clades show greater or lesser tendencies in evolving brood parasitic strategies, in no small part due to the fairly specific requirement of a suitable host species with at least rudimentary parental care. While most brood parasitic lineages are derived from ancestors with parental care themselves, having effectively discovered a way to offload the costs of this behavior, others appear to have derived this strategy by other means. Despite a long history of study for some of the older and more diverse parasitic lineages, however, there is surprisingly little cross-communication in brood parasite research among specialists on different taxa, as noted in the recent “call for integration” from Thorogood et al. (2019). Although no brood parasitic species are of particular medical, agricultural, or economic importance, such organisms play important ecological roles in many ecosystems as natural enemies of their hosts (Pollock et al. 2021) and

much needs to be learned about the evolution of brood parasitic lineages. To this end, we here aim to review the independent origins of brood parasitism across the phylogeny of animals and discuss the evolutionary patterns and consequences of this strategy.

Defining Brood Parasitism Among Other Forms of Exploitation

For purposes of comparison, it is important to have a clear definition of what exactly brood parasitism entails. Within the context of this review, we have applied the following criteria:

Exploitation of parental care

This is the core feature of brood parasitism. Resources (often, but not exclusively, food provisions) which would otherwise benefit the offspring of a host organism are instead co-opted by a brood parasite's young. Based on this criterion, "trophic" parasites, parasitoids, and micropredators which feed directly on a host organism are not brood parasites.

Developmental synchrony

The equivalence of life stages between brood parasite and host offspring is another important feature of brood parasitism. Though the actual rate of development may vary between the two, brood parasites specifically exploit their hosts in the early stages of life, effectively replacing host offspring in the process. Other forms of resource theft by adults (sometimes termed "kleptoparasitism") therefore are not brood parasitism.

Obligate parasitism

While facultative brood parasitism (often between members of the same species) has been documented in a wide range of animals and may serve as a steppingstone in the transition to an obligate relationship, the prevalence of facultative brood parasitism is difficult to ascertain. Additionally, especially in cases of intraspecific parasitism, facultative parasites would not be expected to be measurable as cohesive lineages which could be used for diversification analyses in the same way as larger, obligately brood parasitic clades. We therefore restrict our definition to obligate brood parasites.

Non-eusocial hosts

Although brood parasitism has much in common with the exploitation of social insect colonies exhibited by various taxa, the appropriation of resources from a worker caste is distinguished from parental care in the traditional sense. As a result, social parasites, inquilines, and other categories of organisms which live among colonial insects are not considered brood parasites by our definition.

These criteria for defining brood parasitism are by design restrictive, with the goal of designating a guild of organisms with convergent biological features in spite of very distant evolutionary relationships. Our categorization is generally similar to the "parental-care parasitism" described by Roldán and Soler (2011), with the main difference being their inclusion of facultative and intraspecific brood

parasites. Similarly, our criteria overlap with those considered by Pollock et al. (2021) in excluding "trophic" parasites and requiring obligate parasitism, but ours are narrowed by the exclusion of organisms which exploit eusocial insects; Pollock et al. in fact only consider social parasites among insects to the exclusion of several brood parasitic groups with solitary hosts. Though the biology and evolutionary trends among such species overlap somewhat with those of brood parasites as we define them, the distinction between brood parasitism and social parasitism maintained herein is justified in part by the latter's general treatment as a distinct phenomenon with its own body of literature (reviewed by Rabeling et al. 2020).

The broader parasitism literature includes several attempts to classify or group all parasites based on various details of their biology and life cycles. Anderson and May (1979) first recognized a distinction between "micro-" and "macroparasites", largely based on the later category's larger body size and slower generation time closer to that of the host. Kuris and Lafferty (2000, and subsequently Lafferty and Kuris 2002) create further dichotomies based on whether or not host fitness is reduced to zero, the number of host individuals used over a parasite's life, and whether host-parasite interactions are intensity-dependent or not (see below). Poulin (2011) additionally recognizes divisions based on the mode by which parasites encounter or spread between hosts, defining six "universal" strategies: parasitoids, parasitic castrators, directly transmitted parasites, trophically transmitted parasites, vector-transmitted parasites, and

micropredators. Poulin and Randhawa (2015) further suggest that these six categories represent the “only adaptive peaks in the parasite evolutionary landscape”.

Having now formalized a definition of brood parasitism based on our criteria above, this strategy can be assessed by all of the same dichotomies that are used for these other classes of parasites. However, it is important to recognize that some of these vary depending on the generational perspective taken. In the case of brood parasitism, are adults or offspring best considered the hosts, and likewise is it the adults or offspring that are best considered the parasites? We argue that it is most consistent to consider the interaction between adult brood parasites and adult hosts, despite the fact that the perspective of offspring may result in different implications.

With this in mind, all brood parasites clearly fall under Anderson and May’s definition of macroparasites since generation times between host and parasite are necessarily similar. Under the criteria of Lafferty and Kuris (2002), the question of perspective becomes more evident. An adult brood parasite may attack one or multiple host individuals or nests and may or may not reduce host fitness to zero depending on the number of surviving host offspring. To the extent that the estimation of intensity of “infection” makes sense, brood parasitism should also be intensity-dependent, since multiple “infections” by different brood parasites could reduce host fitness

further, though in reality this is fairly uncommon. In this sense, adult brood parasites overlap with micropredators, social predators, and/or certain types of parasitic castrator within Lafferty and Kuris' framework.

Following the six-strategy system of Poulin and Randhawa (2015), brood parasites may affect one or more host individuals, typically but not always from the same species. Their effective virulence ranges from high to maximum due to the death of some or all host offspring (whether directly or indirectly as a result of the brood parasite's actions). Manipulation of host behavior occurs in some cases, and the size ratio of parasite:host is typically near 1:1. Prevalence of brood parasitic species may be highly variable, but their mean intensity of infection is low, close to or rarely even below one parasite per host.

As this assessment hopefully makes clear, brood parasitism can indeed be placed within existing frameworks of parasitic strategies and compared against other recognized categories. Furthermore, brood parasitism occupies a discrete and distinctive position among parasites as a whole based on the particular combination of life history traits discussed above, and should be included in future analyses and discussions on the nature of parasitism.

Methodology

For our review of brood parasitic taxa, independent origins of brood parasitism were identified through a combination of primary and secondary literature sources. Initial efforts included directed literature searches for the terms “brood parasitism” and “kleptoparasitism/cleptoparasitism” combined with the names of different animal taxa of high rank. However, many brood parasitic clades were ultimately identified through expert knowledge via more general and often taxon-specific secondary sources (e.g. Michener 2007; Roldán and Soler 2011; Mann 2017). Candidate brood parasites were further investigated and evaluated against the criteria discussed above for inclusion, resulting in a final list of 59 brood parasitic clades (Appendix I Table 1.1; see also Appendix I Table 1.2 for examples of excluded taxa).

Further data collection was then carried out on each of the taxa under consideration. Sister groups were identified by searching for phylogenetic studies including the relevant species, and in the event of conflicting topologies, preference was given to more recent phylogenies with more focused taxon sampling and/or larger datasets (Appendix I Table 1.3). Species richness estimates for both brood parasitic clades and their sister groups were obtained, depending on taxa, from sources ranging from larger databases to individual primary research publications (Appendix I Tables 1.1, 1.3). Wherever possible, estimates for parasites and their sister groups were obtained from the same source to minimize potential bias. Divergence times between parasitic clades and their sister groups, as well as crown ages for both when

obtainable, were sourced from previously examined dated phylogenies as well as other studies identified through TimeTree.org (Kumar et al. 2017); if multiple estimates were available, dates were typically averaged (Appendix I Table 1.4).

To help account for uncertainty and possible errors in these data, multiple taxon sets were used which included or excluded different sets of taxa (Appendix I Table 1.5). Specifically, species richness analyses (Appendix I Table 1.6) were conducted on the "SR-Full" set of brood parasitic taxa with identified sister groups (n=46); a "non-monotypic" set (henceforth "SR-NM", n=39) excluding any taxon pairs where either the brood parasitic clade or its sister group consisted of a single species; a "non-problematic" set ("SR-NP", n=31) excluding brood parasites which had competing hypotheses for the sister group's identity and/or uncertainty in species counts of either clade; and a "non-monotypic, non-problematic" set ("SR-NMNP", n=26) applying both criteria. In addition, to investigate possible differences between brood parasites clades from different evolutionary backgrounds, the SR-Full dataset was partitioned into two additional subsets of brood parasitic lineages inferred to have derived from ancestors with parental care ("SR-PC", n=34) and those which transitioned to brood parasitism from other ancestral conditions not including parental care ("SR-NPC", n=12).

Diversification rate comparisons (Appendix I Table 1.7) were conducted on a different "DR-Full" dataset of brood parasite/sister

group pairs with available divergence times (n=42), as well as a separate “non-problematic” set (“DR-NP”, n=23) excluding pairs with only rough age estimates, and a “crown age” dataset (“DR-CA”, n=23) using estimated crown ages for brood parasite/sister group pairs which had these available as opposed to divergence times. As with species richness comparisons, the DR-Full dataset was again partitioned between “parental care-derived” brood parasitic lineages (“DR-PC”, n=32) and those from other ancestral backgrounds (“DR-NPC”, n=10).

Pairwise analyses of species richness and diversification rate comparing brood parasitic clades with their sister groups were conducted in R using linear regressions as well as the “diversity.contrast.test”, “mcconwaysims.test”, and “richness.yule.test” functions available through the “ape” package (Paradis and Schliep 2019). The “sister clade comparison” method of Käfer and Mousset (2014) was also carried out with the “scc_test” R script provided by these authors. Additionally, the packages ape and phytools (Revell 2012) were used to generate a schematic phylogeny of all brood parasitic clades and their sister groups using approximate divergence times along with date estimates for deeper clades obtained from previously discussed sources and/or TimeTree.org (Kumar et al. 2017).

Broad Phylogenetic Patterns

Table 1.1 summarizes the distribution of the fifty-nine independent origins of brood parasitism across various major (not

necessarily monophyletic) groups of animals. Among insects, brood parasitism is best represented in the order Hymenoptera, which includes twenty origins within bees and twelve more in predatory and parasitic wasps. There are also nine origins in flies (Diptera) and eight in beetles (Coleoptera). These insect groups account for over 98% of described species of brood parasite (Figure 1.1). In vertebrates, brood parasitism is both most common and also most rigorously studied in birds, where it has evolved seven times. However, this strategy has also been recorded in at least three species of fish and may well be present in others that have yet to be studied in detail. Figure 1.1 shows the schematic phylogeny of all brood parasitic clades with their respective sister groups.

Table 1.1: Major animal groups exhibiting brood parasitism

	# Origins of Brood Parasitism	**# Brood Parasitic Species**	**Total Extant Species**	**% Brood Parasitic**
Fish (Actinopterygii)	3	4	34,650[1]	0.012
Birds (Aves)	7	105	10,824[2]	0.970
Beetles (Coleoptera)	8	2,183	392,415[3]	0.556
Flies (Diptera)	9	651	160,591[3]	0.405
Wasps* (Hymenoptera)	12	3,286	112,569[4]	2.919
Bees (Anthophila)	20	2,769	19,844[4]	13.95
Total	59	9,024	1,525,728[5]	0.590

Sources. - 1. Fricke et al. 2021; 2. Billerman et al. 2020; 3. Zhang 2013b; 4. Aguiar et al. 2013; 5. Zhang 2013a.
Note. - * Wasps are here defined as the paraphyletic group consisting of the hymenopteran clade Apocrita with the exclusion of bees (Anthophila) and ants (Formicidae).

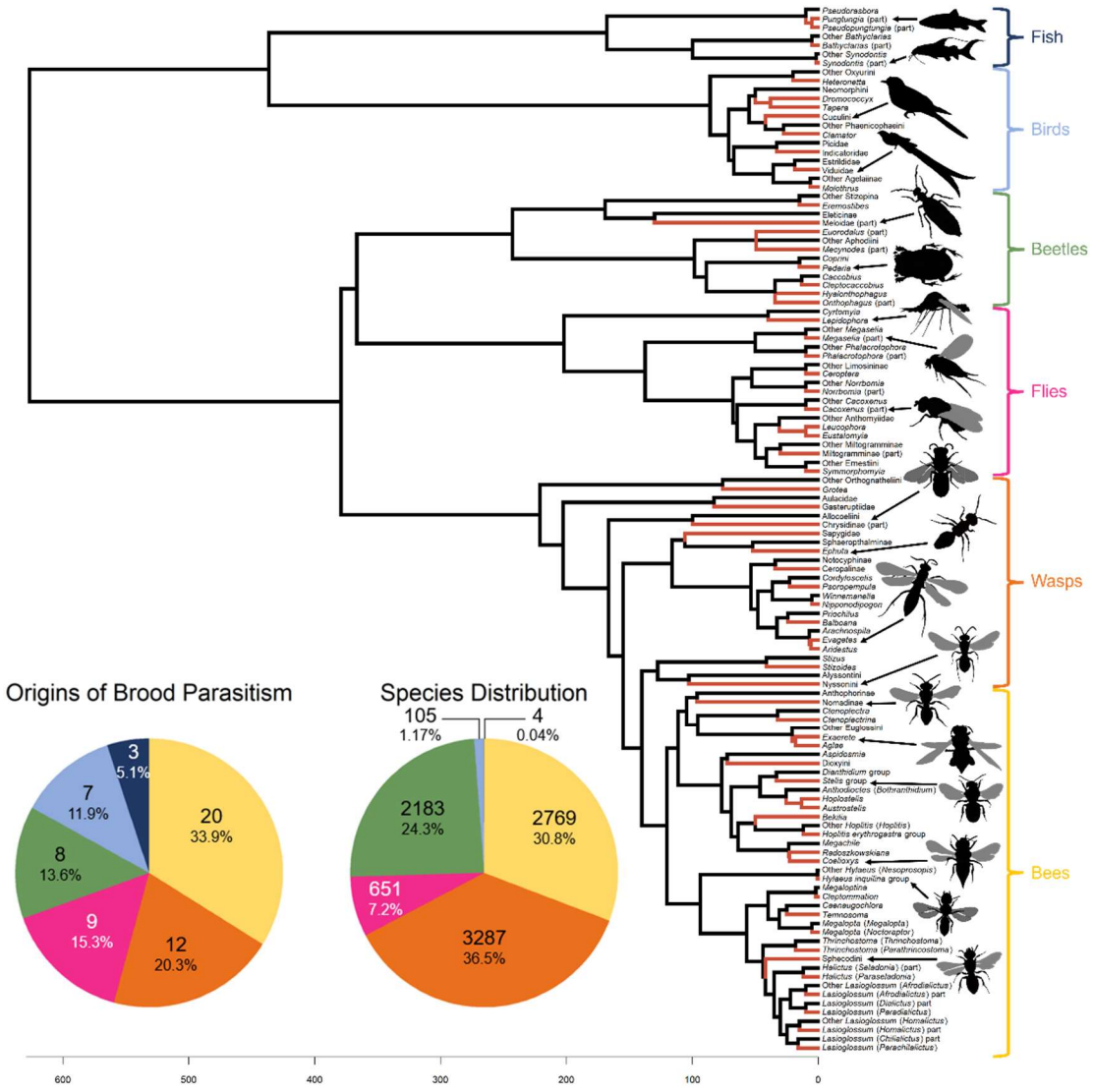


Figure 1.1: Phylogeny of brood parasitic lineages (red) and their respective sister clades. Inset pie charts show number of independent origins of brood parasitism and species diversity estimates broken down by labeled taxonomic groups. Scale bar shows millions of years before present. Organism silhouettes acquired from phylopic.org and other sources, reproduced under CC-BY 3.0 (attributions: Dick Belgers, Melissa Broussard, Patrick Coin, Øivind Gammelmo, Cody Hough, Maija Karala, Udo Schmidt, Arnstein Staverløkk)

Brood parasitic clades show a wide range of variation in their ages of origin. There are multiple such groups within insects dating back over 100 million years, though the fact that these age estimates are divergence times from non-parasitic sister groups rather than crown ages makes it difficult to tell exactly when the transition to this strategy first occurred. On the other extreme, brood parasitic masked bees in the genus *Hylaeus* are necessarily less than one million years old, being part of a recent Hawaiian radiation (Magnacca and Danforth 2006). Though it is difficult to compare with similar datasets for other types of parasites, the general pattern of age distributions in brood parasitic clades does match some very broad features of Weinstein and Kuris' (2016) review of animal parasitism. Their study reports that over 90% of extant parasites are found in a small number of ancient species-rich clades with origins in the Paleozoic, as opposed to a much larger number of more recent, less speciose groups of parasites. Similarly, the six oldest clades of brood parasites (blister beetles [Meloidae], sapygid wasps, cuckoo wasps [Chrysidinae], sand wasps [Nyssonini], nomadine bees, and gasteruptiid wasps) together comprise a clear majority of all living brood parasitic species at over 70% (6,425/8,999).

Knowledge of the diversity of brood parasites is necessarily limited by the poorly understood biology of some groups. In taxa where brood parasitism is relatively common, such as bees, potential parasites may be recognized even in the absence of behavioral data by suites of morphological characteristics known from other parasitic

taxa (e.g., loss of pollen-collecting structures, defensive adaptations against hosts; Michener 2007). The relative ease of identification of putative brood parasites in such cases may partially account for the disproportionately large number of independent origins in bees. Or alternatively, this may hint at the existence of many more brood parasites among the numerous insects and other invertebrates whose life histories are entirely unknown, including undescribed species (Stork 2018). Similarly, it would be surprising if the three unrelated species of fish considered herein are the only brood parasites in such a taxonomically diverse group. Even among groups with clearly known brood parasitism, it is sometimes challenging to ascertain the true number of unique origins and brood parasitic species - especially when further complicated by reversals away from brood parasitism as appear to occur in both blister beetles and cuckoo wasps.

Patterns of Species Richness and Diversification

The relatively large number of parallel origins of brood parasitism presents the opportunity to examine potential effects of this strategy on lineage diversity. To investigate this, pairwise species richness analyses between brood parasitic clades and their sister groups were conducted using a variety of methods for each of the six taxon sets generated ("SR-Full", "SR-NM", "SR-NP", "SR-NMNP", "SR-PC", and "SR-NPC"; see methodology).

Comparison of log-normalized species counts between brood parasites and sister groups (Figure 1.2) resulted in regression models

with slopes significantly less than 1 for all taxon sets except the "non-parental care-derived" brood parasites ($p = 0.00132$, 0.00375, 0.00122, 0.00926, 0.00705, and 0.0881 respectively), implying an overall pattern of lower species richness in brood parasitic groups. This was further supported by diversity contrast tests through R's ape package (see Appendix I Table 1.8 for full statistical results). The "proportion" method based on Barraclough et al. (1995) indicated significantly lower species richness in brood parasitic clades ($p < 0.05$) for all taxon sets except SR-NPC, while the "ratiolog" method (Barraclough et al. 1996) detected significant differences in all but SR-NPC and SR-NP. The "difference" method (Sargent 2004) and McConway-Sims test (McConway and Sims 2004) only found significantly reduced species richness for brood parasites in the SR-Full and SR-PC taxon sets. In contrast, the "sister clade comparison" (SCC) method of Käfer and Mousset (2014), which attempts to account for the asymmetry induced by the waiting time until a novel trait (i.e. brood parasitism) appears, found no significant differences between the species richness of brood parasitic clades and their sister groups in any taxon set except for the "non-parental care-derived" SR-NPC set, which actually showed significantly increased species richness in brood parasitic lineages vs. their sister groups.

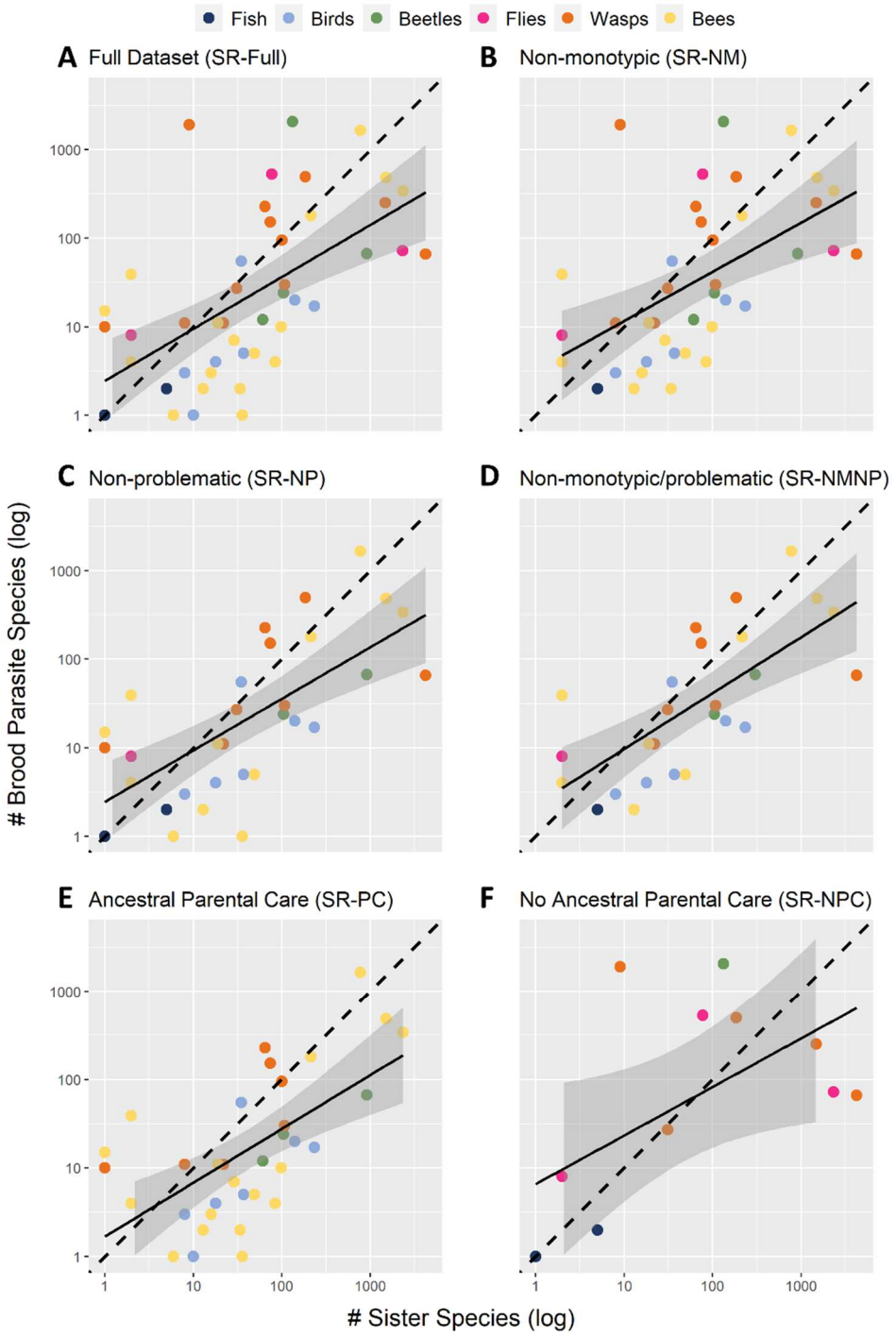


Figure 1.2: Log-log species richness comparisons of brood parasitic clades (Y-axis) and their sister groups (X-axis). Solid lines show linear regression with 95% confidence interval, while dotted lines represents a 1:1 relationship. Brood parasite/sister group pairs are color-coded according to major taxonomic categories: fish = dark blue, birds = light blue, beetles = green, flies = pink, wasps = orange, bees = yellow. Panels show different subsets of data (see methodology): A) SR-Full dataset of 46 taxon pairs, B) SR-NM (n=39), C) SR-NP (n=31), D) SR-NMNP (n=26), E) SR-PC (n=34), F) SR-NPC (n=12).

When incorporating age estimates to more directly measure diversification rate as opposed to simple species richness, results were generally similar (Figure 1.3). Paired t-tests of log-transformed diversification rates in species/million years between brood parasitic clades and their sister groups showed significantly lower rates for brood parasites in the DR-Full divergence time dataset ($p = 0.03429$), but not for the "DR-NP" divergence time dataset ($p = 0.2213$) or the "DR-CA" dataset ($p = 0.1058$). The taxon subset of "parental care-derived" brood parasites (DR-PC) also showed significantly reduced diversification rates in comparison to their sister groups under this method ($p = 0.001414$), while "non-parental care-derived" brood parasites (DR-NPC) did not ($p = 0.1244$). However, the diversification shift test based on the Yule process (Paradis 2012) found significant reductions in diversification rate for all tested datasets (DR-Full $p = 1.022918e-12$, DR-NP $p = 1.479713e-07$, DR-PC $p = 2.400846e-09$, and DR-NPC $p = 5.36107e-05$; the DR-CA dataset is not compatible with this method).

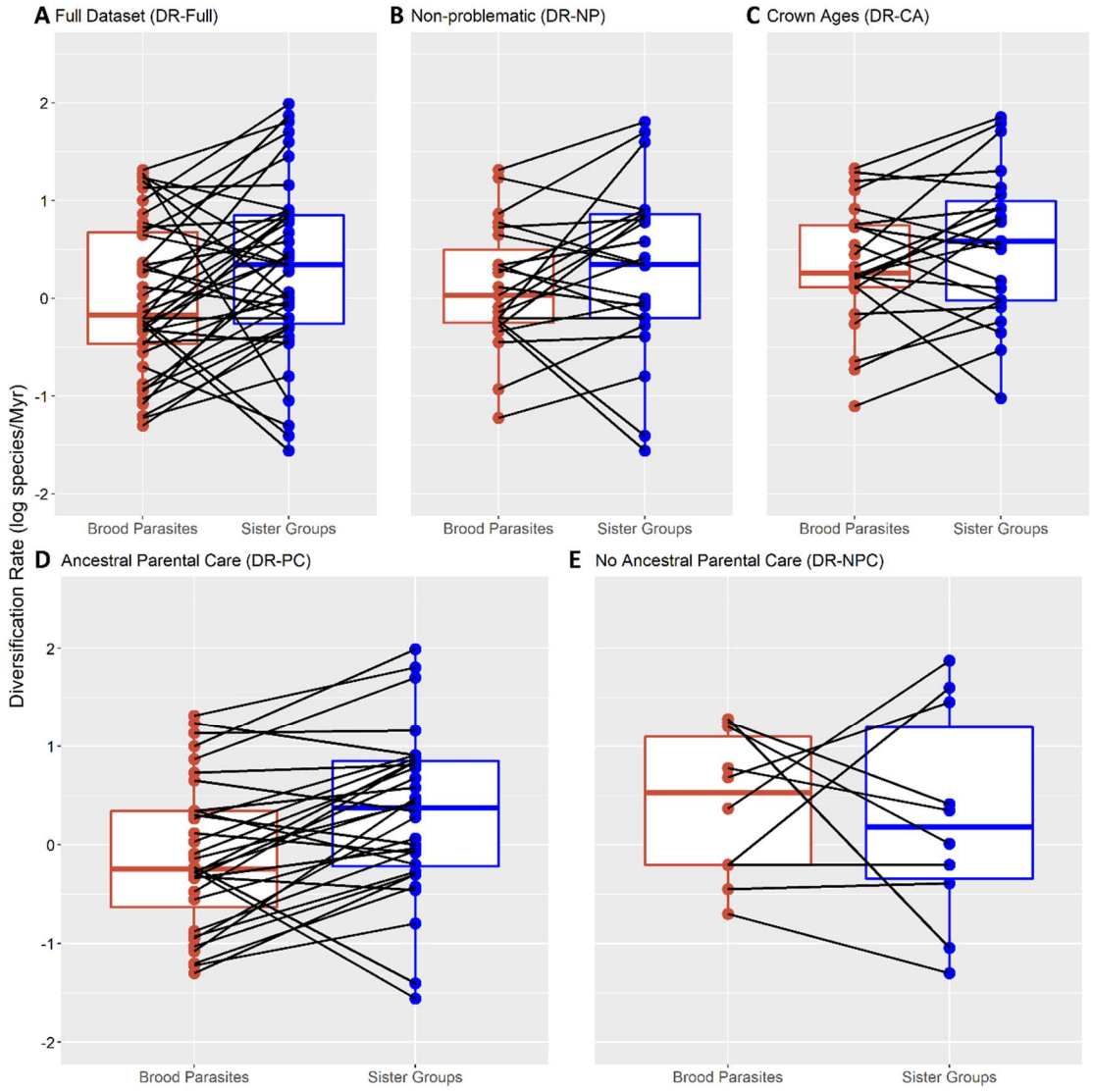


Figure 1.3: Comparison of diversification rates for brood parasitic clades (red) and their sister groups (blue), measured in species per million years. Individual brood parasite/sister clade pairs are connected by black lines. Panels show different subsets of data (see methodology): A) DR-Full dataset of 42 taxon pairs, B) DR-NP (n=23), C) DR-CA (n=23), D) DR-PC (n=32), E) DR-NPC (n=10).

The overall picture provided by these analyses indicates a negative or at least clearly non-positive effect of a brood parasitic life history on lineage diversification and extant species richness, which is generally consistent with previous literature on specific taxa. Krüger et al. (2009), studying cuckoo birds (including three distinct brood parasitic lineages: Cuculini, *Clamator*, and *Dromococcyx* + *Tapera*), reported approximately doubled speciation rates for brood parasites compared to their non-parasitic sister lineages, but quadrupled extinction rates, leading to an overall decrease in net speciation. Medina and Langmore (2015) found increased speciation and extinction rates for brood parasitic indigobirds (Viduidae) vs. their sister lineage under one model, but no significant differences in either rate for cuckoos or honeyguides (Indicatoridae). Among insects, Litman et al. (2013) reported a reduction in diversification rates for eight out of ten brood parasitic lineages of bees. Comparing more broadly across all forms of parasitism presents conflicting results. Jezkova and Wiens' (2017) analysis of animal phyla recovered parasitism as a significant (positive) explanatory factor for diversification rate in their phylogenetic multiple regression. However, Weinstein and Kuris (2016), using a similar sister clade analysis to those presented herein, did not recover any significant differences in species richness between parasitic clades and their sister groups.

The tendency for species richness and diversification rates to be reduced in brood parasites in contrast to their non-parasitic sister groups is an intriguing indication of the effects of this repeated life

history transition, though the relative importance of speciation and extinction cannot be resolved with these methods. More complete diversification rate analyses are unfortunately challenging due to the vast phylogenetic distances between brood parasitic taxa and lack of available phylogenies with sufficient taxonomic scope to be of use for tree-based diversification rate metrics (an obstacle also discussed by Hay et al. 2020). However, the apparent dichotomy in the relative evolutionary trajectories of brood parasites derived from ancestors with parental care and those that have transitioned to this strategy from other behaviors is an interesting finding. The reduction in diversification rates seen in "parental care-derived" brood parasites is consistent with the results discussed above, since all birds and bees fall into this category, but the absence of this effect in "non-parental care-derived brood parasites" - or indeed, even the inverse with apparently increased species richness based on the SCC test - is puzzling.

One potential explanation is that the evolution of parental care itself may be a driver of species diversity, with lineages that abandon this strategy in favor of brood parasitism correspondingly experiencing reduced diversification relative to their sister groups, while other lineages that transition to brood parasitism from ancestral backgrounds of parasitoidy or saprophagy experience no such effect. This difference could also be partly driven by several very recent origins of brood parasitism (represented by just one or a small number of known species within otherwise non-brood parasitic genera) falling

into this latter group, which were excluded from sister group analyses due to insufficient data. The fact that brood parasitic lineages derived from groups with ancestral care apparently never experience reversals back to this state, and thus may end up "trapped" in this strategy while some brood parasites clades from other backgrounds do show evidence of reversal (see below) may also be relevant. The general rarity of reversals among brood parasitic clades may also hint at the existence of a "ratchet" effect in the evolution of this strategy, which has been explored to some extent in birds (e.g. Takasu 1998) but remains uninvestigated among other taxa. More work is needed to further understand the diversification trends uncovered here.

Converging Pathways to Brood Parasitism

The clearly non-random distribution of brood parasites across the animal tree of life raises the question of what prerequisite conditions allow this strategy to become viable. About two-thirds of the independent origins of brood parasitism discussed herein (40/59) occur from an evolutionary background of parental care in some form. Though the specifics may range from simple mass provisioning as in bees, wasps, and dung beetles to active feeding in birds, these brood parasites likely began as species with parental care that opportunistically began parasitizing similar species. The potential for facultative, intraspecific brood parasitism to serve as a stepping stone to fully obligate brood parasitism has long been suggested (e.g., Hamilton and Orians 1965). Opportunistic nest usurpation within a species has been documented in hundreds of birds (Yom-Tov 2001;

Yom-Tov and Geffen 2017) as well as insects (Field 1992; Černá et al. 2013), and previous work has recognized parallels between these groups (Brockmann 1993). The potential for such a situation to result in the establishment of true brood parasitic lineages through sympatric speciation has been supported by mathematical modelling studies (Cichoń 1996; Robert and Sorci 2001), though not universally (Yom-Tov and Geffen 2006).

A logical consequence of an intraspecific pathway to brood parasitism is that brood parasitic lineages derived from an ancestral background of parental care may be expected to be closely related to their hosts. Among insects, this tendency has been noted for brood and socially parasitic species, where it is known as Emery's rule (Emery 1909). This phenomenon is sometimes divided into a loose form, in which parasites and hosts are phylogenetically close in a general sense, and a strict form with hosts and parasites being immediate sister taxa. Emery's rule has most commonly been examined within socially parasitic insect clades, where it is typically not supported in the strict form when considering individual parasite species (e.g. Huang and Dornhaus 2008; Lopez-Osario et al. 2015). However, when applied at the level of parasitic clades - which necessarily must have originated from a single, ancestral parasitic species - and their closest non-parasitic relatives, Emery's rule may provide useful insights into the origins of such strategies. The phylogenetic signal of Emery's rule during the initial transition to parasitism may persist even in ancient clades, despite subsequent

host switching, as in the cuckoo bee subfamily Nomadinae (Sless et al. 2022, or Chapter 2 herein). And indeed, Emery's rule in the loose if not strict sense appears to hold true for nearly every clade of insect brood parasites considered herein with an ancestral background of parental care (Appendix I Table 1.9), with the sand wasp *Stizoides* as the only exception (though in some groups the identity of sister clades and/or hosts is too poorly known to make any conclusions). In birds, close relationships with hosts are less commonly evident, though some clades do follow this pattern (e.g., indigobirds and wydahs exploit hosts in their sister family Estrildidae, and some species of cowbirds specialize on other members of Icteridae; Mann 2017). It is interesting to consider why close relationships between parasites and hosts are less apparent in birds than in insects. Brood parasitic lineages of birds do not differ significantly in age from those of insects on average (though they do span a narrower range of ages), precluding any explanations based on these groups being relatively younger or older. The better fit of Emery's rule in insects may reflect hard-wired behavioral strategies with a genetic change promoting intraspecific parasitism and a rather small further genetic change enabling parasitism of a close relative that provisions young in a similar way or with similar resources. In birds, greater behavioral flexibility may permit natural selection to favor brood parasitism of the most easily exploited host among a whole suite of species, with the chosen host not necessarily a close relative. This difference between birds and insects is ripe for further analysis with scenario testing using mathematical models.

Conversely, about a third of the brood parasitic clades considered herein did not arise from an evolutionary background of parental care, but rather transitioned from other ways of life. This can be seen in several groups of wasps which have switched from direct parasitoidy on other wasps or bees to parasitoidy of immobilized insects used as food stores created by the host for its larvae. The relative ease of this transition may explain the existence of reversals away from brood parasitism in cuckoo wasps (Pauli et al. 2019), contrasting with the complete absence of reversal in brood parasites derived from ancestors with parental care. However, in at least two cases (the ichneumonid wasp genus *Grotea* and some Gasteruptiidae), the shift to brood parasitism has been accompanied by a more drastic dietary shift from insects as food to the utilization of plant material (namely pollen) in exploited larval food supplies of bees. Different again, flies, blister beetles and the darkling beetle genus *Eremostibes* appear to have arrived at brood parasitism from saprophagy or scavenging. It seems likely that such associations initially began as opportunistic exploitation of food provisions gathered by hosts before their transition to a more obligate strategy. The tachinid fly *Symmorphomyia* is an interesting exception among dipteran brood parasites, however, apparently having evolved from parasitoidy of individual chrysomelid beetle larvae to clearing out entire nests of host wasps that use multiple beetles as food provisions (Hamanishi 1996). Finally, all three brood parasitic groups of fish appear to exploit the parental care of other species without any evolutionary history of such

behaviors themselves. These strategies probably arose from accidental mixing of eggs with those of brood-guarding or mouthbrooding hosts which provided advantages over simple mass spawning (Yamane et al. 2009; Polačik et al. 2019). By necessity, the groups discussed above parasitize unrelated organisms and do not follow Emery's rule, since their sister clades have no exploitable forms of parental care.

Thus, while parental care within a lineage is not *per se* a prerequisite for brood parasitism to arise, it does seem to substantially increase the likelihood of this strategy evolving. Following from this observation, the conspicuous absence of brood parasites from some other groups of animals with long histories of parental care is a point of interest. Parental care is obviously ubiquitous in mammals, and while intraspecific allonursing behavior has been recorded in more than 120 species (Riedman 1982), no obligate "milk parasites" exist. A sizeable fraction of reptiles and amphibians also care for their young (Balshine 2012), and some poison dart frogs do engage in exploitation of parental care of a sort, though is not a form of brood parasitism considered here due to its facultative nature (Brown et al. 2008). Among invertebrates, parental care is seen in about 1% of insect species across over a dozen orders, as well as some spiders, scorpions, centipedes, crustaceans, and segmented worms (Trumbo 2012). Intraspecific brood parasitism in the form of "egg dumping" exists within several of these groups (e.g. Tallamy 2005), and some have been demonstrated to practice kin recognition in the context of maternal care, which may have evolved partially as a defense against brood

parasitism (Patterson et al. 2008). It is a certainty that there are undescribed cases of brood parasitism as we have defined it, and maybe in a novel major vertebrate or invertebrate taxonomic group. Despite the emphasis on brood parasitism in birds and the ubiquity of parental care in that group, the ~1% of avian brood parasites are actually less than double the overall frequency of ~0.6% for all animals (Table 1.1). When comparing major taxa on these grounds, the order Hymenoptera is the clear standout, with a prevalence of about 4% of species being obligate brood parasites, and collectively forming a majority of the world's brood parasitic species (both in terms of absolute species and number of independent origins). Perhaps the commonness of this strategy among wasps and bees is the result of an optimal combination of widely distributed parental care and a level of cognitive or sensory complexity that is more tolerant of brood parasitism than in most vertebrates, though this is largely speculative.

Host Specificity and Co-evolution

As with any guild of parasitic taxa, brood parasitic species vary in the degree to which they specialize on a certain group of hosts. For example, the brown-headed cowbird *Molothrus ater* has been recorded parasitizing nearly 300 species across over 30 host families, while the congeneric screaming cowbird *M. rufoaxillaris* utilizes just a few hosts from within its own family (Mann 2017). As discussed above, some conventional wisdom suggests that initial transitions to brood parasitism may often begin by exploiting close relatives, at least in the case of brood parasites derived from ancestors with parental care.

Under the strict form of Emery's rule, this implies sympatric speciation between parasitic and non-parasitic lineages of a single ancestral species, with the former becoming reproductively isolated from the latter while still depending on them for survival - unless they are able to expand their host breadth. Simulation studies by Cichoń (1996) shows that obligate parasites arising spontaneously within a population that also serves as their hosts face a significant risk of driving both groups to extinction, so the initial evolution of facultative parasitism may be a necessary prerequisite to allow hosts to build up some defenses before the obligate strategy can spread. Under the loose form of Emery's rule, however, parasitism of one or more related but already reproductively isolated host species presumably arises polymorphically and spreads to fixation (though still potentially in waves of facultative followed by obligate interspecific parasitism) without the risk of extinction through outcompeting its own hosts. In either case, it would seem that incipient brood parasites that do follow Emery's rule should be fairly host-specific, at least at first. However, among those lineages which have evolved brood parasitism from ancestral backgrounds which do not involve parental care, these dynamics do not apply, and it is more difficult to form expectations as to the generalist or specialist tendencies of such cases.

There are some indications that host preferences may have a tendency to trend towards generalization as a brood parasitic clade ages, as has been reported in the aforementioned cowbird genus *Molothrus* (Lanyon 1992). Certainly, it is true that older brood parasitic

clades often collectively attack a wider range of hosts than younger ones, as in the bee subfamily Nomadinae (Sless et al. 2022, or Chapter 2 herein). However, the question of whether this is due to increased host generalization at the specific level or merely niche partitioning and expansion by a growing number of specialized species can be difficult to ascertain. This is further complicated by the fact that some generalist brood parasites clearly contain specialized subpopulations (termed “host races” or “gentes”), as documented in the common cuckoo *Cuculus canorus* (Gibbs et al. 2000) as well as blood bees in the genus *Sphecodes* (Bogusch et al. 2006) - though others like the brown-headed cowbird show no signatures of this specialization, and are generalists at both the level of species and individual (Gibbs et al. 1997). The role of brood parasitism as a driver of intraspecific diversity in birds, for both parasites and their hosts, is also noted by Medina et al. (2020). Overall though, the evolutionary dynamics of host specialization remain poorly studied within most groups of brood parasites, and further study of the population genetics of generalists as well as mathematical modelling should prove useful in understanding one of the major mechanisms proposed to drive speciation in such taxa.

The intimate relationship between brood parasites and their hosts often leads to striking cases of co-evolution. Brood parasites often evolve to imitate their hosts, as exemplified in some birds by impressively faithful egg mimicry (Spottiswoode and Stevens 2010; Stoddard and Stevens 2010), visual imitation of host nestlings

(Langmore et al. 2011) and behavioral mimicry to induce feeding by hosts (Soler 2018). Chemical mimicry is also used in at least some brood parasitic insects (Litman 2019), though this is perhaps more common among social parasites (e.g. Uboni et al. 2012). Conversely, some brood parasitic birds use visual crypsis to hide themselves or their offspring from their hosts (Davies 2011), and some insects physically obscure their eggs in the host nest or mask them with chemical signatures (Litman et al. 2019).

The "evolutionary arms race" paradigm, in which two competing species continuously evolve offensive and defensive innovations, has often been invoked to describe the co-evolution of brood parasites and their hosts (e.g. Davies et al. 1989; Takasu 1998). The dynamics of this interaction differ to a degree from co-evolution in some other forms of host-parasite interactions. This is in part due to the fact that brood parasites still exist as free-living adults, and so cannot "commit" to traits that may be useful for parasitism but are detrimental to independent survival in the same way that an internal parasite might. Additionally, the general congruence of body size and generation time between brood parasites and their hosts likely assures that two such species evolve at generally similar rates. In fact, it may be the case that brood parasites are limited to smaller population sizes, and thus potentially slower rates of evolution, than their hosts (at least among specialist species), as has been suggested for some social parasites (Bousjein et al. 2016). However, the contrast between a parasite whose fitness hinges entirely on success and a host who may be capable of

tolerating a certain brood parasite load may swing the balance back in the favor of brood parasites. This dynamic is reminiscent of some cases of Batesian mimicry, which like brood parasitism involves an asymmetrical dependence between mimic species and their models (Joron and Mallet 1998).

Conclusion and Future Considerations

Brood parasitism is a fascinating strategy employed by species spanning vastly disparate - though unevenly distributed - branches in the animal tree of life. Though often excluded from conversations and studies on parasitism more generally, the life history features of brood parasites and their effects on hosts can be integrated with existing frameworks that seek to classify other strategies of exploitation. Following the criteria for brood parasitism as defined herein, a total of 59 independent origins of this strategy are identified in animals. The majority of these are seen in hymenopteran insects, with bees particularly well-represented, but also occur in other groups of insects, along with several lineages of birds and fish among vertebrates.

Comparing the evolution of brood parasitism in these groups allows for exploration of the effects this trait may have on the diversity of such lineages. Though conclusions vary depending on the specific methods used, there is a tendency for brood parasitic groups to have reduced species richness in direct comparison with their sister groups, and potentially also a negative diversification rate shift associated with the transition to this strategy. This is particularly manifest for brood

parasitic clades that are derived from an ancestral background of parental care, which make up about two thirds of considered groups, and in many cases these taxa utilize their close relatives as hosts - though this pattern is more evident among insects than vertebrates. Conversely, several lineages of insect brood parasites evolved as a result of switching from other forms of parasitism, and these cases tend to show much less predictable host associations as well as potentially higher evolutionary lability, with some examples of reversals to these ancestral strategies.

Beyond the brood parasitic lineages considered herein, it is likely that many more examples of this strategy remain undiscovered - and that the diversity of several known lineages is underestimated. The discovery and description of new species, in addition to further behavioral investigation into many poorly known ones, is essential to expand our knowledge of brood parasitism in animals. There is a substantial need for further phylogenetic studies covering a range of taxonomic scopes to investigate diversification rates. There are also several outstanding questions with respect to the evolutionary dynamics of brood parasite populations, which may be usefully investigated through both mathematical modelling and real-world population genetic study. Finally, we further echo recommendations such as those of Thorogood (2019) for further comparative study of brood parasites in the future, as well as the inclusion of brood parasitism within the wider literature on the behavior, ecology, and evolution of parasitism.

REFERENCES

Aguiar, A. P., Deans, A. R., Engel, M. S., Forshage, M., Huber, J. T., Jennings, J. T., Johnson, N. F., Lelej, A. S., Longino, J. T., Lohrmann, V., Mikó, I., Ohl, M., Rasmussen, C., Taeger, A., and Yu, D. S. K. 2013. Order Hymenoptera. Pages 51-62 *in* Zhang, Z.-Q., ed. Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness (addenda 2013). Zootaxa 3703(1). https://doi.org/10.11646/zootaxa.3703.1.12

Anderson, R. M. and May, R. M. 1979. Population biology of infectious diseases: part I. Nature 280:361-367. https://doi.org/10.1038/280361a0

Balshine, S. 2012. Patterns of parental care in vertebrates. Pages 62-80 *in* Royle, N. J., Smiseth, P. T. and Kölliker, M. eds. The evolution of parental care. Oxford University Press, Oxford, UK.

Barraclough, T. G., Harvey, P. H., and Nee, S. 1995. Sexual selection and taxonomic diversity in passerine birds. Proceedings of the Royal Society of London B 259:211–215. https://doi.org/10.1098/rspb.1995.0031

Barraclough, T. G., Harvey, P. H., and Nee, S. 1996. Rate of *rbc*L gene sequence evolution and species diversification in flowering plants (angiosperms). Proceedings of the Royal Society of London B 263:589–591. https://doi.org/10.1098/rspb.1996.0088

Billerman, S. M., Keeney, B. K., Rodewald, P. G., and Schulenberg, T. S. 2020. Birds of the world. Cornell Lab of Ornithology, Ithaca, NY, USA. Retrieved December 1st, 2021, from https://birdsoftheworld.org/bow/home

Bogusch, P., Kratochvíl, L., and Straka, J. 2006. Generalist cuckoo bees (Hymenoptera: Apoidea: *Sphecodes*) are species-specialist at the individual level. Behavioral Ecology and Sociology 60(3):422-429. https://doi.org/10.1007/s00265-006-0182-4

Bousjein, N. S., Gardner, M. G., and Schwarz, M. P. 2016. Small effective population sizes of bee social parasites compared to their hosts raise important questions for evolutionary arms race. Journal of Zoology 299(1):17-26. https://doi.org/10.1111/jzo.12325

Brockmann, H. J. 1993. Parasitizing conspecifics: comparisons between Hymenoptera and birds. Trends in Ecology and Evolution 8(1):2-4. https://doi.org/10.1016/0169-5347(93)90120-E

Brown, J. L., Morales, V., and Summers, K. 2008. Tactical reproductive parasitism via larval cannibalism in Peruvian poison frogs. Biology Letters 5(2):148-151. https://doi.org/10.1098/rsbl.2008.0591

Černá, K., Zemenová, M., Macháčková, L., Kolínová, Z., and Straka, J. 2013. Neighbourhood society: nesting dynamics, usurpations and social behaviour in solitary bees. PLoS ONE 8(9):e73806. https://doi.org/10.1371/journal.pone.0073806

Cichoń, M. 1996. The evolution of brood parasitism: the role of facultative parasitism. Behavioral Ecology 7(2):137-139. https://doi.org/10.1093/beheco/7.2.137

Davies, N. B., Bourke, A. F. G., and Brooke, M. de L. 1989. Cuckoos and parasitic ants: interspecific brood parasitism as an evolutionary arms race. Trends in Ecology and Evolution 4(9):274-278. https://doi.org/10.1016/0169-5347(89)90202-4

Dawkins, R. 1982. The extended phenotype: the gene as the unit of selection. Oxford University Press, Oxford, UK.

Emery, C. 1909. Über den Ursprung der dulotischen, parasitischen und myrmikophilen Ameisen. Biologisches Centralblatt 29:352-362.

Field, J. 1992. Intraspecific parasitism as an alternative reproductive tactic in nest-building wasps and bees. Biological Reviews 67:79-126. https://doi.org/10.1111/j.1469-185X.1992.tb01659.x

Fricke, R., Eschmeyer, W. N., and Fong, J. D. 2021. Species by family/subfamily. Eschmeyer's Catalog of Fishes. Retrieved December 8th, 2021, from http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp

Gibbs, H. L., Miller, P., Alderson, G., and Sealy, S. G. 1997. Genetic analysis of brown-headed cowbirds *Molothrus ater* raised by

different hosts: Data from mtDNA and microsatellite DNA markers. Molecular Ecology 6(2):189–193. https://doi.org/10.1046/j.1365-294X.1997.00171.x

Gibbs, H. L., Sorenson, M. D., Marchetti, K. Brooke, M. D., Davies, N. B., and Nakamura, H. 2000. Genetic evidence for host-specific races of the common cuckoo. Nature 407(6801):183-186. https://doi.org/10.1038/35025058

Hamanishi, Y. 1996. Cleptoparasitic life of the tachinid fly, *Symmorphomyia katayamai* Mesnil et Shima, in the nest of host wasps hunting the chrysomelid-larva prey. Japanese Journal of Entomology 64(4):843–860.

Hamilton, W. J., III and Orians, G. H. 1965. Evolution of brood parasitism in altricial birds. Condor 67(4):361-382.

Hay, E. M., Poulin, R., and Jorge, F. 2020. Macroevolutionary dynamics of parasite diversification: a reality check. Journal of Evolutionary Biology 33(12):1758-1769. https://doi.org/10.1111/jeb.13714

Huang, M. H. and Dornhaus, A. 2008. A meta-analysis of ant social parasitism: host characteristics of different parasitism types and a test of Emery's rule. Ecological Entomology 34(4):589-596. http://doi.org/10.1111/j.1365-2311.2008.01005.x

Jezkova, T. and Wiens, J. J. 2017. What explains patterns of diversification and richness among animal phyla? American Naturalist 189(3):201-212. https://doi.org/10.1086/690194

Joron, M. and Mallet, J. L. B. 1998. Diversity in mimicry: paradox or paradigm? Trends in Ecology and Evolution 13(11):461-466. https://doi.org/10.1016/S0169-5347(98)01483-9

Käfer, J. and Mousset, S. 2014. Standard sister clade comparison fails when testing derived character states. Systematic Biology 63(4):601-609. https://doi.org/10.1093/sysbio/syu024

Kimsey, L. S. and Bohart, R. M. 1990. The chrysidid wasps of the world. Oxford University Press, Oxford, UK.

Krüger, O., Sorenson, M. D., and Davies, N. B. 2009. Does coevolution promote species richness in parasitic cuckoos? Proceedings of

the Royal Society B 276(1674):3871-3879. https://doi.org/10.1098%2Frspb.2009.1142

Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. Molecular Biology and Evolution 34(7):1812-1819. http://doi.org/10.1093/molbev/msx116

Kuris, A. M. and Lafferty, K. D. 2000. Parasite-host modeling meets reality: adaptive peaks and their ecological attributes. Pages 9-26 *in* Poulin, R., Morand, S., and Skorping, A. eds. Evolutionary biology of host-parasite relationships: theory meets reality. Elsevier, Amsterdam, Netherlands.

Lafferty, K. D. and Kuris, A. M. 2002. Trophic strategies, animal diversity, and body size. Trends in Ecology and Evolution 17(11):507-513. https://doi.org/10.1016/S0169-5347(02)02615-0

Langmore, N. E., Stevens, M., Maurer, G., Heinsohn, R., Hall, M. L., Peters, A., and Kilner, R. M. 2011. Visual mimicry of host nestlings by cuckoos. Proceedings of the Royal Society B 278(1717):2455-2463. http://doi.org/10.1098/rspb.2010.2391

Lanyon, S. M. 1992. Interspecific brood parasitism in blackbirds (Icterinae): a phylogenetic perspective. Science 255(5040):77-79. http://doi.org/10.1126/science.1553533

Litman, J. R., Praz, C. J., Danforth, B. N., Griswold, T. L., and Cardinal, S. 2013. Origins, evolution, and diversification of cleptoparasitic lineages in long-tongued bees. Evolution 67(10):2982–2998. https://doi.org/10.1111/evo.12161

Litman, J. R. 2019. Under the radar: detection avoidance in brood parasitic bees. Philosophical Transactions of the Royal Society B 374:20180196. https://doi.org/10.1098/rstb.2018.0196

Lopez-Osario, F., Perrard, A., Pickett, A. M., Carpenter, J. M., and Agnarsson, I. 2015. Phylogenetic tests reject Emery's rule in the evolution of social parasitism in yellowjackets and hornets (Hymenoptera: Vespidae, Vespinae). Royal Society Open Science 2(9):150159. https://doi.org/10.1098/rsos.150159

Magnacca, K. N., and Danforth, B. N. 2006. Evolution and biogeography of native Hawaiian *Hylaeus* bees (Hymenoptera: Colletidae). Cladistics 22(5):393-411. https://doi.org/10.1111/j.1096-0031.2006.00119.x

Mann, C.F. 2017. A taxonomic review of obligate and facultative interspecific avian brood parasitism. Pages 61-92 *in* Soler, M. ed. Avian brood parasitism: behavior, ecology, evolution, and co-evolution. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-319-73138-4_4

McConway, K. J. and Sims, H. J. 2004. A likelihood-based method for testing for nonstochastic variation of diversification rates in phylogenies. Evolution 58:12-23. https://doi.org/10.1111/j.0014-3820.2004.tb01569.x

Medina, I., and Langmore, N. E. 2015. Coevolution is linked with phenotypic diversification but not speciation in avian brood parasites. Proceedings of the Royal Society B 282(1821):20152056. https://doi.org/10.1098/rspb.2015.2056

Medina, I., Kilner, R. M., and Langmore, N. E. 2020. From micro- to macroevolution: brood parasitism as a driver of phenotypic diversity in birds. Current Zoology 66(5):515-526. https://doi.org/10.1093/cz/zoaa033

Michener, C. D. 2007. The bees of the world. 2nd edition. John Hopkins University Press, Baltimore, Maryland, USA.

Paradis, E. 2012. Shift in diversification in sister-clade comparisons: A more powerful test. Evolution 66(1):288–295. https://doi.org/10.1111/j.1558-5646.2011.01429.x

Paradis, E. and Schliep, K. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526-528. https://doi.org/10.1093/bioinformatics/bty633

Patterson, L., Dick, J. T. A., and Elwood, R. W. 2008. Embryo retrieval and kin recognition in an amphipod (Crustacea). Animal Behaviour 76(3):717-722. https://doi.org/10.1016/j.anbehav.2008.03.018

Pauli, T., Castillo-Cajas, R. F., Rosa, P., Kukowka, S., Berg, A., van der Berghe, E., Fornoff, F., Hopfenmüller, S., Niehuis, M., Peters, R.

S., Staab, M., Strumia, F., Tischendorf, S., Schmitt, T., and Niehuis, O. 2018. Phylogenetic analysis of cuckoo wasps (Hymenoptera: Chrysididae) reveals a partially artificial classification at the genus level and a species-rich clade of bee parasitoids. Systematic Entomology 44(2):322-335. https://doi.org/10.1111/syen.12323

Polačik, M., Reichard, M., Smith, C., and Blažek, R. 2019. Parasitic cuckoo catfish exploit parental responses to stray offspring. Philosophical Transactions of the Royal Society B 374:20180412. https://doi.org/10.1098/rstb.2018.0412

Pollock, H. S., Hoover, J. P., Uy, F. M. K., and Hauber, M. E. 2021. Brood parasites are a heterogeneous and functionally distinct class of natural enemies. Trends in Parasitology 37(7):588-596. https://doi.org/10.1016/j.pt.2021.02.005

Poulin, R. 2011. The many roads to parasitism: a tale of convergence. Pages 1-40 *in* Rollinson, D. and Hay, S. I. eds. Advances in Parasitology 74. Academic Press, Cambridge, Massachusetts, USA. https://doi.org/10.1016/B978-0-12-385897-9.00001-X

Poulin, R., and Randhawa, H. S. 2015. Evolution of parasitism along convergent lines: from ecology to genomics. Parasitology 142:S6–S15. https://doi.org/10.1017/S0031182013001674

Rabeling, C. 2020. Social Parasitism. *In* Starr, C. ed. Encyclopedia of social insects. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-319-90306-4_175-1

Revell, L. J. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217-223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Riedman, M. L. 1982. The evolution of alloparental care and adoption in mammals and birds. The Quarterly Review of Biology 57(4):405-435. https://doi.org/10.1086/412936

Robert, M. and Sorci, G. 2001. The evolution of obligate interspecific brood parasitism in birds. Behavioral Ecology 12(2):128-133. https://doi.org/10.1093/beheco/12.2.128

Roldán, M. and Soler, M. 2011. Parental-care parasitism: how do unrelated offspring attain acceptance by foster parents? Behavioral Ecology 22(4):679-691. https://doi.org/10.1093/beheco/arr041

Sargent, R. D. 2004. Floral symmetry affects speciation rates in angiosperms. Proceedings of the Royal Society of London B 271:603-608. https://doi.org/10.1098/rspb.2003.2644

Sless, T. J. L., Branstetter, M. G., Gillung, J. P., Krichilsky, E. A., Tobin, K. B., Straka, J., Rozen Jr., J. G., Freitas, F. V., Martins, A. C., Bossert, S., Searle, J. B., Danforth, B. N. 2022. Phylogenetic relationships and the evolution of host preferences in the largest clade of brood parasitic bees (Apidae: Nomadinae). Molecular Phylogenetics and Evolution 166:107326. https://doi.org/10.1016/j.ympev.2021.107326

Soler, M. 2018. Begging behaviour, food delivery, and food acquisition in nests with brood parasitic nestlings. Pages 493-515 *in:* Soler, M. ed. Avian brood parasitism: behavior, ecology, evolution, and co-evolution. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-319-73138-4_27

Spottiswoode, C. N. and Stevens, M. 2010. Visual modeling shows that avian host parents use multiple visual cues in rejecting parasitic eggs. Proceedings of the National Academy of Sciences USA 107(19):8672-8676. https://doi.org/10.1073/pnas.0910486107

Stoddard, M. C. and Stevens, M. 2010. Pattern mimicry of host eggs by the common cuckoo, as seen through a bird's eye. Proceedings of the Royal Society B 277(1686):1387-1393. https://doi.org/10.1098/rspb.2009.2018

Stork, N. E. 2018. How many species of insects and other terrestrial arthropods are there on Earth? Annual Review of Entomology 63:31-45. https://doi.org/10.1146/annurev-ento-020117-043348

Takasu, F. 1998. Modelling the arms race in avian brood parasitism. Evolutionary Ecology 12:969-987. https://doi.org/10.1023/A:1006520323251

Tallamy, D. W. 2005. Egg dumping in insects. Annual Review of Entomology 50:347-370. https://doi.org/10.1146/annurev.ento.50.071803.130424

Thorogood, R., Spottiswoode, C. N., Portugal, S. J., and Gloag, R. 2019. The coevolutionary biology of brood parasitism: A call for integration. Philosophical Transactions of the Royal Society B 374:20180190. https://doi.org/10.1098/rstb.2018.0190

Trumbo, S. T. 2012. Patterns of parental care in invertebrates. Pages 81-100 *in* Royle, N. J., Smiseth, P. T., and Kölliker, M. eds. The evolution of parental care. Oxford University Press, Oxford, UK.

Uboni, A., Bagnères, A.-G., Christidès, J.-P., and Lorenzi, M. C. 2012. Cleptoparasites, social parasites, and a common host: chemical insignificance for visiting host nests, chemical mimicry for living in. Journal of Insect Physiology 58(9):1259-1264. https://doi.org/10.1016/j.jinsphys.2012.06.013

Weinstein, S. B., and Kuris, A. M. 2016. Independent origins of parasitism in Animalia. Biology Letters 12(7):20160324. https://doi.org/10.1098/rsbl.2016.0324

Windsor, D. A. 1998. Most of the species on Earth are parasites. International Journal for Parasitology 28(12):1939–1941. https://doi.org/10.1016/s0020-7519(98)00153-2

Yamane, H., Watanabe, K., and Nagata, Y. 2009. Flexibility of reproductive tactics and their consequences in the brood parasitic fish *Pungtungia herzi* (Teleostei: Cyprinidae). Journal of Fish Biology 75(3):563–574. https://doi.org/10.1111/j.1095-8649.2009.02306.x

Yom-Tov, Y. 2001. An updated list and some comments on the occurrence of intraspecific nest parasitism in birds. Ibis 143:133-143. https://doi.org/10.1111/j.1474-919X.2001.tb04177.x

Yom-Tov, Y. and Geffen, E. 2006. On the origin of brood parasitism in altricial birds. Behavioral Ecology, 17(2):196-205. https://doi.org/10.1093/beheco/arj013

Yom-Tov, Y. and Geffen, E. 2017. Conspecific brood parasitism among birds: the effects of phylogeny, mode of reproduction, and

geographic distribution. Pages 95-103 *in* Soler, M. ed. Avian brood parasitism: behavior, ecology, evolution, and co-evolution. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-319-73138-4_5

Zhang, Z.-Q. 2013a. Animal biodiversity: An update of classification and diversity in 2013. Pages 5-11 *in* Zhang, Z.-Q. ed. Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness (addenda 2013). Zootaxa 3703(1). https://doi.org/10.11646/zootaxa.3703.1.3

Zhang, Z.-Q. 2013b. Phylum Arthropoda. Pages 17-26 *in*: Zhang, Z.-Q. ed. Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness (addenda 2013). Zootaxa 3703(1). https://doi.org/10.11646/zootaxa.3703.1.6

CHAPTER 2

PHYLOGENETIC RELATIONSHIPS AND THE EVOLUTION OF HOST PREFERENCES IN THE LARGEST CLADE OF BROOD PARASITIC BEES (APIDAE: NOMADINAE)

Abstract

Brood parasites (also known as cleptoparasites) represent a substantial fraction of global bee diversity. Rather than constructing their own nests, these species instead invade those of host bees to lay their eggs. Larvae then hatch and consume the food provisions intended for the host's offspring. While this life history strategy has evolved numerous times across the phylogeny of bees, the oldest and most speciose parasitic clade is the subfamily Nomadinae (Apidae). However, the phylogenetic relationships among brood parasitic apids both within and outside the Nomadinae have not been fully resolved. Here, we present new findings on the phylogeny of this diverse group of brood parasites based on ultraconserved element (UCE) sequence data and extensive taxon sampling with 114 nomadine species representing all tribes. We suggest a broader definition of the subfamily Nomadinae to describe a clade that includes almost all parasitic members of the family Apidae. The tribe Melectini forms the sister group to all other Nomadinae, while the remainder of the subfamily is composed of two sister clades: a "nomadine line" representing the former Nomadinae *sensu stricto*, and an "ericrocidine line" that unites several, mostly Neotropical, lineages. We find the tribe Osirini Handlirsch to be polyphyletic, and divide it into three lineages,

including the newly described Parepeolini *trib. nov*. In addition to our taxonomic findings, we use our phylogeny to explore the evolution of different modes of parasitism, detecting two independent transitions from closed-cell to open-cell parasitism. Finally, we examine how nomadine host-parasite associations have evolved over time. In support of Emery's rule, which suggests close relationships between hosts and parasites, we confirm that the earliest nomadines were parasites of their close free-living relatives within the family Apidae, but that over time their host range broadened to include more distantly related hosts spanning the diversity of bees. This expanded breadth of host taxa may also be associated with the transition to open-cell parasitism.

Introduction

Bees are known to display a wide variety of life history strategies, including diverse plant associations, ways of collecting food and building nests, and varying levels of sociality. The evolution of eusociality in particular has been the focus of much research attention in bees, in part due to the close association between some social species and human agriculture. Brood parasitism is another fascinating life history which has received substantially less research attention than eusociality despite being taxonomically more frequent, with about one in eight of the 20,000 species of bee adopting this strategy compared with fewer than one in ten for eusociality (Danforth et al. 2019). Brood parasitic bees have important ecological consequences for their bee hosts, in some cases causing greater brood

loss than any other group of nest predators/parasites, including beetles, flies, and other hymenopterans (Minckley and Danforth 2019). The prevalence of both brood parasitic (also known as “cuckoo”) species and eusocial species within bees is higher than almost all other animal lineages, which may partially be explained by the early evolution of food provisioning in bees - something that may reasonably be considered a prerequisite for both strategies (at least insofar as brood parasites must target a host that exhibits such behavior).

The exact number of origins of brood parasitism within bees is unclear. Michener (2007) posited as many as thirty independent transitions from solitary species to brood parasites, although more recent studies reduce this number while still recognizing several convergent transition events (Cardinal et al. 2010; Litman et al. 2013). One thing that is broadly agreed, however, is that there is no evidence for a reversal from brood parasitism back to pollen provisioning. Additionally, it is clear that there is a high degree of variation in species richness across brood parasitic groups, with some being much more species-rich than others. The reasons for this variation in species diversity have been the subject of some previous work (Litman et al. 2013; Policarová et al. 2019) but are still not fully understood. The oldest and most diverse brood parasitic group is the subfamily Nomadinae Latreille (1802) within the family Apidae. As traditionally defined, this group contains approximately 1200 species – nearly half of all brood parasitic bees – and has a cosmopolitan distribution (Michener 2007; Danforth et al. 2019).

In addition to the tribes that make up the Nomadinae *sensu stricto* as described by Latreille, various studies over the past decade (Cardinal et al. 2010 and 2018, Litman et al. 2013, Policarová et al. 2019, Bossert et al. 2019) have indicated that several other brood parasitic lineages within the Apidae are closely related to this group. The expanded Nomadinae *sensu* Bossert et al. 2019, including these taxa, is itself a monophyletic group of brood parasites containing closer to 1600 species (Ascher and Pickering 2020). We will henceforth use this broader definition of Nomadinae and refer to Nomadinae *sensu stricto* as the "nomadine line" within this clade. As a whole, this newly defined Nomadinae contains approximately 60% of all brood parasitic bees and has a crown age of 77.2-109.7 million years (Litman et al. 2013). Though the monophyly of Nomadinae is supported by all other recent studies, the relationships within this group are more contentious. Previous studies into the phylogenetic relationships of the Nomadinae have either relied on relatively little molecular data (Cardinal et al. 2010, Litman et al. 2013, Policarová et al. 2019) or a limited number of taxa (Bossert et al. 2019).

As a group, parasites in the subfamily Nomadinae attack a wide range of hosts across the phylogeny of bees. These include other subfamilies of Apidae, as well as hosts in the families Andrenidae, Halictidae, Colletidae, and Melittidae. As of yet, no members of the Nomadinae are known to parasitize members of family Megachilidae. This may be due to differences in nesting biology; most megachilids

nest in above-ground cavities, while all other hosts attacked by nomadines are ground-nesting (Danforth et al. 2019). Additionally, there are no known associations between nomadines and members of the depauperate Australian family Stenotritidae. Most likely, this is a result of the low abundance and high endemism of this family, as well as the relative paucity of apid brood parasites in Australia, being represented on the continent by just one species of *Nomada* and about ten species of *Thyreus* (Houston 2018).

At the level of individual genera and species, it is difficult to ascertain to what extent nomadines have specialized on their hosts. Most nomadine genera are consistent in parasitizing a set of closely related hosts or a single host genus, but reliable host association data are rare or absent in the literature for many species. In some cases, such as a few of the better-studied *Nomada*, a single species has been recorded attacking hosts from multiple bee families, though their most common host by far appears to be the genus *Andrena* (Snelling 1986, Alexander 1991). Within-species size variation has in some cases been interpreted as evidence of multiple hosts, though this may also represent cryptic diversity or simply environmental effects (Michener 2007). It is, however, clear that apid brood parasites like the Nomadinae are more specialized than some other cuckoo bees, such as the generalist genus *Sphecodes* in the family Halictidae (Habermannová et al. 2013).

While all members of the Nomadinae are obligate brood

parasites, they differ in the details of how they exploit their hosts. Some species wait for the host to finish building, provisioning, and sealing up a nest before invading. Females of these "closed-cell" parasites will then break into the brood cell, lay their own eggs, and reseal it. In contrast, others are "open-cell" parasites. Females of these species invade a nest while it is still under construction or being provisioned and then lay their eggs in brood chambers, but do not seal up the nest afterwards. Exceptions to this dichotomy do exist, such as the genus *Epeoloides*, which has been observed invading unfinished cells but closing them off afterwards, combining aspects of both strategies (Straka and Bogusch 2007). The discovery of up to four *Epeoloides* eggs/larvae of different ages within a single cell indicates that this strategy also allows females to attack nests that have already been parasitized by other individuals. Different modes of parasitism have resulted in corresponding differences in behavior and morphology in the Nomadinae. For example, open-cell parasites lay smaller eggs than non-parasitic species of similar body size (Iwata and Sakagami 1966, Rozen 2003). These eggs often have conspicuous tubercules, flanges, or other modifications, and are typically hidden against the brood cell wall, presumably to avoid detection and removal by hosts (Rozen and Özbek 2003). Closed-cell parasites, meanwhile, have average-sized eggs, likely because hosts will not return to investigate a finished nest. In some brood parasitic bees, adult females will kill or remove host eggs/larvae. However, in all members of the Nomadinae regardless of the mode of parasitism, host eggs or larvae are killed directly by the parasitic larva instead. These so-called "hospicidal"

larvae typically have enormous mandibles during their first instar which are used for this purpose but lost after molting (Michener 2007).

Over a century ago, Carlo Emery suggested that, in socially parasitic or brood parasitic insect species, hosts are typically closely related to the parasites themselves (Emery, 1909). The rationale behind this idea stems largely from the possibility that these types of parasites may at first evolve intraspecifically, as suggested by some models (Zink 2000) and observed cases in both parasitic birds and ants (Petrie and Møller 1991; Rabeling et al. 2014). Additionally, the possibility of shared chemical signals for mimicry, as is sometimes seen with parasitic ants (Lenoir et al. 2001), as well as the higher likelihood of sharing other aspects of life history (e.g. diet, habitat, seasonality) represent plausible factors which may create an expectation of close relationships between parasites and hosts. Within some literature, "Emery's rule", as it has come to be known, is often divided into a "strict" form (requiring that a parasitic lineage be sister to its host) and a "loose" form, which merely suggests that parasites and hosts are generally closely related (Ward 1989; Huang and Dornhaus 2008). In either sense, Emery's rule suggests that the earliest brood parasites within a group were likely associated with their close relatives, which raises an interesting question: can the signal of these ancestral host-parasite relationships still be detected in a group as diverse as the Nomadinae, which transitioned to brood parasitism tens of millions of years ago?

In this study, we present the most comprehensive exploration of the nomadine phylogeny to date, including an unprecedented level of taxon sampling and a wealth of molecular data provided by ultraconserved element (UCE) sequencing. The resulting phylogenetic tree is subsequently used as a framework to analyze traits of interest, including the evolution of host preferences (providing an opportunity to examine the applicability of Emery's rule) and transitions between open- and closed-cell modes of parasitism.

Materials and Methods

Sample collection

A total of 114 samples of brood parasites from within the family Apidae were obtained, including several species that have not been included in previous phylogenetic studies. Collectively, these represent 56 of the 62 total genera that comprise the Nomadinae, and original collection localities ranged across all six continents where this subfamily can be found. An additional five outgroup taxa representing other major groups of Apidae (*Apis mellifera*, *Bombus nevadensis*, *Centris hoffmanseggiae*, *Eufriesea surinamensis*, and *Habropoda laboriosa*) were also included, bringing the total number of samples used in phylogenetic analyses up to 119.

Samples were assembled from a combination of museum and laboratory collections. DNA was extracted from pinned or ethanol-preserved voucher specimens (n=40), or in some cases was already available as a result of previously conducted extractions (n=68).

Additionally, data for some samples (including outgroup taxa) were obtained in the form of already processed sequence data from published or in-preparation datasets (Bossert et al. 2019 (n=5), Grab et al. 2019 (n=4), Freitas et al. 2020 (n=2)). Collection methods varied by specimen and are unknown in some cases, but most were caught by hand-netting. See Appendix II Table 2.1 for full details on voucher specimen sources, collection dates, and localities.

DNA extraction

For samples in which DNA extraction had not already been carried out (n=40), a phenol-chloroform based protocol was used (Danforth 2011a). First, tissue samples were placed into a 2x CTAB solution and ground with a pestle for 30 seconds. When possible, a single leg was taken from pinned or ethanol-preserved specimens for DNA extraction, though in some cases multiple legs were used. For a few exceptionally small individuals, the entire body was destructively sampled. After grinding, proteinase K was added, and samples were incubated at 55°C overnight.

DNA extraction continued the next day with the addition of chloroform:isoamyl alcohol (24:1), followed by centrifugation and aspiration of the supernatant. This was followed with a phenol:chloroform:isoamyl alcohol (25:24:1) treatment, and then a final chloroform:isoamyl alcohol treatment to wash out any remaining phenol. DNA pellets were precipitated in 100% ethanol with sodium acetate, then washed in 80% ethanol again before final resuspension

in Tris-EDTA buffer.

For the samples obtained in the form of previously extracted DNA, extraction methods varied, but most were obtained by either the same phenol-chloroform extraction as detailed above, or with the use of either a DNeasy Blood and Tissue kit (Qiagen) or a Quick-DNA Miniprep Plus kit (Zymo Research).

All samples were subsequently measured on a Nanodrop 2000 and Qubit 3 or 4 Fluorometer, both Thermo Fisher Scientific, to estimate DNA quantity and quality. A subset of samples was further analyzed on an Agilent 2200 TapeStation machine with D1000 HS tapes (Agilent Technologies) to estimate the size distribution of DNA samples based on age.

UCE library preparation and enrichment

We used a targeted UCE approach to generate sequence data following previous literature (Faircloth et al. 2012, 2015). The protocols used followed those outlined in Branstetter et al. (2017) and were carried out at the USDA-ARS Pollinating Insects Research Unit in Logan, Utah, USA.

We targeted a set of 2,527 UCE loci and additional "legacy" loci using baits based on the Hymenoptera v2 probe set outlined by Branstetter et al. (2017), with ant-bee specific probes as described in Grab et al. (2019). Probes were synthesized by Arbor Biosciences,

previously MYcroarray.

First, extracted DNA samples were sheared to reduce the average fragment size to a target of ~400-600 bp. Older or more degraded samples were not sheared, while other samples were sheared in a Q800R3 sonicator (Qsonica) for 30, 60, or 90 seconds depending on sample quality and predicted size distribution. Mean final DNA input mass for all samples was 102 ng but ranged from <1 ng – 1,630 ng.

Library preparation involved the use of a KAPA HyperPrep kit (Roche Sequencing Systems) for enzymatic steps including repair of fragment ends and addition of A-tails, and Illumina TruSeq-style adapters (Glenn et al. 2019) for dual-indexing with redundantly unique sequences. DNA-binding magnetic beads were made following an in-house protocol based on Rohland and Reich (2012) and were used to clean and concentrate samples at various steps in the process. After the final bead cleaning, samples were measured for DNA concentration with a Qubit 3 fluorometer and pooled in groups of 8-10 at equimolar concentrations.

These pooled samples were enriched following a protocol from Arbor Biosciences (v4) for the first day of UCE enrichment, and a standard UCE protocol for the second day (enrichment protocol v1.5 available at ultraconserved.org, based on Blumenstiel et al. (2010)). Post-enrichment samples were measured on a TapeStation machine

for fragment size distributions and size selected with a Blue Pippin machine (Sage Science) for a range of 200-700 base pairs, if necessary. Finally, pooled libraries were quantified with an Applied Biosystems qPCR machine and KAPA reagents, pooled together, and sent off for sequencing.

Sequencing

After library preparation and enrichment, a final total of 97 samples were sent to Novogene inc. for multiplexed sequencing on a single lane of an Illumina HiSeq X instrument (paired-end 150 bp). A total of approximately 360 Gb of sequencing data was received. Of these, 3 samples were ultimately not used due to the presence of redundant taxa. The remaining 94 UCE assemblies were combined with 20 previously generated datasets as well as UCE sequences for 5 outgroup taxa from Bossert et al. (2019), as mentioned above.

In silico processing

Sequence data were demultiplexed to sort reads to their respective samples using BBMap (accessed from https://sourceforge.net/projects/bbmap/). Most processing was done through the PHYLUCE pipeline (Faircloth 2016). First, reads were trimmed and adapters were removed using illumiprocessor (Faircloth 2013), which is a wrapper software based on Trimmomatic (Bolger et al. 2014). Removal of adapters was assessed using FastQC (Andrews 2010) for quality control. Reads for each sample were then assembled into contigs using SPAdes (Bankevich et al. 2012). The software LastZ

(Harris 2007) was then used to identify contigs containing UCE sequences which matched the probes from the Hymenoptera v2 "ant-bee" probe set, with "min-identity" and "min-coverage" parameters set to 80.

These contigs were then aligned using MAFFT v7.31 (Katoh and Standley 2013), followed by internal trimming using Gblocks (Talavera and Castresana 2007) as recommended by Faircloth (2019), in both cases using default parameters. Data matrices were then created to test different levels of taxon completeness. Separate alignments were made which included all UCE loci present in greater than or equal to 75%, 85%, and 95% of samples, respectively.

Phylogenetic trees

Concatenated data matrices for the three different levels of taxon completeness were used to generate unpartitioned phylogenies in IQ-TREE v1.6.9 (Nguyen et al. 2015). In each case, ModelFinder (Kalyaanamoorthy et al. 2017) was first used to select an appropriate model, and then a phylogeny was created with 1000 replicates for approximate maximum likelihood ratio tests (Guindon et al. 2010) and ultra-fast bootstraps (Hoang et al. 2018). Further analyses were conducted only on the 95% taxon-completeness matrix, which was selected due to the minimal amount of missing data. First, gene trees were generated for all 1247 UCE loci in this matrix using IQ-TREE v2.1.2 (Minh et al. 2020a). These were subsequently used to calculate gene and site concordance factors in IQ-TREE (Minh et al. 2020b), and

also to create a coalescent species tree with ASTRAL v5.7.4 (Zhang et al. 2018) on default settings (including calculation of local posterior probabilities). Finally, a partitioning analysis of loci in the 95% taxon-completeness matrix was conducted using the sliding-window site characteristics method (SWSC; Tagliacollo and Lanfear 2018). The resulting partitioning scheme by entropy (SWSC-EN), along with a separate by-locus partitioning scheme, were both fed into IQ-TREE v2.1.2 using the "testmerge" option (Chernomor et al. 2016) and a GTR+G model to generate partitioned trees. All phylogenies were then edited for clarity in FigTree v1.4.4 (Rambaut 2014).

Trait evolution analysis

To reconstruct the ancestral hosts parasitized by Nomadinae, each brood parasitic species' host preferences were identified at the family level through a literature search (see Appendix II Tables 2.2 and 2.3 for character state data). Additionally, mode of parasitism was analyzed as a character, distinguishing open- and closed-cell parasites. Ancestral state reconstructions were conducted on the SWSC-EN partitioned 95% taxon-completeness matrix phylogeny in Mesquite v3.61 (Maddison and Maddison 2019). The "trace characters over trees" option was used after inputting relevant phylogenies and character data, using a maximum likelihood reconstruction method with the Markov k-state 1 parameter model ("Mk1"). Since Mesquite is not able to work with polymorphic states at tips, a few species that have been recorded attacking multiple host families were coded according to the most commonly recorded host (if clear), or as

"unknown" otherwise.

Nomenclature

Names of the tribes were adopted from Michener (2007), Engel (2005), a subsequent revision of Neolarrini (Bossert et al. 2020), and the tribe Coelioxoidini (Martins et al. 2018, Bossert et al. 2019, Engel et al. 2020). For the new names, including lineages called "lines", we applied family group name rules (ICZN 1999). The "nomadine line" used herein is identical to the *sensu stricto* definition of Nomadinae Latreille 1802, and the "ericrocidine line" corresponds to the group of the same name first used by Litman et al. (2013).

Results

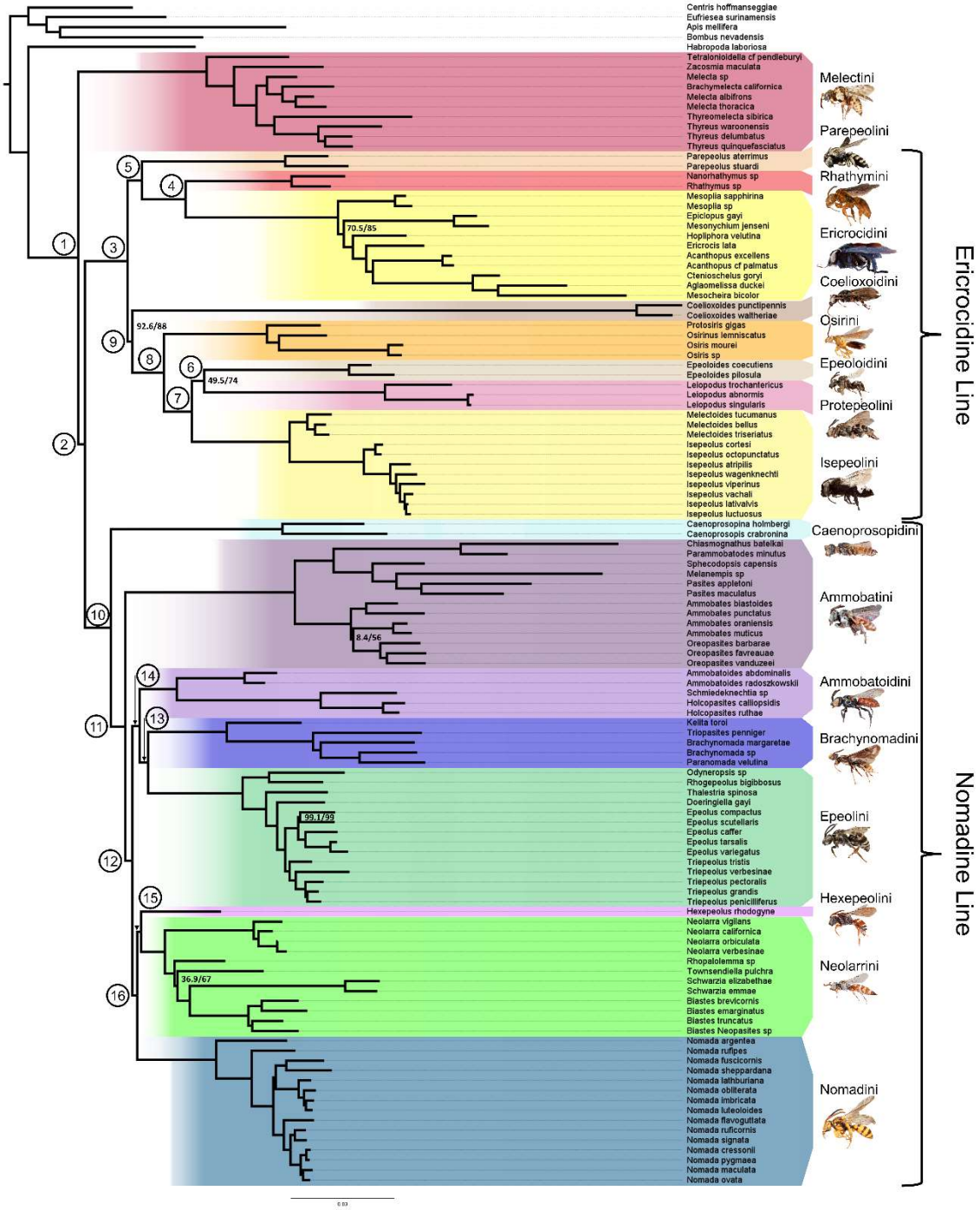
Topologies of generated phylogenetic trees

We created three data matrices consisting of loci that were recovered from 75%, 85%, and 95% of sampled taxa. These resulted in final datasets of 2048, 1833, and 1247 UCE loci respectively. The latter of these, with a total alignment length of 366,640 bp and the lowest proportion of missing data at approximately 4.96%, was then used to generate two partitioning schemes, one by locus, and one using the SWSC-EN method (Tagliacollo and Lanfear 2018). Within-locus partitioning has been shown to improve phylogenetic inference in similar datasets (e.g. Freitas et al. 2020), and so for this reason we used the SWSC-EN partitioned phylogeny for the ancestral state reconstructions discussed below, but the overall conclusions of these analyses are consistent with all generated topologies. This final

phylogeny consisted of 114 nomadine species as tips, as well as five outgroup species (Figure 2.1). A majority of nodes in the tree were recovered with 100/100 support values according to SH-aLRT (Guindon et al. 2010) and ultra-fast bootstrap (Hoang et al. 2018) metrics respectively, though six nodes had less than 100/100 support.

Figure 2.2: Phylogeny of Nomadinae based on concatenated, SWSC-EN partitioned phylogeny (with partitions merged) generated with IQ-TREE2 using the 95% taxon-completeness matrix of 1247 UCE loci. SH-aLRT and ultra-fast bootstrap values are indicated for some nodes; all unlabeled nodes have 100/100 support. All tribes are highlighted for clarity; members of the nomadine line are highlighted in cool colors, while the ericrocidine line and tribe Melectini are shown in warm colors. Numbered circles at some nodes correspond to clades listed in Table 1. From top to bottom, images depict *Zacosmia maculata* ♀, *Parepeolus stuardi* ♀, *Rhathymus* sp. ♀, *Acanthopus* sp. ♂, *Coelioxoides waltheriae* ♀, *Osiris* sp. ♀, *Epeoloides pilosula* ♀, *Leiopodus singularis* ♀, *Isepeolus wagenknechti* ♀, *Caenoprosopina holmbergi* ♀, *Oreopasites favreauae* ♀, *Holcopasites calliopsidis* ♀, *Paranomada velutina* ♀, *Triepeolus pectoralis* ♀, *Hexepeolus rhodogyne* ♀, *Neolarra verbesinae* ♀, and *Nomada luteoloides* ♀. Images courtesy of Laurence Packer and USGS Bee Inventory and Monitoring Lab.

Figure 2.1:


Centris hoffmanseggiae
Eufriesea surinamensis
Apis mellifera
Bombus nevadensis
Habropoda laboriosa
Tetraloniodella cf pendleburyi
Zacosmia maculata
Melecta sp
Brachymelecta californica
Melecta albifrons
Melecta thoracica
Thyreomelecta sibirica
Thyreus waroonensis
Thyreus delumbatus
Thyreus quinquefasciatus
Parepeolus aterrimus
Parepeolus stuardi
Nanorhathymus sp
Rhathymus sp
Mesoplia sapphirina
Mesoplia sp
Epiclopus gayi
Mesonychium jenseni
Hopliphora velutina
Ericrocis lata
Acanthopus excellens
Acanthopus cf palmatus
Ctenioschelus goryi
Aglaomelissa duckei
Mesocheira bicolor
Coelioxoides punctipennis
Coelioxoides waltheriae
Protosiris gigas
Osirinus lemniscatus
Osiris mourei
Osiris sp
Epeoloides coecutiens
Epeoloides pilosula
Leiopodus trochantericus
Leiopodus abnormis
Leiopodus singularis
Melectoides tucumanus
Melectoides bellus
Melectoides triseriatus
Isepeolus cortesi
Isepeolus octopunctatus
Isepeolus atripilis
Isepeolus wagenknechti
Isepeolus viperinus
Isepeolus vachali
Isepeolus lativalvis
Isepeolus luctuosus
Caenoprosopina holmbergi
Caenoprosopis crabronina
Chiasmognathus bateikai
Parammobatodes minutus
Sphecodopsis capensis
Melanempis sp
Pasites appletoni
Pasites maculatus
Ammobates blastoides
Ammobates punctatus
Ammobates oraniensis
Ammobates muticus
Oreopasites barbarae
Oreopasites favreauae
Oreopasites vanduzeei
Ammobatoides abdominalis
Ammobatoides radoszkowskii
Schmiedeknechtia sp
Holcopasites calliopsidis
Holcopasites ruthae
Keita lorai
Triopasites penniger
Brachynomada margaretae
Brachynomada sp
Paranomada velutina
Odyneropsis sp
Rhogepeolus bigibbosus
Thalestria spinosa
Doeringiella gayi
Epeolus compactus
Epeolus scutellaris
Epeolus caffer
Epeolus tarsalis
Epeolus variegatus
Triepeolus tristis
Triepeolus verbesinae
Triepeolus pectoralis
Triepeolus grandis
Triepeolus penicilliferus
Hexepeolus rhodogyna
Neolarra vigilans
Neolarra californica
Neolarra orbiculata
Neolarra verbesinae
Rhopalolemma sp
Townsendiella pulchra
Schwarzia elizabethae
Schwarzia emmae
Biastes brevicornis
Biastes emarginatus
Biastes truncatus
Biastes Neopasites sp
Nomada argentea
Nomada rufipes
Nomada fuscicornis
Nomada sheppardana
Nomada lathburiana
Nomada obliterata
Nomada imbricata
Nomada luteoloides
Nomada flavoguttata
Nomada ruficornis
Nomada signata
Nomada cressonii
Nomada pygmaea
Nomada maculata
Nomada ovata
Melectini
Parepeolini
Rhathymini
Ericrocidini
Coelioxoidini
Osirini
Epeoloidini
Protepeolini
Isepeolini
Caenoprosopidini
Ammobatini
Ammobatoidini
Brachynomadini
Epeolini
Hexepeolini
Neolarrini
Nomadini
Ericrocidine Line
Nomadine Line
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
70.5/85
92.6/88
49.5/74
8.4/56
99.1/99
36.9/67

The unpartitioned phylogenies (Appendix III Figures 2.1-2.3), locus-partitioned phylogeny (Appendix III Figure 2.4), and ASTRAL coalescent phylogeny (Appendix III Figure 2.5) were all topologically similar to the concatenated SWSC-EN partitioned tree, differing at only a few nodes. While the tribe Coelioxoidini was recovered as the sister group to the clade consisting of Osirini, Epeoloidini, Protepeolini, and Isepeolini in the SWSC-EN partitioned phylogeny, the unpartitioned 75% and 85% taxon-completeness matrix trees, as well as the ASTRAL coalescent phylogeny, recovered it instead as the sister group to all other members of the ericrocidine line. The tribe Epeoloidini was typically recovered as the sister to Protepeolini, but instead appeared as the sister to Protepeolini + Isepeolini in the unpartitioned 95% matrix and ASTRAL trees. The relationship among *Townsendiella*, *Rhopalolemma*, and the clade consisting of *Biastes* and *Schwarzia* also varied. While both partitioned phylogenies recovered *Townsendiella* and *Rhopalolemma* as successive sister groups to this latter clade, the unpartitioned phylogenies reversed their positions, and the ASTRAL phylogeny recovered a sister relationship between *Townsendiella* and *Rhopalolemma* instead. Within the tribe Ammobatini, the SWSC-EN partitioned and all unpartitioned phylogenies recovered a weakly-supported paraphyletic *Ammobates* with respect to *Oreopasites*, but this was not recovered in either the locus-partitioned or ASTRAL phylogenies. Finally, the ASTRAL phylogeny uniquely recovered *Epeolus scutellaris* as the sister group to *E. caffer*, *E. tarsalis*, and *E. variegatus*, while all other analyses instead recovered a sister relationship between *E. scutellaris* and *E. compactus*. Unsurprisingly,

these nodes had lower gene and site concordance factors than most other parts of the phylogeny (Appendix III Figure 2.6).

Monophyletic Nomadinae includes almost all brood parasites within Apidae

As suggested by other recent molecular studies, the subfamily Nomadinae in the broad sense forms a monophyletic clade consisting of almost all the brood parasitic species within Apidae (Figure 2.1). The tribes Melectini, Isepeolini, Protepeolini, Ericrocidini, Osirini, Rhathymini, and Coelioxoidini are descendants of a single parasitic common ancestor shared with the Nomadinae *sensu stricto* (henceforth "nomadine line"). Though this study does not have as broad a selection of outgroup taxa as some previous analyses (e.g. Cardinal et al. 2010, Bossert et al. 2019, Policarová et al. 2019), our recovered topology is consistent with suggestions that the subfamily Anthophorinae *sensu* Bossert et al. (2019) is the sister group to Nomadinae.

Melectini is the sister group to all other Nomadinae

All phylogenies generated as part of this study place the tribe Melectini as the sister group to all other members of Nomadinae (Figure 2.1). This result is consistent with one previous study (Litman et al. 2013) but differs from others which found Melectini to be the sister group to the nomadine line (Cardinal et al. 2010, Policarová et al. 2019) or to Caenoprosopidini (Cardinal et al. 2018).

Other apid parasites form a predominantly Neotropical "ericrocidine line"

Besides Melectini, the other parasitic tribes not included within the nomadine line form a monophyletic group (Figure 2.1). This "ericrocidine line" (including the tribes Osirini, Rhathymini, Ericrocidini, Coelioxoidini, Protepeolini, and Isepeolini, as well as the revived Epeoloidini and Parepeolini *trib. nov.*) is also suggested by Cardinal et al. (2010), Litman et al. (2013), and Policarová et al. (2019). This clade is almost entirely Neotropical, with the exceptions of *Ericrocis*, which extends north to the southern Nearctic, and *Epeoloides*, which is found in the Nearctic and Palearctic. The tribes in this ericrocidine line mainly consist of parasites on oil-collecting bees and may represent a radiation associated with these hosts as discussed by Policarová et al. (2019).

Internal relationships among nomadine tribes

To the extent that individual tribes within the subfamily Nomadinae have been studied, our phylogeny generally supports previously suggested topologies. Within the tribe Melectini, the finding of *Brachymelecta* (*sensu* Onuferko et al. 2021) nested within the genus *Melecta* is somewhat unexpected, though there have been some preliminary suggestions that the latter genus may be paraphyletic (M. Orr, personal communication).

Our phylogeny also identified the tribe Osirini as polyphyletic, with the genera *Epeoloides* and *Parepeolus* individually appearing

distinct from the type genus *Osiris*, which clusters with *Osirinus* and *Protosiris*. For this reason, we propose the elevation of the genus *Epeoloides* to the tribe Epeoloidini Linsley and Michener (1939), and the elevation of *Parepeolus* to the new tribe Parepeolini (see below).

Within the nomadine line, the tribe Epeolini shows similar genus-level relationships to those recovered by Onuferko et al. (2019), with the additional recovery of a sister relationship between the genera *Odyneropsis* and *Rhogepeolus*. The monophyly of the tribe Neolarrini *sensu* Bossert et al. (2020) is also recovered in this study, subsuming the former tribes Biastini and Townsendiellini. Our topology for the tribe Ericrocidini differs slightly from that of Martins et al. (2018), which recovered *Ericrocis* as the sister genus to all other ericrocidines, instead of *Mesoplia* as in the present study.

Finally, two genera (*Brachynomada* and *Ammobates*) are recovered as paraphyletic in our phylogeny due to nested members of the genera *Paranomada* and *Oreopasites* respectively, though in the latter case the node in question is poorly supported, and *Ammobates* was recovered as monophyletic in some analyses.

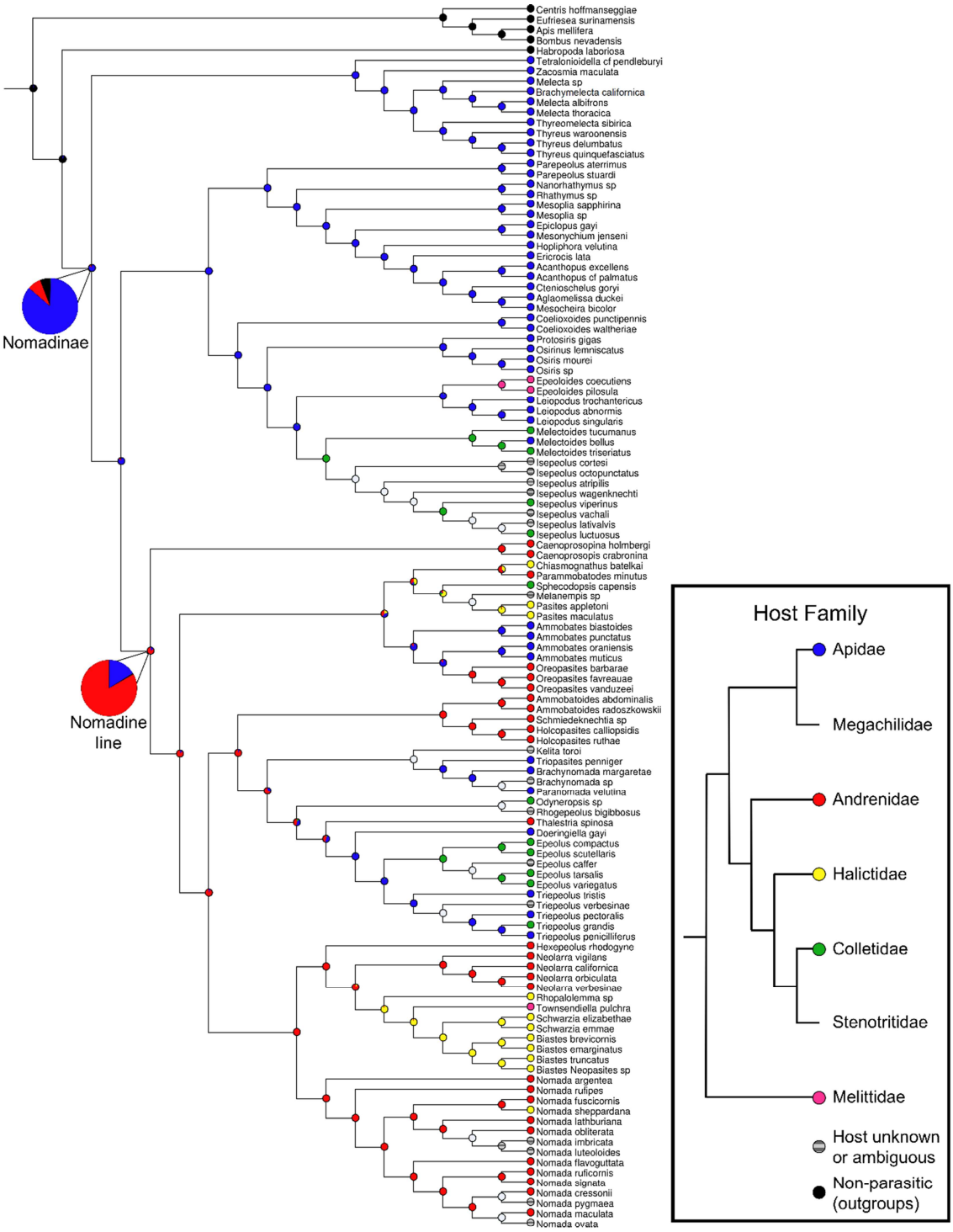
Nomadinae followed Emery's rule during initial origins of brood parasitism

The question of how host preferences of the Nomadinae have evolved over time has not been examined in detail previously, and the application of ancestral state reconstruction techniques to the

phylogeny generated here provides some interesting new insights in this respect. The rule suggested by Emery (1909) holds that certain types of parasitic insects should be closely related (or even sister clades) to their host taxa. In the case of the Nomadinae, at least the "loose" form of this rule does appear to hold true when considering the context in which parasitism first evolved. The reconstructed ancestral hosts for Nomadinae are other members of family Apidae, with a calculated proportional likelihood of 86.2% (Figure 2.2), and it was not until long after the initial transition to parasitism that nomadines began branching out to more distantly related hosts.

Figure 2.2: Cladogram showing ancestral state reconstruction of nomadine host preferences based on the concatenated, SWSC-EN partitioned phylogeny. Tips are colored according to known host families. Blue = Apidae, green = Colletidae, red = Andrenidae, yellow = Halictidae, pink = Melittidae, black = non-parasitic (outgroups), gray = unknown/polymorphic. Ancestral states at each node are shown by pie charts of proportional likelihood value for each state. Two of these, at the nodes ancestral to all Nomadinae and the nomadine line respectively, are enlarged for clarity. Inset: phylogeny of host bee families, following Danforth et al. (2011b).

Figure 2.2:


Centris hoffmanseggiae
Eufriesea surinamensis
Apis mellifera
Bombus nevadensis
Habropoda laboriosa
Tetralonioidella cf pendleburyi
Zacosmia maculata
Melecta sp
Brachymelecta californica
Melecta albifrons
Melecta thoracica
Thyreomelecta sibirica
Thyreus waroonensis
Thyreus delumbatus
Thyreus quinquefasciatus
Parepeolus aterrimus
Parepeolus stuardi
Nanorhathymus sp
Rhathymus sp
Mesoplia sapphirina
Mesoplia sp
Epiclopus gayi
Mesonychium jenseni
Hopliphora velutina
Ericrocis lata
Acanthopus excellens
Acanthopus cf palmatus
Ctenioschelus goryi
Aglaomelissa duckei
Mesocheira bicolor
Coelioxoides punctipennis
Coelioxoides waltheriae
Protosiris gigas
Osirinus lemniscatus
Osiris mourei
Osiris sp
Epeoloides coecutiens
Epeoloides pilosula
Leiopodus trochantericus
Leiopodus abnormis
Leiopodus singularis
Melectoides tucumanus
Melectoides bellus
Melectoides triseriatus
Isepeolus cortesi
Isepeolus octopunctatus
Isepeolus atripilis
Isepeolus wagenknechti
Isepeolus viperinus
Isepeolus vachali
Isepeolus lativalvis
Isepeolus luctuosus
Caenoprosopina holmbergi
Caenoprosopis crabronina
Chiasmognathus batelkai
Parammobatodes minutus
Sphecodopsis capensis
Melanempis sp
Pasites appletoni
Pasites maculatus
Ammobates biastoides
Ammobates punctatus
Ammobates oraniensis
Ammobates muticus
Oreopasites barbarae
Oreopasites favreauae
Oreopasites vanduzeei
Ammobatoides abdominalis
Ammobatoides radoszkowskii
Schmiedeknechtia sp
Holcopasites calliopsidis
Holcopasites ruthae
Kelita toroi
Triopasites penniger
Brachynomada margaretae
Brachynomada sp
Paranomada velutina
Odyneropsis sp
Rhogepeolus bigibbosus
Thalestria spinosa
Doeringiella gayi
Epeolus compactus
Epeolus scutellaris
Epeolus caffer
Epeolus tarsalis
Epeolus variegatus
Triepeolus tristis
Triepeolus verbesinae
Triepeolus pectoralis
Triepeolus grandis
Triepeolus penicilliferus
Hexepeolus rhodogyne
Neolarra vigilans
Neolarra californica
Neolarra orbiculata
Neolarra verbesinae
Rhopalolemma sp
Townsendiella pulchra
Schwarzia elizabethae
Schwarzia emmae
Biastes brevicornis
Biastes emarginatus
Biastes truncatus
Biastes Neopasites sp
Nomada argentea
Nomada rufipes
Nomada fuscicornis
Nomada sheppardana
Nomada lathburiana
Nomada obliterata
Nomada imbricata
Nomada luteoloides
Nomada flavoguttata
Nomada ruficornis
Nomada signata
Nomada cressonii
Nomada pygmaea
Nomada maculata
Nomada ovata
Nomadinae
Nomadine line
Host Family
Apidae
Megachilidae
Andrenidae
Halictidae
Colletidae
Stenotritidae
Melittidae
Host unknown or ambiguous
Non-parasitic (outgroups)

Similarly to the Nomadinae as a whole, both the tribe Melectini and the ericrocidine line are recovered as ancestrally attacking other members of Apidae. The former does not contain any transitions away from this state, but within the ericrocidine line, there are host switching events in the genus *Epeoloides* to attacking Melittidae, and in the tribe Isepeolini, which predominantly attacks members of Colletidae. The nomadine line as a whole experienced a transition in host preference at its origin, with members of the family Andrenidae recovered as its ancestral hosts (prop. likelihood 83.3%). Within the nomadine line, the initial host preferences of some clades are unclear; both the tribes Ammobatini and Epeolini are somewhat equivocal at their ancestral nodes and experience multiple transition events internally (variously to hosts of the families Halictidae, Colletidae, Andrenidae, or Apidae, with some reversals likely). There is also a clear transition from parasitism of Andrenidae to Halictidae within the tribe Neolarrini (*sensu* Bossert et al. 2020).

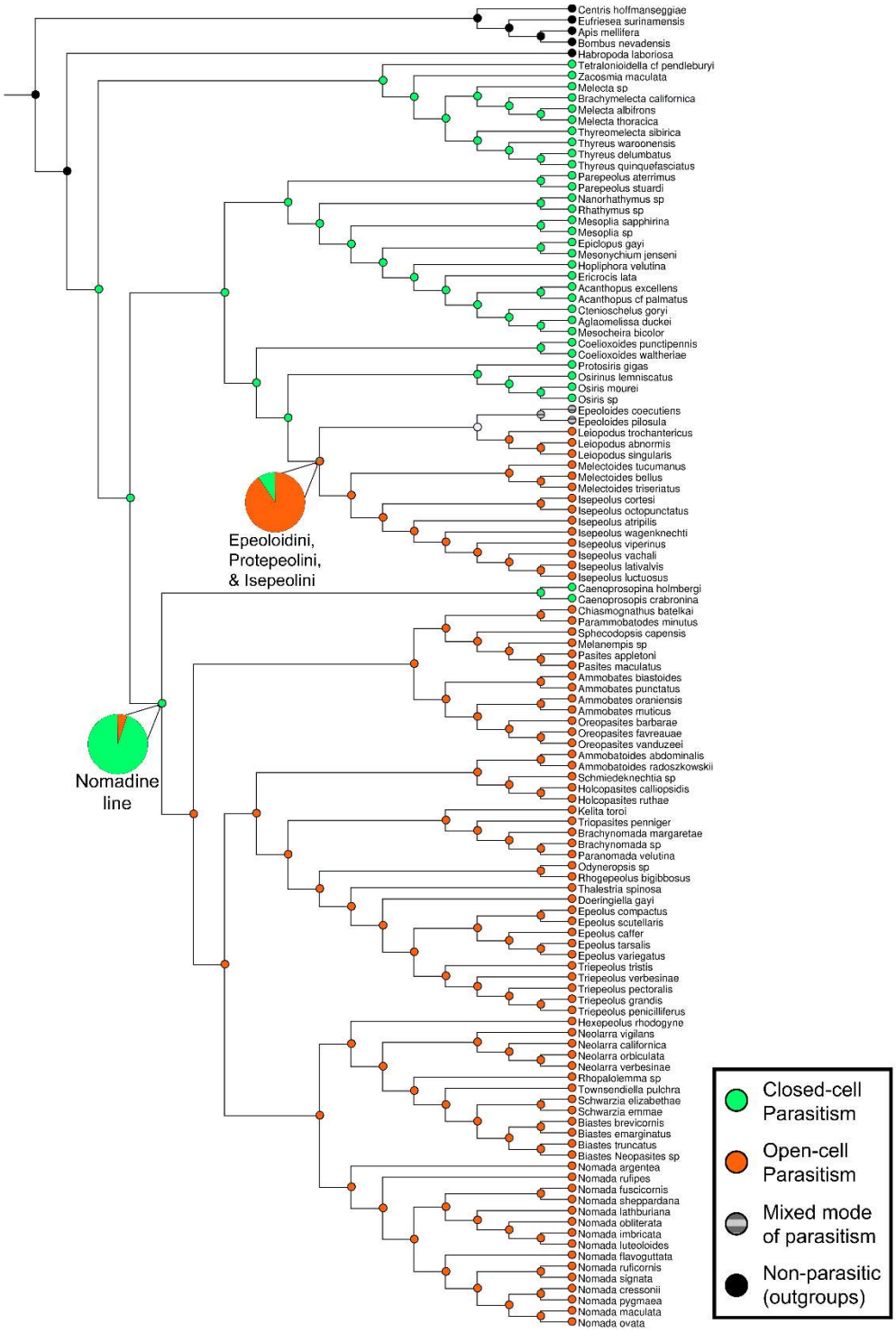
Closed-cell parasitism ancestral, with two transitions to open-cell parasitism

We recovered closed-cell parasitism as the most likely ancestral mode which was adopted during the initial transition to brood parasitism (proportional likelihood 99.4%; Figure 2.3). This was subsequently followed by two transitions to open-cell parasitism: one in the subsection of the nomadine line that forms the sister group to the tribe Caenoprosopidini (prop. likelihood 98.6%), and another ancestral to the tribes Isepeolini, Protepeolini, and Epeoloidini (prop.

likelihood 90.7%). The somewhat ambiguous mode of parasitism exhibited by *Epeoloides* may represent a partial reversion to a strategy employing some characteristics of closed-cell parasitism, however the uncertainty of this node complicates the reconstruction of this trait. Some of our other phylogenetic analyses instead recovered Epeoloidini as the sister group to both Isepeolini and Protepeolini, in which case the strategy observed in *Epeoloides* may be a transitional state between closed-cell parasitism and the "fully" open-cell mode of parasitism suggested for the latter two tribes.

Figure 2.3: Cladogram showing ancestral state reconstruction of mode of parasitism based on the concatenated, SWSC-EN partitioned phylogeny. Tips are colored according to known or suspected mode of parasitism. Green = closed-cell parasitism, orange = open-cell parasitism, white = non-parasitic (outgroups), gray = ambiguous mode of parasitism. Ancestral states at each node are shown by pie charts of likelihood value for each state. Two of these, at the nodes ancestral to the nomadine line and to the clade consisting of Epeoloidini, Protepeolini, and Isepeolini, are enlarged for clarity.

Figure 2.3:


Centris hoffmanseggiae
Eufriesea surinamensis
Apis mellifera
Bombus nevadensis
Habropoda laboriosa
Tetralonioidella cf pendleburyi
Zacosmia maculata
Melecta sp
Brachymelecta californica
Melecta albifrons
Melecta thoracica
Thyreomelecta sibirica
Thyreus waroonensis
Thyreus delumbatus
Thyreus quinquefasciatus
Parepeolus aterrimus
Parepeolus stuardi
Nanorhathymus sp
Rhathymus sp
Mesoplia sapphirina
Mesoplia sp
Epiclopus gayi
Mesonychium jenseni
Hopliphora velutina
Ericrocis lata
Acanthopus excellens
Acanthopus cf palmatus
Ctenioschelus goryi
Aglaomelissa duckei
Mesocheira bicolor
Coelioxoides punctipennis
Coelioxoides waltheriae
Protosiris gigas
Osirinus lemniscatus
Osiris mourei
Osiris sp
Epeoloides coecutiens
Epeoloides pilosula
Leiopodus trochantericus
Leiopodus abnormis
Leiopodus singularis
Melectoides tucumanus
Melectoides bellus
Melectoides triseriatus
Isepeolus cortesi
Isepeolus octopunctatus
Isepeolus atripilis
Isepeolus wagenknechti
Isepeolus viperinus
Isepeolus vachali
Isepeolus lativalvis
Isepeolus luctuosus
Caenoprosopina holmbergi
Caenoprosopis crabronina
Chiasmognathus batelkai
Parammobatodes minutus
Sphecodopsis capensis
Melanempis sp
Pasites appletoni
Pasites maculatus
Ammobates biastoides
Ammobates punctatus
Ammobates oraniensis
Ammobates muticus
Oreopasites barbarae
Oreopasites favreauae
Oreopasites vanduzeei
Ammobatoides abdominalis
Ammobatoides radoszkowskii
Schmiedeknechtia sp
Holcopasites calliopsidis
Holcopasites ruthae
Kelita toroi
Triopasites penniger
Brachynomada margaretae
Brachynomada sp
Paranomada velutina
Odyneropsis sp
Rhogepeolus bigibbosus
Thalestria spinosa
Doeringiella gayi
Epeolus compactus
Epeolus scutellaris
Epeolus caffer
Epeolus tarsalis
Epeolus variegatus
Triepeolus tristis
Triepeolus verbesinae
Triepeolus pectoralis
Triepeolus grandis
Triepeolus penicilliferus
Hexepeolus rhodogyne
Neolarra vigilans
Neolarra californica
Neolarra orbiculata
Neolarra verbesinae
Rhopalolemma sp
Townsendiella pulchra
Schwarzia elizabethae
Schwarzia emmae
Biastes brevicornis
Biastes emarginatus
Biastes truncatus
Biastes Neopasites sp
Nomada argentea
Nomada rufipes
Nomada fuscicornis
Nomada sheppardana
Nomada lathburiana
Nomada obliterata
Nomada imbricata
Nomada luteoloides
Nomada flavoguttata
Nomada ruficornis
Nomada signata
Nomada cressonii
Nomada pygmaea
Nomada maculata
Nomada ovata
Epeoloidini,
Protepeolini,
& Isepeolini
Nomadine
line
Closed-cell
Parasitism
Open-cell
Parasitism
Mixed mode
of parasitism
Non-parasitic
(outgroups)

Nomenclature of Nomadinae

In light of the polyphyletic status of the tribe Osirini Handlirsch, we divide the five genera represented therein into three tribes. The tribe Osirini is reduced to the three genera *Osiris*, *Osirinus*, and *Protosiris*. Additionally, we reinstate the tribe Epeoloidini Linsley and Michener (1939) containing the genus *Epeoloides*, and describe the following new tribe:

Parepeolini, Straka and Sless, *trib. nov.*

Type genus: *Parepeolus* Ducke (1912)

Diagnosis: This tribe includes epeoliform cuckoo bees with three submarginal cells from the former Osirini tribe. As in Osirini and Epeoloidini, they have a ventral neck sclerite, a carina along the inner and basal margins of the forecoxa, a very large stigma (several times larger than prestigma), and a marginal cell distinctly separated from the wing margin. They also share mouthparts typical for most Apidae, but absent in the nomadine line, including a ridge on the outer surface of the stipes and a more or less anteroventrally emarginated stipes with comb. The posterior margin of the metasternum is translucent and impunctate. The tribe is differentiated from other Nomadinae by the additional combination of the following characters. Jugal lobe of the hind wing is rounded. Parocular carina is reduced to completely missing. Labrum with paired tubercules. Sternum 6 of female with longitudinal ridge medially. Gonostyli of male genitalia large and complex in structure, ventral gonostylus bifid, and dorsal gonostylus large and flattened.

Discussion

Comparison to other phylogenies

Several previous phylogenetic studies have included at least some nomadine representatives, though most of these used datasets consisting of a small number of highly conserved protein-coding or ribosomal genes. Specifically, Cardinal et al. (2010, 2018), and Policarová et al. (2019) used wingless (wg), RNA polymerase II (pol II), long-wavelength rhodopsin (LWR), sodium-potassium ATPase (NaK), elongation factor 1 alpha (EF-1α), and both 18S and 28S rRNAs, resulting in a total alignment length of approximately 7500 bp. Litman et al. (2013) also used NaK, LWR, EF-1α, and 28S rRNA, but added carbamoyl phosphate synthetase 2 (CAD) in place of the other genes, giving an alignment length of approximately 6000 bp. Meanwhile, Bossert et al. (2019) used two UCE datasets with 129 and 561 loci respectively, resulting in alignments of about 79,293 and 302,379 bp. In contrast, the 95% taxon-completeness UCE dataset in the present study consists of 1247 loci and a final alignment length of 366,640 bp. The difference in average UCE locus length between Bossert et al.'s study and ours may be due in part to the extraction of UCEs from genomes with large (up to 3200 bp) flanking regions and more relaxed parameters during trimming with Gblocks by Bossert et al., or simply a consequence of older average specimen age in the present study. In terms of taxon sampling, these previous studies included 63 (Cardinal et al. 2010), 32 (Litman et al. 2013), 44 (Cardinal et al. 2018), 35 (Policarová et al. 2019), and 12 (Bossert et al. 2019) representative

species from the broader Nomadinae respectively, compared to the 114 species included in the present study. We believe that our substantially greater taxon sampling provides more reliability to our data and gives us confidence with respect to instances of variance in our taxonomy in relation to previous studies (see below).

Overall, the topology recovered in the present study has many similarities with these earlier ones, though it is not identical to any of them (Table 2.1). For example, Cardinal et al. (2010) and Policarová et al. (2019) recovered the tribe Melectini as the sister group to the nomadine line, rather than the sister to all other Nomadinae (Table 2.1, row 2). The tribe Protepeolini is placed as the sister group to Epeoloidini in our main analysis (Table 2.1, row 6), in agreement with Cardinal et al. (2010, 2018) and Litman et al. (2013), but contrary to its position as the sister to Isepeolini in Policarová et al. 2019. However, some of our other analyses did recover a sister relationship between Protepeolini and Isepeolini to the exclusion of Epeoloidini (Appendix III Figures 2.1, 2.5), making the placement of this group somewhat problematic. Another part of the former Osirini, the tribe Parepeolini newly described herein, is uniquely recovered in our analysis as sister to Rhathymini + Ericrocidini, whereas previous studies (Cardinal et al. 2010 and 2018, Litman et al. 2013) have often recovered it as the sister group to Coelioxoidini (Table 2.1, row 5). This latter group is itself inconsistently placed across past studies, and also appears to be of uncertain placement in our analyses, appearing as by

far the longest branch in our phylogeny and changing position with different methodologies (Appendix III Figures 2.2, 2.3, & 2.5).

We recover the enigmatic tribe Caenoprosopidini (Table 2.1, row 11) as the sister group to the rest of the nomadine line, similarly to Policarová et al. (2019), while other studies have instead considered it the sister to Ammobatini (Cardinal et al. 2010) or Melectini (Cardinal et al. 2018). Additionally, the present study is the only one so far to recover the tribe Ammobatoidini as the sister group to Epeolini + Brachynomadini, while other studies have more commonly recovered Ammobatini as the sister to the Neolarrini (*sensu* Bossert et al. 2020) + Hexepeolini group (Table 2.1, row 14). This Neolarrini + Hexepeolini clade also includes some of the lowest gene and site concordance factors across the phylogeny (Appendix III Figure 2.6), which may be a result of the short internode lengths separating these species, but the broader topology of this tree is generally supported by these metrics.

Table 2.1: Comparison of clades recovered in this and previous phylogenetic studies. Presence of a monophyletic group composed of the listed clades is indicated with a "Y", absence with an "N". Cases where a study had insufficient taxon sampling to identify a clade as present or absent are indicated with "n/a". Clades labelled with an asterisk were recovered in our primary SWSC-EN partitioned phylogeny but were absent from at least one other analysis in the present study.

	Group recovered as monophyletic?	**This study**	**Cardinal et al. 2010**	**Litman et al. 2013**	**Cardinal et al. 2018**	**Policarová et al. 2019**	**Bossert et al. 2019**
	Total # nomadine species	114	63	32	44	35	12
	Dataset type	UCEs	7 genes	5 genes	7 genes	7 genes/ morphology	UCEs/ transcriptomes
1	Single large parasitic clade [2-16]	Y	Y	Y	Y	Y	Y
2	All Nomadinae excluding Melectini [3-16]	Y	N	Y	N	N	Y
3	"Ericrocidine line" [4-9]	Y	Y	Y	Y	Y	Y
4	Rhathymini + Ericrocidini	Y	Y	Y	Y	Y	Y
5	Parepeolini + [4]	Y	N	N	N	N	n/a
6	Protepeolini + Epeoloidini	Y*	Y	Y	Y	N	n/a
7	Isepeolini + [6]	Y*	Y	Y	Y	N	n/a
8	Osirini + [7]	Y	Y	Y	Y	Y	n/a
9	Coelioxoidini + [8]	Y*	N	N	N	N	n/a
10	"Nomadine line" [Nomadinae *sensu stricto*, 11-16]	Y	Y	Y	Y	Y	Y
11	Nomadine line excluding Caenoprosopidini [12-16]	Y	N	n/a	N	Y	n/a
12	Remaining Nomadine line excluding Ammobatini [13-16]	Y	N	Y	Y	Y	N
13	Epeolini + Brachynomadini	Y	Y	Y	Y	Y	Y
14	Ammobatoidini + [13]	Y	N	N	N	N	n/a
15	Neolarrini *sensu* Bossert et al. (2020) + Hexepeolini	Y	Y	Y	Y	Y	n/a
16	Nomadini + [15]	Y	Y	N	N	N	n/a

Taxonomic implications

The findings of this study have several important implications for the classification of the subfamily Nomadinae and its component subclades. As has been suggested by previous literature, there is a clear signal of a large, entirely parasitic clade consisting of both the Nomadinae *sensu stricto* (our nomadine line) and several other tribes of apid brood parasites. Previously referred to as the "large cleptoparasitic clade" by Cardinal et al. (2010), this group contains all brood parasitic members of the family Apidae with the exceptions of the orchid bee genera *Aglae* and *Exaerete* (Euglossini) and the genus *Ctenoplectrina* (Ctenoplectrini). Due to the inclusion of several lineages traditionally classified in the subfamily Apinae, viz. the tribes Melectini, Rhathymini, Ericrocidini, Coelioxoidini, Osirini, Protepeolini, and Isepeolini as well as the resurrected Epeoloidini and newly named Parepeolini, this large parasitic clade thus renders the Apinae paraphyletic. For this reason, we support the recommendation by Bossert et al. (2019) to revise the definition of the subfamily Nomadinae such that it includes all of the above-mentioned tribes in addition to those that currently form the nomadine line (Nomadinae *s.s.*). Correspondingly, Apinae should be redefined to exclude these taxa and should therefore only include the tribes Centridini, Euglossini, Apini, Bombini, and Meliponini, again following Bossert et al. (2019). We also find support for the reclassification of Neolarrini proposed by Bossert et al. (2020) to include Biastini and Townsendiellini, though that study shares some data with the present one and thus these findings are not fully independent.

Considering the broader internal classification of the Nomadinae, we recover three major lineages which make up the group. The tribe Melectini on its own is one descendant of the earliest branching event within Nomadinae. The other branch then further splits into the "ericrocidine" and "nomadine" lines. The ericrocidine line includes the tribes Rhathymini, Ericrocidini, Isepeolini, Protepeolini, Coelioxoidini, and all lineages previously included in Osirini (see below: Osirini, Epeoloidini, and Parepeolini *trib. nov.*). Such a grouping has also been suggested by previous studies (Litman et al. 2013; Martins et al. 2018). This group is tentatively united by the absence of the epistomal suture past the anterior tentorial pits in adult bees (Martins et al. 2018). Meanwhile, the nomadine line (former Nomadinae *sensu stricto*) includes the tribes Ammobatini, Ammobatoidini, Epeolini, Brachynomadini, Caenoprosopidini, Hexepeolini, Nomadini, and Neolarrini *sensu* Bossert et al. 2020 (i.e., Neolarrini along with the former tribes Biastini and Townsendiellini). From a biogeographic perspective, the ericrocidine line is almost entirely Neotropical in distribution, with the exception of a few species that reach the extreme southern Nearctic and the genus *Epeoloides*, found in both the Palearctic and Nearctic realms. The nomadine line, in contrast, is cosmopolitan in distribution but generally most diverse in the Holarctic region, with only a few *Nomada* species reaching the Australasian realm (Michener 2007).

Additionally, the continued recognition of one of the tribes is not supported by our analyses. Our phylogeny, the first to include representatives of all five genera traditionally included in the tribe Osirini Handlirsch, finds the relationship among these to be polyphyletic. To clarify the nomenclature of this group, we propose to include within the tribe Osirini only the three genera which do form a clade in our analyses, namely *Osiris*, *Protosiris*, and *Osirinus*. Representatives of the tribe Osirini *sensu stricto* are described in detail by Roig-Alsina (1989). Meanwhile, the genus *Parepeolus* is recovered as the sister group to Ericrocidini + Rhathymini. Due to a lack of morphological characters uniting *Parepeolus* with Ericrocidini and Rhathymini, we diagnose the new tribe Parepeolini. The erstwhile osirine genus *Epeoloides* is recovered with a weakly-supported sister relationship to Protepeolini, but similarly could not be included with this tribe. Thus, we suggest that the previously proposed tribe Epeoloidini Linsley and Michener (1939) should be used for this genus. Our analyses also support previous suggestions that the enigmatic genus *Coelioxoides* should not be considered a close relative of its host *Tetrapedia*, but rather should be placed into its own tribe of Coelioxoidini (Martins et al. 2018; Engel et al. 2020). While it might be optimal to include these newly recognized small tribes within larger clades, this solution is problematic due to the divergent morphology of the aforementioned genera from their sister groups, and so we propose the above solution as a more stable one.

Host preferences and Emery's rule

The ability to map traits associated with parasitism onto a phylogenetic tree of the Nomadinae with unprecedented resolution in turn allows for more detailed investigation of the evolutionary dynamics which drive the evolution and diversification of such brood parasitic groups. Perhaps the most interesting question of this type relates to how host preferences are determined, and how they change over time. In his 1909 paper, Carlo Emery noted that social and brood parasites have a tendency to attack closely related species, and this principle has come to be known as Emery's rule. Other parasitic bees, such as those in the families Halictidae and Megachilidae as well as the apine tribe Euglossini, typically follow this rule (Michener, 2007). However, the age and diversity of the Nomadinae compared to other brood parasitic clades has been a complicating factor in evaluating their adherence to this concept. The use of ancestral state reconstruction techniques, combined with the broad taxon sampling utilized in this study, provide the first chance to look back in time to the origins of parasitism in the Nomadinae, estimated to have occurred approximately 100 million years ago (Litman et al. 2013).

As these analyses show, it appears that nomadine bees do follow Emery's rule when considering their initial transition to parasitism. The ancestral hosts for the earliest nomadines are indeed recovered to be other members of the family Apidae. Furthermore, the tribe Melectini, which forms the sister group to all other Nomadinae, are entirely parasitic on members of the apid subfamily Anthophorinae, which is recovered as the sister group to Nomadinae as a whole in

both this and other studies (Cardinal et al. 2010, Policarová et al. 2019, Bossert et al. 2019). This suggests the possibility that not only the loose but also the strict form of Emery's rule (requiring a direct sister relationship between host and parasite) may have been the case during the origins of this group. As time went on and the Nomadinae diversified and grew more speciose, they evolved to attack a more diverse range of hosts, which at present span five families of bees. Unfortunately, sufficiently detailed phylogenies for all of these host groups do not exist, so that a comprehensive co-evolutionary analysis is not currently possible. Though difficult to quantify, there does appear to be a general trend towards increased host diversity demonstrated by more recent parasitic clades. However, it is still unclear whether this trend may be due to an increased number of host-switching events, or simply increased generalization in host use at the level of particular parasitic groups.

The applicability of Emery's rule to the Nomadinae presents an important contrast with some other brood parasitic and social parasitic insects, where the rule has been considered in relation to much more recent origins of parasitism in smaller taxonomic groups. For example, Smith et al. (2013) and Sumner et al. (2004) both report that Emery's rule in the loose sense is broadly observed in socially parasitic allodapine bees and leafcutter ants respectively, though the exact sister relationships expected under the strict form of the rule are not always seen. Conversely, Lopez-Osorio et al. (2015) fail to find support for this principle at all in socially parasitic vespine wasps. As

for other lineages of bees outside the Apidae, there are several examples of apparently recent origins of brood parasitism with a close relationship between parasite and host, sometimes within a single genus. The family Halictidae in particular features several examples, including certain species of *Lasioglossum* (*Dialictus*), *Megalopta* (*Noctoraptor*), and *Megommation* (*Cleptommation*) all parasitic on congeneric species, as well as parasitic representatives of *Parathrincostoma* and *Temnosoma* which attack other members of their respective tribes (*Thrinchostoma* and other Augochlorini; Michener 2007).

Some nomadines have remained fairly specialized on a narrow range of hosts. In some cases, there are clear biological factors which may account for this, such as parasites in the tribes Ericrocidini, Rhathymini, and Osirini which target hosts that are themselves specialized in collecting floral oils as food resources (Martins et al. 2018, Policarová et al. 2019). Other genera of brood parasites have become much wider generalists, attacking diverse host groups. For example, while the genus *Nomada* is most commonly associated with mining bee hosts in the genus *Andrena*, it has been recorded parasitizing members of five different families of bees (Michener 2007, Westrich 1989). However, this level of generalization is not necessarily unique among brood parasitic bees; outside the Nomadinae, brood parasites of the genus *Sphecodes* have even been shown to display individual differences among their preferred hosts (Bogusch et al.

2006) as well as flexible host switches over evolutionary time (Habermannová et al. 2013).

It is interesting to note that, while the five families of bees attacked by nomadines represent a substantial diversity of hosts, there is a large group of species which remain as a potential, yet unused, resource. To our knowledge, no records exist of any nomadine species attacking a member of the family Megachilidae as a host, despite the fact that this group contains over 4,000 species and is widely distributed. The reasons for this can only be speculated on, though the existence of several brood parasitic megachilids which attack other members of their family (including *Coelioxys*, *Radoszkowskiana*, the *Stelis* group, the tribe Dioxyini, and certain species of *Hoplitis*; Litman et al. 2013) rules out any advanced defenses that make these bees immune to parasitism entirely. The most obvious potential explanation seems to be the diversity of nesting strategies employed by members of the Megachilidae, which includes the use of cavities in wood, stone, or even snail shells as well as a variety of structures created from mud or plant material (Danforth et al. 2019). In contrast, almost all hosts of Nomadinae are ground-nesting, though the existence of several ground-nesting megachilids leaves the question somewhat open.

Mode of parasitism

In addition to studying the historical patterns in changing host preferences, the phylogeny presented in this study also allows for

research into changes in different strategies and forms of brood parasitism. Michener (2007) outlines some general modes of parasitism, which were further expanded by Litman et al. (2013). In some parasites, the host larvae or eggs are killed by the adult female parasite before her own eggs are laid, though this strategy is not found in any members of the Nomadinae, where parasitic larvae kill their nestmates instead. The main dichotomy in this group exists between so-called "closed-cell" parasites – those which invade a nest after it has been completed and close the cell themselves – and "open-cell" parasites, which infiltrate a nest that is still under construction and leave it open for the host to complete.

In line with the findings of Litman et al. (2013), we recover the closed-cell mode of brood parasitism as ancestral for Nomadinae, with two transitions to open-cell parasitism. However, in contrast to this study, we recover the tribe Caenoprosopidini as the sister group to all other members of the nomadine line. Though the tribe is poorly studied, the description of possible egg insertion holes by Rozen and Roig-Alsina (1991) suggests that they are closed-cell parasites. Thus, this origin of open-cell parasitism in our study is detected at a slightly later date, and in fewer taxa, than in Litman et al. (2013).

The other transition to open-cell parasitism in both the present study and Litman et al. occurs in the common ancestor of the tribes Isepeolini, Protepeolini, and Epeoloidini. However, the genus *Epeoloides* has been observed to use a strategy that does not neatly fit

either the open-cell or closed-cell modes. As described by Straka and Bogusch (2007), *Epeoloides coecutiens* females were observed entering nests of their *Macropis* hosts that were still being provisioned, like open-cell parasites, yet closing them afterwards in the same way as closed-cell parasites. This combination of behaviors may represent a partial reversion to closed-cell parasitism, or perhaps a derived but transitional state between obligate open- and closed-cell modes. The somewhat poorly resolved location of Epeoloidini in our phylogenetic analyses further complicates the interpretation of this trait. Such cases demonstrate the importance of field observations of the invasion behaviors of brood parasites to further verify and record the strategies used by lesser-studied groups.

Comparisons of host preference in concert with mode of parasitism also reveal some noteworthy patterns. Both transitions from closed-cell to open-cell parasitism in our phylogeny occur within one node of inferred shifts in host preference: from Apidae to Colletidae in the Isepeolini, and from Apidae to Andrenidae in the nomadine line. Indeed, with the exception of the poorly known tribe Caenoprosopidini, all closed-cell parasites in our phylogeny appear restricted to the use of confamilial hosts (i.e., other species of Apidae). This suggests that the open-cell strategy allows for greater evolutionary lability in host preferences and could in part account for the much greater diversity of hosts utilized by the nomadine line in contrast to the ericrocidine line and tribe Melectini, though it is difficult to speculate as to the exact mechanism by which this might

be achieved. The existence of many open-cell parasites within the nomadine line which have reverted to attacking apids would seem to preclude the possibility that these hosts are simply better at defending against this mode of parasitism. Perhaps the open-cell strategy instead allows for the circumvention of host defenses used by some non-apids that only become effective upon completion of nest cells, such as the complex glandular secretions of many colletids (Almeida 2008). This pattern may also relate to Emery's rule in the sense that the ability to locate, enter, and reseal a finished nest as in closed-cell parasitism is potentially a more straightforward strategy against close relatives with similar nest architecture to a brood parasite's most recent free-living ancestors. As already expressed, however, further investigation into the nesting biology of brood parasites and their hosts is essential to the continued exploration of such questions.

Conclusion

This study provides the first comprehensive analysis of the phylogeny of the oldest and largest clade of brood parasitic bees, the subfamily Nomadinae in the broad sense. While the specific relationships among most of the members of this group are consistent with findings from previous research, there are some notable differences. Additionally, the unprecedented level of taxon sampling included herein has allowed us to explore novel questions related to the evolution of brood parasitism. Emery's rule is supported, at least in the loose sense, by ancestral state reconstruction of other members of the family Apidae as the earliest hosts of nomadine parasites. Similarly, these techniques shed

light on the most likely behavioral strategies used by the first brood parasitic bees over 100 million years ago, with closed-cell parasitism inferred to be ancestral, followed by the independent evolution of open-cell parasitism in two lineages. Finally, this study also amends the nomenclature of subgroups within the Nomadinae to remove poly- and paraphyletic taxa by partitioning the tribe Osirini into three tribes.

REFERENCES

Alexander, B. (1991). *Nomada* phylogeny reconsidered (Hymenoptera: Anthophoridae). *Journal of Natural History*, 25(2), 315-330. https://doi.org/10.1080/00222939100770231

Almeida, E. A. B. (2008). Colletidae nesting biology (Hymenoptera: Apoidea). *Apidologie*, 39(1), 16-29. https://doi.org/10.1051/apido:2007049

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ascher, J. S. & Pickering, J. (2020). Discover life bee species guide and world checklist (Hymenoptera: Apoidea: Anthophila). Available at http://www.discoverlife.org/mp/20q?guide=Apoidea_species

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5), 455-477. https://doi.org/10.1089/cmb.2012.0021

Blumenstiel, B., Cibulskis, K., Fisher, S., DeFelice, M., Barry, A., Fennell, T., Abreu, J., Minie, B., Costello, M., Young, G., Maquire, J., Kernytsky, A., Melnikov, A., Rogov, P., Gnirke, A., & Gabriel, S. (2010) Targeted exon sequencing by in-solution hybrid selection. *Current Protocols in Human Genetics*, Chapter 18, Unit 18.4. https://doi.org/10.1002/0471142905.hg1804s66

Bogusch, P., Kratochvíl, L., & Straka, J. (2006). Generalist cuckoo bees (Hymenoptera: Apoidea: Sphecodes) are species-specialist at the individual level. *Behavioral Ecology and Sociobiology*, 60(3), 422-429. https://doi.org/10.1007/s00265-006-0182-4

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Bossert, S., Murray, E. A., Almeida, E. A. B., Brady, S. G., Blaimer, B. B., & Danforth, B. N. (2019). Combining transcriptomes and

ultraconserved elements to illuminate the phylogeny of Apidae. *Molecular Phylogenetics and Evolution*, 130, 121-131. https://doi.org/10.1016/j.ympev.2018.10.012

Bossert, S., Copeland, R. S., Sless, T. J. L., Branstetter, M. G., Gillung, J. P., Brady, S. G., Danforth, B. N., Policarová, J. & Straka, J. (2020). Phylogenomic and morphological reevaluation of the bee tribes Biastini, Neolarrini, and Townsendiellini (Hymenoptera: Apidae) with description of three new species of *Schwarzia*. *Insect Systematics and Diversity*, 4(6). https://doi.org/10.1093/isd/ixaa013

Branstetter, M. G., Longino, J. T., Ward, P. S., & Faircloth, B. C. (2017). Enriching the ant tree of life: enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. *Methods in Ecology and Evolution*, 8(6), 768-776. https://doi.org/10.1111/2041-210X.12742

Cardinal, S., Straka, J., & Danforth, B. N. (2010). Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. *Proceedings of the National Academy of Sciences USA*, 107(37), 16207-16211. https://doi.org/10.1073/pnas.1006299107

Cardinal, S., Buchmann, S. L., & Russell, A. L. (2018). The evolution of floral sonication, a pollen foraging behavior used by bees (Anthophila). *Evolution*, 72(3), 590-600. https://doi.org/10.1111/evo.13446

Chernomor, O., von Haeseler, A., & Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*, 65(6), 997–1008. https://doi.org/10.1093/sysbio/syw037

Danforth, B. N., Fang, J., Sipes, S., Brady, S. G., Almeida, E. A. B., Litman, J. R., & Cardinal, S. (2011a). Phylogeny and molecular systematics of bees (Hymenoptera: Apoidea). Cornell University, Ithaca, NY. Available from http://www.danforthlab.entomology.cornell.edu/research/resources/

Danforth, B. N., Cardinal, S., Praz, C., Almeida, E. A. B., & Michez, D. (2011b). The impact of molecular data on our understanding of bee phylogeny and evolution. *Annual Review of Entomology*, 58(1), 57-78. https://doi.org/10.1146/annurev-ento-120811-

153633

Danforth, B., Minckley, R., & Neff, J. (2019). The Solitary Bees: Biology, Evolution, Conservation. Princeton; Oxford: Princeton University Press.

Emery, C. (1909) Über den Ursprung der dulotischen, parasitischen und myrmikophilen Ameisen. *Biologisches Centralblatt* 29, 352-362

Engel, M. S. (2005). Family-group names for bees (Hymenoptera: Apoidea). *American Museum Novitates* 2005(3476), 1-33. https://doi.org/10.1206/0003-0082(2005)476[0001:FNFBHA]2.0.CO;2

Engel, M. S., Gonzalez, V., & Rasmussen, C. (2020). An overlooked family-group name among bees: Availability of Coelioxoidini (Hymenoptera: Apidae). *Journal of Melittology*, (94), 1-3. https://doi.org/10.17161/jom.vi94.13385

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, 61(5), 717-726. https://doi.org/10.1093/sysbio/sys004

Faircloth, B. C. (2013). illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. http://dx.doi.org/10.6079/J9ILL

Faircloth, B. C., Branstetter, M. G., White, N. D., & Brady, S. G. (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Molecular Ecology Resources*, 15(3), 489-501. https://doi.org/10.1111/1755-0998.12328

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, 32(5), 786-788. https://doi.org/10.1093/bioinformatics/btv646

Faircloth, B. C. (2019). Phyluce: software for UCE (and general) phylogenomics. Last updated 8 Oct. 2019. Available at http://phyluce.readthedocs.io/en/latest/

Freitas, F. V., Branstetter, M. G., Griswold, T., & Almeida, E. A. B. (2020). Partitioned gene-tree analyses and gene-based topology testing help resolve incongruence in a phylogenomic study of host-specialist bees (Apidae: Eucerinae). *Molecular Biology and Evolution*, msaa277. https://doi.org/10.1093/molbev/msaa277

Glenn, T. C., Nilsen, R. A., Kieran, T. J., Sanders, J. G., Bayona-Vásquez, N. J., Finger, J. W., Pierson, T. W., Bentley, K. E., Hoffberg, S. L., Louha, S. et al. (2019). Adapterama I: universal stubs and primers for 384 unique dualindexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ*, 7: e7755. http://doi.org/10.7717/peerj.7755

Grab, H., Branstetter, M. G., Amon, N., Urban-Mead, K. R., Park, M. G., Gibbs, J., Blitzer, E. J., Poveda, K., Loeb, G., & Danforth, B. N. (2019). Agriculturally dominated landscapes reduce bee phylogenetic diversity and pollination services. *Science,* 363:282-284. https://doi.org/10.1126/science.aat6016

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307-321. https://doi.org/10.1093/sysbio/syq010

Habermannová, J., Bogusch, P., & Straka, J. (2013). Flexible host choice and common host switches in the evolution of generalist and specialist cuckoo bees (Anthophila: *Sphecodes*). *PLoS ONE,* 8(5), e64537. https://doi.org/10.1371/journal.pone.0064537

Harris, R. S. (2007). Improved pairwise alignment of genomic DNA. Ph.D. Dissertation, Pennsylvania State University, State College, PA. Available at https://etda.libraries.psu.edu/catalog/7971

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2), 518-522. https://doi.org/10.5281/zenodo.854445

Houston, T. (2018). A Guide to Native Bees of Australia. 1st Edition, CSIRO Publishing, Clayton South, Victoria, Australia.

Huang, M. H. & Dornhaus, A. (2008). A meta-analysis of ant social parasitism: host characteristics of different parasitism types and

a test of Emery's rule. *Ecological Entomology* 33, 589-596. https://doi.org/10.1111/j.1365-2311.2008.01005.x

Iwata, K. & Sakagami, S. F. (1966). Gigantism and dwarfism in bee eggs in relation to the modes of life, with notes on the number of ovarioles. *Japanese Journal of Ecology*, 16(1), 4-16

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587-589. https://doi.org/10.1038/nmeth.4285

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772-780. https://doi.org/10.1093/molbev/mst010

Lenoir, A., D'Ettorre, P., Errard, C., & Hefetz, A. (2001). Chemical ecology and social parasitism in ants. *Annual Review of Entomology*, 46(1), 573-599. https://doi.org/10.1146/annurev.ento.46.1.573

Linsley, E. G., & Michener, C. D. (1939). A generic revision of the North American Nomadidae (Hymenoptera). *Entomological Society,* 65(3), 265-305.

Litman, J. R., Praz, C. J., Danforth, B. N., Griswold, T. L., & Cardinal, S. (2013). Origins, evolution, and diversification of cleptoparasitic lineages in long-tongued bees. *Evolution*, 67(10), 2982-2998. https://doi.org/10.1111/evo.12161

Lopez-Osorio, F., Perrard, A., Pickett, K. M., Carpenter, J. M., & Agnarsson, I. (2015). Phylogenetic tests reject Emery's rule in the evolution of social parasitism in yellowjackets and hornets (Hymenoptera: Vespidae, Vespinae). *Royal Society Open Science*, 2: 150159. https://doi.org/10.1098/rsos.150159

Maddison, W. P. & Maddison, D. R. 2019. Mesquite: a modular system for evolutionary analysis. Version 3.61. Available at http://www.mesquiteproject.org

Martins, A. C., Luz, D. R., & Melo, G. A. R. (2018). Palaeocene origin of the Neotropical lineage of cleptoparasitic bees Ericrocidini-Rhathymini (Hymenoptera, Apidae). *Systematic Entomology*, 43(3), 510-521. https://doi.org/10.1111/syen.12286

Michener, C. D. (2007). *The Bees of the World*. 2nd Edition, John Hopkins University Press, Baltimore.

Minckley, R. L., Danforth, B. N. (2019). Sources and frequency of brood loss in solitary bees. *Apidologie*, 50, 515–525. https://doi.org/10.1007/s13592-019-00663-2

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020a). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution,* 37(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Minh, B. Q., Hahn, M. W., & Lanfear, R. (2020b). New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution*, 37(9), 2727-2733. https://doi.org/10.1093/molbev/msaa106

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268-274. https://doi.org/10.1093/molbev/msu300

Onuferko, T. M., Bogusch, P., Ferrari, R. R., & Packer, L. (2019). Phylogeny and biogeography of the cleptoparasitic bee genus *Epeolus* (Hymenoptera: Apidae) and cophylogenetic analysis with its host bee genus *Colletes* (Hymenoptera: Colletidae). *Molecular Phylogenetics and Evolution*, 141, 106603. https://doi.org/10.1016/j.ympev.2019.106603

Onuferko, T. M., Packer, L., & Genaro, J. A. (2021). *Brachymelecta* Linsley, 1939, previously the rarest North American bee genus, was described from an aberrant specimen and is the senior synonym for *Xeromelecta* Linsley, 1939. *European Journal of Taxonomy*, 754(1), 1-51. https://doi.org/10.5852/ejt.2021.754.1393

Petrie, M., & Møller, A. P. (1991). Laying eggs in others' nests: Intraspecific brood parasitism in birds. *Trends in Ecology and Evolution,* 6(10), 315-320. https://doi.org/10.1016/0169-5347(91)90038-Y

Policarová, J., Cardinal, S., Martins, A. C., & Straka, J. (2019). The role of floral oils in the evolution of apid bees (Hymenoptera:

Apidae). *Biological Journal of the Linnean Society*, 128(2), 486-497. https://doi.org/10.1093/biolinnean/blz099

Rabeling, C., Schultz, T. R., Pierce, N. E., & Bacci, M. (2014). A social parasite evolved reproductive isolation from its fungus-growing ant host in sympatry. *Current Biology*, 24(17), 2047-2052. https://doi.org/10.1016/j.cub.2014.07.048.

Rambaut, A. (2014). FigTree v1.4.4. Available at http://tree.bio.ed.ac.uk/software/figtree/.

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, 22(5), 939-946. https://doi.org/10.1101/gr.128124.111

Roig-Alsina, A. (1989). The tribe Osirini, its scope, classification, and revisions of the genera *Parepeolus* and *Osirinus* (Hymenoptera, Apoidea, Anthophoridae). *The University of Kansas Science Bulletin*, 54, 1-23. https://doi.org/10.5962/bhl.part.19636

Rozen, J. G. (2003). Eggs, ovariole numbers, and modes of parasitism of cleptoparasitic bees, with emphasis on neotropical species (Hymenoptera: Apoidea). *American Museum Novitates*, 3413, 1-36. https://doi.org/10.1206/0003-0082(2003)413<0001:eonamo>2.0.co;2

Rozen, J. G., & Özbek, H. (2003). Oocytes, eggs, and ovarioles of some long-tongued bees (Hymenoptera: Apoidea). *American Museum Novitates*, 3393, 1-35 https://doi.org/10.1206/0003-0082(2003)393<0001:oeaoos>2.0.co;2

Rozen, J. G., & Roig-Alsina, A. (1991). Biology, larvae, and oocytes of the parasitic bee tribe Caenoprosopidini (Hymenoptera: Anthophoridae: Nomadinae). *American Museum Novitates*, 3004, 1-10

Smith, J. A., Chenoweth, L. B., Tierney, S. M., & Schwarz, M. P. (2013). Repeated origins of social parasitism in allodapine bees indicate that the weak form of Emery's rule is widespread, yet sympatric speciation remains highly problematic. *Biological Journal of the Linnean Society*, 109(2), 320-331. https://doi.org/10.1111/bij.12043

Snelling, R. R. (1986). Contributions toward a revision of the new world nomadine bees. A partitioning of the genus *Nomada* (Hymenoptera: Anthophoridae). *Contributions in Science: Natural History Museum of Los Angeles County*, 376, 1-32

Straka, J., & Bogusch, P. (2007). Description of immature stages of cleptoparasitic bees *Epeoloides coecutiens* and *Leiopodus trochantericus* (Hymenoptera: Apidae: Osirini, Protepeolini) with remarks to their unusual biology. *Entomologica Fennica,* 18(4), 242-254. https://doi.org/10.33338/ef.84405

Sumner, S., Aanen, D. K., Delabie, J., & Boomsma, J. J. (2004). The evolution of social parasitism in Acromyrmex leaf-cutting ants: a test of Emery's rule. *Insectes Sociaux*, 51(1), 37-42. https://doi.org/10.1007/s00040-003-0723-z

Tagliacollo, V. A., & Lanfear, R. (2018). Estimating improved partitioning schemes for ultraconserved elements. *Molecular Biology and Evolution*, 35(7), 1798-1811. https://doi.org/10.1093/molbev/msy069

Talavera, G. & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, 56(4), 564-577. https://doi.org/10.1080/10635150701472164

Ward P. S. (1989). Genetic and social changes associated with ant speciation. In: *The Genetics of Social Evolution* (Breed M.D. & Page R.E., Eds). Westview Press, Boulder, 123-148

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics,* 19, 153. https://doi.org/10.1186/s12859-018-2129-y

Zink, A. G. (2000). The evolution of intraspecific brood parasitism in birds and insects. *American Naturalist*, 155(3), 395-405. https://doi.org/10.1086/303325

CHAPTER 3

GENOME OF THE BEE *HOLCOPASITES CALLIOPSIDIS* - A SPECIES SHOWING THE COMMON APID TRAIT OF BROOD PARASITISM

Abstract

Brood parasites represent a substantial but often poorly studied fraction of the wider diversity of bees. Brood parasitic bees complete their life cycles by infiltrating the nests of solitary host bees thereby enabling their offspring to exploit the food provisions intended for the host's offspring. Here, we present the draft assembly of the bee *Holcopasites calliopsidis*, the first brood parasitic species to be the subject of detailed genomic analysis. Consistent with previous findings on the genomic signatures of parasitism more broadly, we find that *H. calliopsidis* has the smallest genome currently known among bees (179 Mbp). This small genome does not appear to be the result of purging of repetitive DNA, with some indications of novel repetitive elements which may show signs of recent expansion. Nor does *H. calliopsidis* demonstrate any apparent net loss of genic content in comparison with non-parasitic species, though many individual gene families do show significant contractions. Although the basis of the small genome size of this species remains unclear, the identification of over 12,000 putative genes -with functional annotation for nearly 10,000 of these - is an important step in investigating the genomic basis of brood parasitism and provides a valuable dataset to be compared against new genomes that remain to be sequenced.

Introduction

Though bees are particularly well-known for eusociality, this way of life is in fact only seen in about 10% of bee species (Danforth, Minckley, & Neff 2019). Non-social bees demonstrate a wide range of behavioral strategies including specialized plant associations, diverse nesting strategies, and the parasitic exploitation of other bees. With over 2,700 species, bees include a higher proportion of brood parasites than any animal taxon of comparable size (Sless et al. in review, or Chapter 1 herein). Yet despite this prevalence, brood parasitic species have received relatively little attention, including in the field of genomics. While some seventy bee species now have publicly available genomes through GenBank, nearly three-quarters of these represent social species (Sayers et al. 2020). Currently, one brood parasite (*Nomada fabriciana*) has been sequenced (Wellcome Sanger Institute 2021), however this genome is unannotated and has not been the subject of further analysis.

Broad shifts in genomic organization and architecture have been discovered that can be associated with the evolution of sociality (Kapheim et al. 2015; Shell et al. 2021) and social parasitism (Schrader et al. 2021) in other groups of hymenopterans, and the evolution of brood parasitism similarly involves phenotypic and behavioral shifts which must have a genomic basis. Some general patterns in genomic evolution have been identified among parasitic animals more generally (reviewed by Poulin & Randhawa 2013). One of the clearest of these is a trend towards reduced genome size in

parasitic species, which has also been noted for other parasitic hymenopterans (Ardila-Garcia et al. 2010; Sundberg & Pulkkinen 2015). The typical explanation for this pattern involves relaxed selection on parts of the genome necessary for survival in a free-living organism, but which may be effectively "off-loaded" by a parasite to its host. While this may involve the loss of metabolic or developmental genes in endoparasites that spend their entire lives within a host, parasitic Hymenoptera (including brood parasites) retain a free-living adult stage without losses in basic functionality. Despite this differing dynamic, brood parasites may be thought of as off-loading *behavioral* responsibilities rather than metabolic attributes (in the form of parental care through nest-building and food provisioning) to the host organisms on which they rely for survival. The question then follows: do brood parasitic bees show similar patterns of broad-scale genomic evolution to other parasites?

The genus *Holcopasites* is a member of the oldest and most diverse group of brood parasitic bees, the subfamily Nomadinae (Apidae) and the only Nearctic representative of the tribe Ammobatoidini. The predominant species in northeastern North America, *H. calliopsidis*, is a specialist parasite of the solitary mining bee *Calliopsis andreniformis*. Though the study of this single species presents a limited opportunity for inferences about the evolutionary signatures of brood parasitism on insect genomes more generally, it represents an important starting point that can serve as the basis for further comparative study as more genomes become available.

Therefore, we here provide broad-scale characterization of the genome of *H. calliopsidis*.

Methods

Specimen Collection and DNA Extraction

Individuals of *Holcopasites calliopsidis* were obtained in June of 2020 from the Lime Hollow Nature Center in Cortland, New York, USA (42.57°N, 76.25°W). Specimens were collected by hand-netting, flash-frozen in liquid nitrogen, and sexed with the assistance of a dissecting microscope before storing at -80°C. Only male specimens were used for subsequent steps, due to the benefit of their haploid genomes. High molecular weight DNA extractions were conducted using a Qiagen 20/G genomic-tip kit (catalog #10223) and associated buffers (catalog #19060). The protocol used followed that recommended by Qiagen's Genomic DNA Handbook (accessed from https://www.qiagen.com/us/resources/resourcedetail?id=d2b85b26-16dd-4259-a3a7-a08cbd2a08a3&lang=en). Briefly, whole specimens were mechanically homogenized using a sterile pestle before being treated with RNase A (NEB catalog #T3018L) and proteinase K for two hours. The resulting lysate was added to the equilibrated genomic tip columns, washed three times, and eluted. The eluted genomic DNA was then washed one more time with 70% ethanol before final resuspension in TE buffer. Samples were then assayed for quality and DNA concentration using a Qubit 4 fluorimeter.

Sequencing and Assembly

Four samples were sent to the University of Maryland's Institute for Genome Sciences for sequencing. These were processed using Pacific Biosciences' low-input library preparation and analyzed for DNA quality and size distribution. The highest-quality sample was then sequenced with a 30-hour run on a Sequel II 8M SMRT Cell in CCS/HiFi mode. Raw data, subreads, and circular consensus sequences were subsequently received from the sequencing facility. The *H. calliopsidis* genome was assembled from the CCS reads described above using SPAdes v3.14.0 (Bankevich et al. 2012) in "assembler only" mode. The resulting assembly was analyzed for completeness and quality using QUAST v5.0.2 (Gurevich et al. 2013) and BUSCO v5.0.0 (Seppi, Manni, and Zdobnov 2019).

Repetitive Element Identification

The genome was searched for repetitive content using two passes of the program RepeatMasker v4.1.0 (Smit, Hubley, & Green 2013-2015) in softmasking mode. An initial run was conducted using a canonical repeat library from RepBase, DFam 5.0 (Storer et al. 2021). The result was then passed into a second run of RepeatMasker with a custom, species-specific repeat library generated using RepeatModeler v2.0.1 (Flynn et al. 2020).

Gene Annotation

The repeat-masked genome was annotated using BRAKER v2.1.5 (Brůna et al. 2021) in "EP" mode, along with a database of orthologous arthropod protein sequences obtained from ensembl.org

(Howe et al. 2021). Annotation files from different lines of evidence within the BRAKER2 pipeline (Augustus and GeneMark) were combined using EVidenceModeler v1.1.1 (Haas et al. 2008) to generate a final annotation file. Transcripts and protein sequences were then extracted from the genome with the software GffRead (Pertea and Pertea 2020) using this annotation file as a guide.

Ortholog Detection

Proteomes from thirteen other hymenopteran species (*Habropoda laboriosa*, *Apis mellifera*, *Bombus impatiens*, *Ceratina calcarata*, *Osmia bicornis*, *Megachile rotundata*, *Megalopta genalis*, *Nomia melanderi*, *Camponotus floridanus*, *Polistes dominula*, *Chelonus insularis*, *Nasonia vitripennis*, *Athalia rosae*), with *Drosophila melanogaster* as an outgroup, were obtained from GenBank for comparison with newly annotated protein sequences from the *H. calliopsidis* genome. The software OrthoFinder v2.5.4 (Emms & Kelly 2019) was used to identify orthogroups (consisting of both orthologous and paralogous sequences) among these protein sequences, running with default settings and using DIAMOND (Buchfink, Xie, & Huson 2015) for sequence comparisons.

Gene Family Evolution

Expansion and contraction of gene families was analyzed with CAFE v4.2.1 (Han et al. 2013) using the orthogroups detected by OrthoFinder as the input dataset. The initial species tree produced by OrthoFinder did not produce the correct topology in comparison with

established phylogenies and cannot produce an ultrametric tree when provided with a preset topology. As a result, the time-calibrated tree required by CAFE was generated manually using divergence time estimates from a previous phylogenomic study (Peters et al. 2017).

Functional Analysis

Initial gene ontology annotations were conducted with both InterProScan v5.51-85.0 (Jones et al. 2014) and DIAMOND v2.0.9 (Buchfink, Xie, & Huson 2015) using the Uniref90 reference dataset. The resulting .xml files from both programs were passed on to Blast2GO v1.5.1 (Götz et al. 2008). Functional enrichment analyses were conducted in Cytoscape (Shannon et al. 2003) using the BiNGO application v3.0.4 (Maere, Heymans, & Kuiper 2005) with a custom annotation file modified from the Blast2GO output.

Results

Genome Assembly:

The final genome assembly after running SPAdes on the circular consensus sequencing data included 5,627 contigs, with 491 >1 kbp in length (Figure 3.1). This resulted in a total assembly size of just over 179 Mbp, corresponding to an expected C-value of 0.183 pg. Though the estimated coverage of 47X was lower than some other recent assemblies, the N50 value of nearly 4.8 Mbp compares favorably (Table 3.1). The L50 value of 14 indicates that the largest several contigs likely represent entire chromosomes or major sections thereof, however the exact chromosome count and structure cannot be

accurately determined from the assembly. The considerable number of extremely small scaffolds can be effectively ignored, as the smallest 5,188 of these (<1 kbp) contain just 0.56% of the overall assembly despite making up 92.2% of all scaffolds. This genome assembly appears to be largely complete, with 97% of the 5991 highly conserved single-copy orthologs in BUSCO's Hymenoptera dataset identified (the remainder comprising 0.3 % duplicated orthologs, 0.6% fragmented, and 2.1% not detected). It is also noteworthy that the genome of *H. calliopsidis* has one of the highest percentages of GC-content of any known bee at approximately 42.5%.

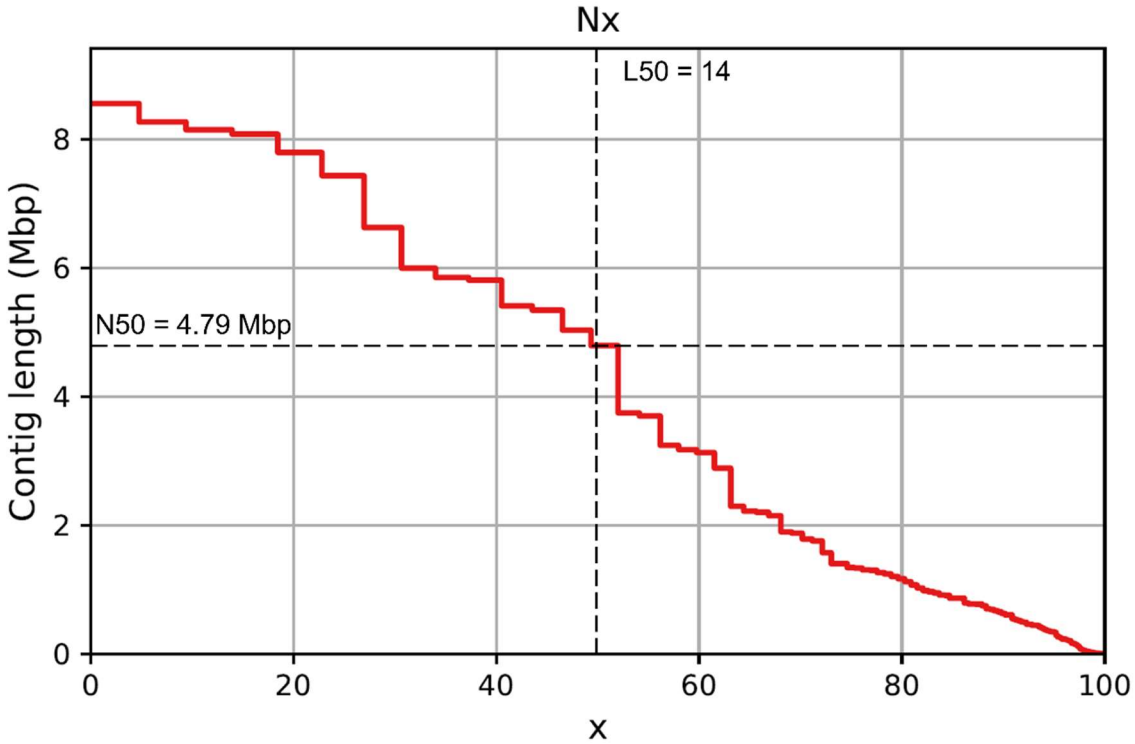


Figure 3.1: Nx plot generated by QUAST (Gurevich et al. 2013) showing contigs ranked by size vs. cumulative assembly completeness. N50 and L50 values (minimum contig size and minimum number of contigs containing 50% of the assembly respectively) for the *H. calliopsidis* genome are indicated by horizontal and vertical dashed lines.

Table 3.1: Comparison of assembly statistics for *Holcopasites calliopsidis* and recent assemblies from a selection of non-parasitic species spanning the diversity of bees (in approximate order of increasing phylogenetic distance from *H. calliopsidis*).

Species	Family: Subfamily	Assembly Size (Mbp)	# Scaffolds	N50 (bp)	Coverage (X)	Source
Holcopasites calliopsidis	Apidae: Nomadinae	179	5627 (491 > 1kb)	4,790,652	~47	This study
Habropoda laboriosa	Apidae: Anthophorinae	297	27,566	1,784,116	113	Kapheim et al. 2015
Apis mellifera	Apidae: Apinae	225	177	13,619,445	192	Wallberg et al. 2019
Bombus impatiens	Apidae: Apinae	248	5,559	1,399,493	108	Sadd et al. 2015
Megachile rotundata	Megachilidae: Megachilinae	273	6,266	1,699,680	272	Kapheim et al. 2015
Nomia melanderi	Halictidae: Nomiinae	299.6	3,194 > 1kb	2,054,768	75	Kapheim et al. 2019
Colletes gigas	Colletidae: Colletinae	273	326	8,109,000	147.5	Zhou et al. 2020

Genome Size:

Based on a comparison of other assemblies published in GenBank, the 179 Mbp *H. calliopsidis* genome assembly is the smallest of any bee species (Figure 3.2), and among other Hymenopterans is exceeded in this respect only by some parasitoid wasps of the family Braconidae. The high proportion (97%) of orthologs detected by the BUSCO analysis above indicates that this result is not simply an artifact due to missing DNA, coverage problems, or a poor assembly. The true genome size should therefore be very close to the

size of this assembly, and possibly even smaller if some trailing contigs represent microbial sequences rather than *H. calliopsidis* DNA.

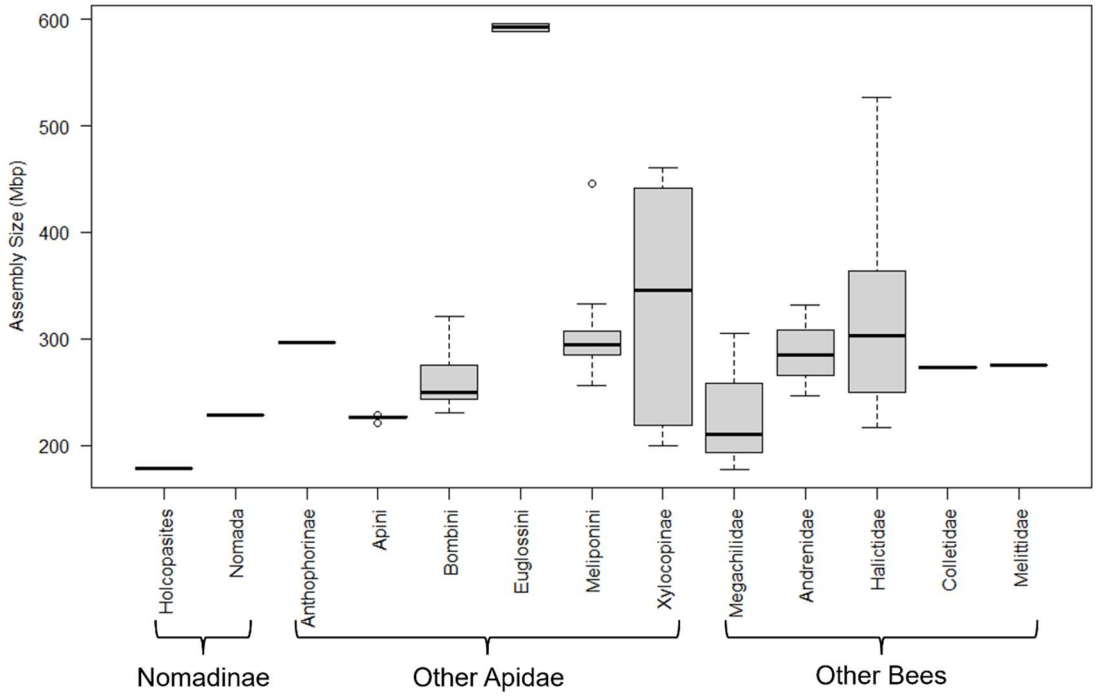


Figure 3.2: Size comparison in megabases of all bee species with genome assemblies available on GenBank (n = 70), arranged by approximate phylogenetic distance from *Holcopasites calliopsidis*.

Repetitive Content:

The initial run of RepeatMasker using the Dfam canonical repeat library identified a relatively small proportion of repetitive sequences in the *H. calliopsidis* genome. Less than 1% of the assembly is composed of known non-interspersed elements including satellite sequences and short simple repeats (SSRs), in contrast to ~4% in the genome of the Western honey bee, *Apis mellifera*. Similarly, *H. calliopsidis* has proportionately about half as many canonical retroelements than *A. mellifera* (2.12% vs. 4.18%), though slightly more DNA transposons in

direct comparison (1.79% vs. 0.57%). However, the second run using a species-specific library produced by RepeatModeler identified a sizeable proportion of the *H. calliopsidis* genome as "unclassified" repeats not yet represented in any database of repetitive sequences, comprising a total of nearly 14% of the assembly. These unclassified repetitive elements specifically included a total of 698 distinct families with an average size of 200.9 bp and average of 154.6 copies masked throughout the genome.

Gene Annotation:

The BRAKER2 pipeline identified 13,028 putative protein sequences derived from 12,364 predicted genes throughout the *H. calliopsidis* genome. Overall, this is similar to the number of genes reported from most other bee species. Though exact identity of these genes is often not possible to ascertain, the overall number and density indicates no net loss of genes in the *H. calliopsidis* genome in contrast to non-parasitic bees. Additionally, the average number and length of introns is similar between *H. calliopsidis* and other species. For example, *H. calliopsidis* averages 5.34 introns of 524.3 bp each per transcript, while the comparable figures for the genome of *Colletes gigas* average 4.99 introns of 665 bp each (Zhou et al. 2020), resulting in comparable proportions of intronic content (62% and 67% of overall genic content respectively).

Orthology Detection:

OrthoFinder identified a total of 21,872 orthogroups across the proteomes of the included species (fourteen hymenopterans + *Drosophila melanogaster*), which together accounted for 97.5% of all genes. The mean number of genes per orthogroup was 15.4 (median = 7), and a total of 4,298 orthogroups contained genes from every included species (of which 101 were single-copy orthogroups). A total of 10,913 putative genes from the *H. calliopsidis* genome were placed into exactly 9,700 orthogroups which were conserved at a range of phylogenetic scales (Figure 3.3). Although this is similar to the average number of orthogroups for all species (mean = 9,910.47), *H. calliopsidis* had a noticeably higher fraction of genes which could not be assigned to any orthogroup (16.2%) than other species.

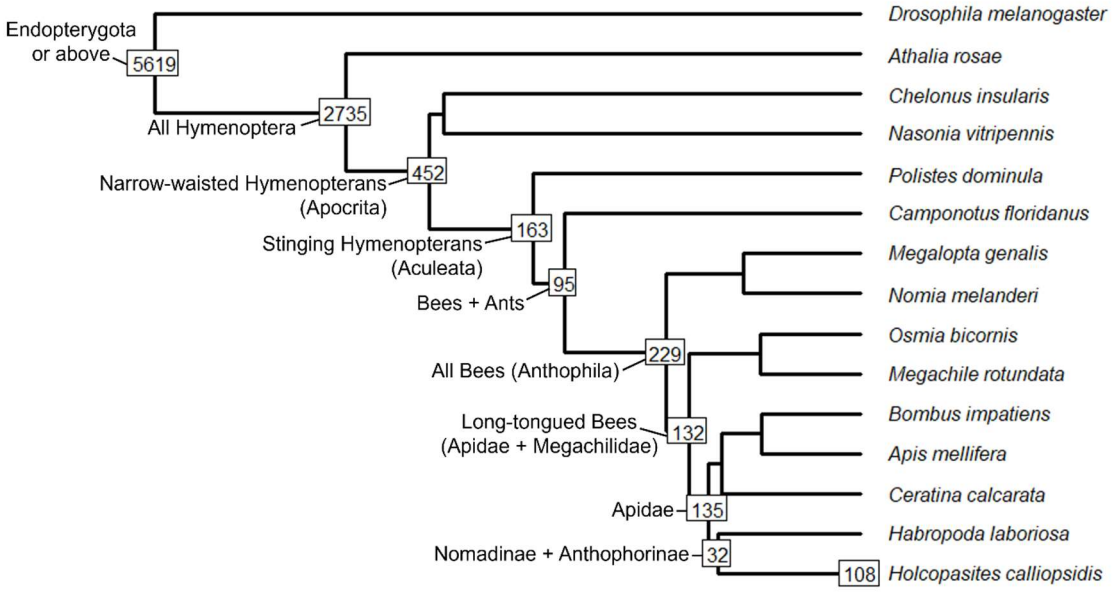


Figure 3.3: Phylogeny of *Holcopasites calliopsidis* and fourteen additional species used in orthology analysis. Boxes at nodes indicate the number of orthogroups containing at least one *H. calliopsidis* gene (9,700 in total) which were phylogenetically conserved at the level of the indicated clade. The box at the tip of the *H. calliopsidis* branch indicates orthogroups not found in any other species. Branch lengths

are to scale, with approximate divergence times taken from Peters et al. (2017).

A total of 108 orthogroups containing 393 genes were identified that appear to be unique to the *H. calliopsidis* genome - henceforth referred to as "*H. calliopsidis*-specific" genes. An important caveat of this terminology is that, due to the taxonomic bias of proteomes available for comparison, it is not possible to determine whether these genes are truly unique to *H. calliopsidis*, or more broadly conserved across the entire genus *Holcopasites*, tribe Ammobatoidini, subfamily Nomadinae, or an intermediate clade between these. However, these numbers are somewhat smaller than the overall mean of 390.6 species-specific orthogroups (and 1570 species-specific genes) among all fourteen hymenopteran genomes, despite the longer branch length of *Holcopasites* in contrast to several other included bees. Outside of the Nomadinae, 32 orthogroups were shared with their closest sister group (digger bees in subfamily Anthophorinae), 135 conserved across the family Apidae, 229 conserved across all bees, and 2,735 shared across all Hymenoptera. The majority of orthogroups containing *H. calliopsidis* genes (5,619) were identified as conserved at the level of holometabolous insects (represented by *D. melanogaster*) or higher.

Gene Family Evolution:

The orthogroups described above were further analyzed with CAFE to identify nodes in the phylogeny of the fifteen included species with significant gene family expansions or contractions (Figure 3.4). A

single birth-death rate parameter (λ = 0.00259067) was estimated for the entire tree. Significantly elevated rates of gene family evolution (family-wide p-value < 0.01) were detected in 831 (3.80%) of orthogroups, with 212 rapidly evolving gene families identified in *H. calliopsidis* specifically (15 expansions, 197 contractions). In comparison with the other included species, this represents the second-highest number of rapidly evolving gene families (after *Nomia melanderi* with 248), but by far the most gene family contractions. Interestingly, when ranked by statistical significance, nine of the top ten gene families which have undergone the most rapid evolution in *H. calliopsidis* are expansions rather than contractions. Of these, several families include genes with blast2GO annotations suggesting involvement with transposable or retroviral elements, which may indicate a recent or ongoing spread of such features in the genome.

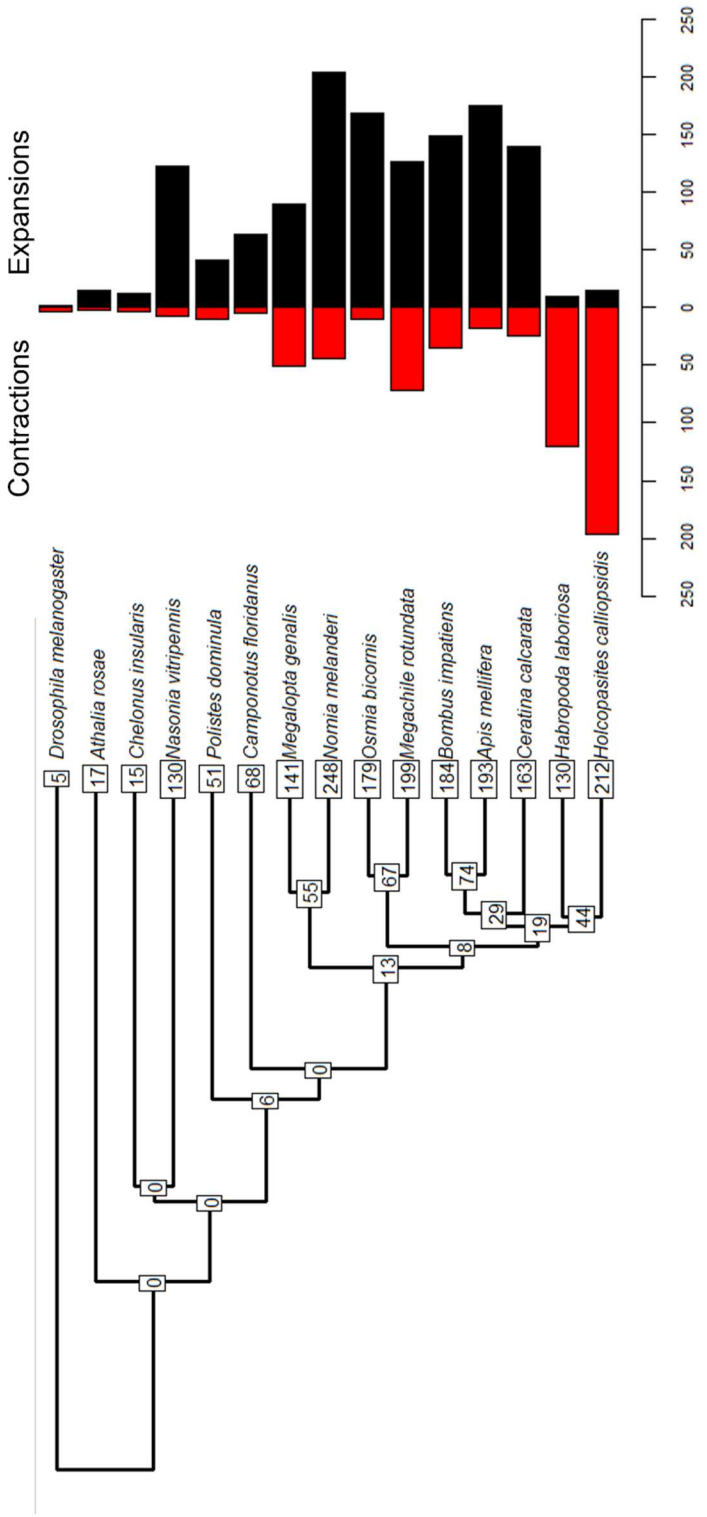


Figure 3.4: Phylogeny of *Holcopasites calliopsidis* and fourteen additional species used in analysis of gene family evolution. Numbers at nodes and tips indicate how many rapidly evolving gene families were detected by CAFE for each species/lineage respectively. Bars further separate rapidly evolving gene families for each tip into contractions (red) or expansions (black) in gene family size.

Functional Analysis:

In total, InterProScan and blast2GO assigned over 60,000 GO terms to 9,943 genes annotated from the *H. calliopsis* genome, leaving about one quarter of predicted genes without functional annotation. Analysis of the distribution of level 3 GO terms indicates few surprising features; various metabolic processes account for the most commonly assigned biological processes, and the most common molecular functions included enzymatic activity and binding of proteins and other compounds. A BiNGO analysis of the 393 *Holcopasites*-specific genes identified through orthology detection compared to the genome as a whole identified several GO terms that appear enriched for these loci (Figure 3.5). Several of these (e.g. transposition, DNA integration, and various classes of endonuclease activity) may in fact be signatures of the large fraction of the genome containing presumed *Holcopasites*-specific repetitive elements as identified above.

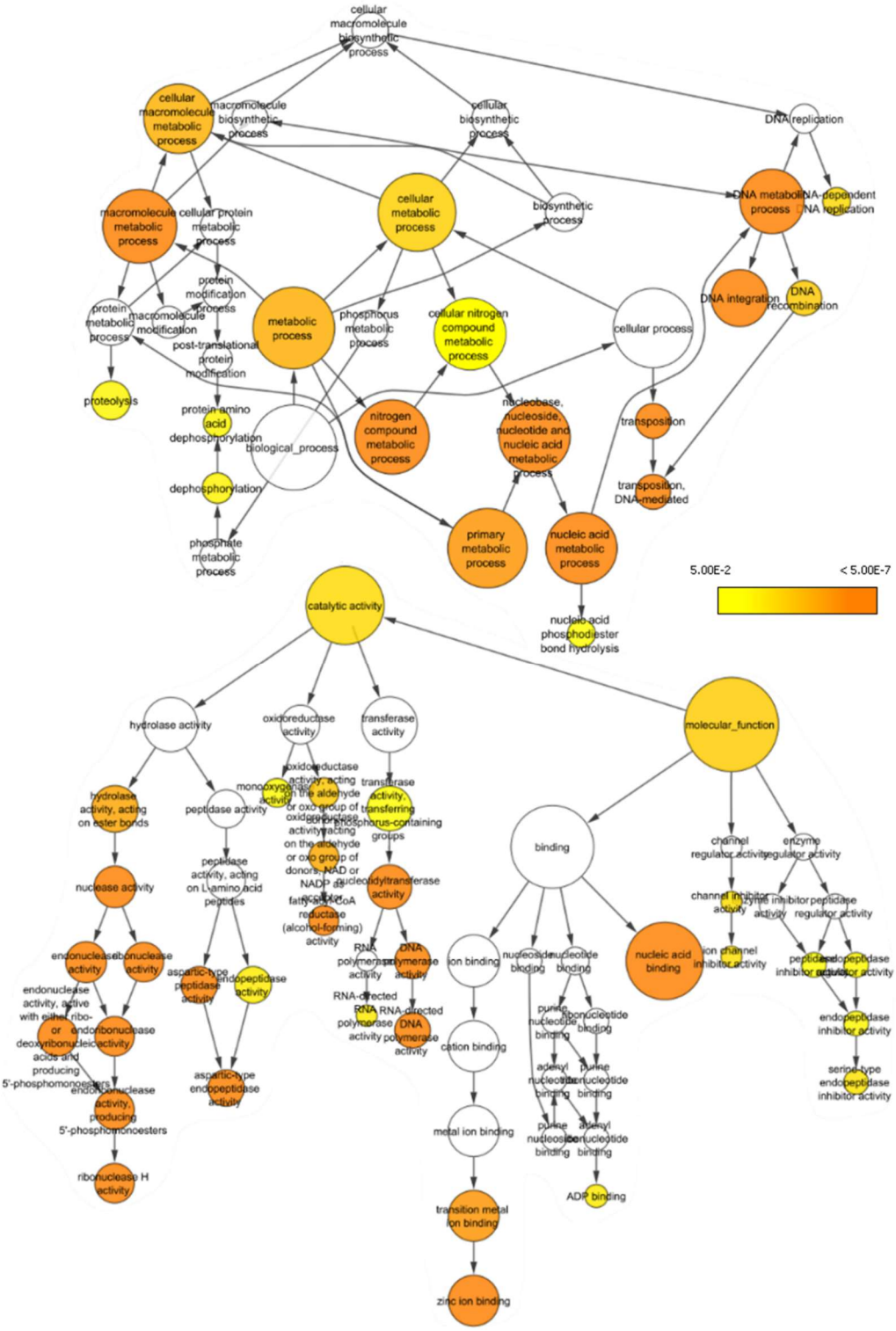


Figure 3.5: Selected sections of gene ontology network showing terms with significant enrichment among the set of 393 *Holcopasites calliopsidis*-specific genes compared to all annotated genes. Color gradient indicates degree of statistical significance from p = 0.05 (yellow) to p = 5 x 10-7 (dark orange)

A second BiNGO analysis compared the entire proteomes of our *H. calliopsidis* assembly with those of *Habropoda laboriosa* (the phylogenetically closest well-annotated genome to *H. calliopsidis*) as well as *Apis mellifera* to identify the GO terms which appear enriched in *H. calliopsidis* in contrast to its non-parasitic relatives or vice versa. Overall, this analysis was inconclusive. Many GO terms were identified as being significantly over- or underrepresented in *H. calliopsidis* in direct comparison to the other two species, yet in some cases these findings may be artefactual. Differences in the depth of gene ontology annotations achieved for the three genomes can result in artificial overrepresentation of GO terms which may not appear in all species, but which are simply child nodes of other well-represented terms. For example, GO terms for synthesis of all essential amino acids were identified as overrepresented in *H. calliopsidis* over *Ha. laboriosa*/*A. mellifera*, though it is difficult to explain this biologically. Some terms (e.g. neurogenesis/neuron development, oogenesis) were underrepresented in *H. calliopsidis* in ways which appear superficially consistent with *a priori* expectations about adaptations to the brood parasitic life history strategy. However, this must be considered with the caveat that enrichment of certain ontology terms, which may represent expansion of genes involved in a given biological process, is still separated from the actual control of such processes by several layers of regulation (e.g. transcriptional regulation, splicing, post-translational modifications).

Discussion

The most striking feature of our broad-scale characterization of the *H. calliopsidis* genome is its small genome size, and consequently we focus the discussion on this attribute. However, as the first brood parasitic bee for which a well-characterized genome has been produced, there is much which remains to be learned. Brood parasitism has evolved at least twenty times independently in bees, with *H. calliopsidis* representing the oldest and most diverse such clade in the subfamily Nomadinae (Sless et al. in review, or Chapter 1 herein). In this sense, our study has opened a new avenue of genomic research in relation to a well-known and fascinating behavioral strategy. This work is also complementary to nascent genomic studies on brood parasitic birds (such as cuckoos and cowbirds), which share a common life history despite their great phylogenetic distance from bees, and similarly include recently sequenced genomes available through GenBank but awaiting in-depth analysis (Wolf et al. unpublished data; Wuitchik et al. unpublished data; Zhang unpublished data).

Genome Size Reduction:

While most bee genomes fall into a fairly consistent range of ~250-350 Mbp, with a mean of 287.05 Mbp for all 70 species with genomes available on GenBank, our assembly for *H. calliopsidis* is noticeably smaller at just 179 Mbp. Methodological explanations for this finding seem unlikely, but corroboration of this estimate with cytological techniques could serve as an additional line of evidence.

Assuming that the ancestral genome size for Apidae (or indeed for all bees) falls within the typical 250-350 Mbp range, this represents a reduction of ~25-45% in *H. calliopsidis*. The only other member of Nomadinae with a known genome size, *Nomada fabriciana*, falls on the lower end of the "typical" bee range at 233 Mbp, indicating that at least some of this genomic contraction may be a recent change in the ancestors of *H. calliopsidis*, rather than a feature of all brood parasitic bees in the subfamily Nomadinae. Significant changes in genome size, especially when they occur over brief time periods, have often been attributed to the expansion or purging of repetitive elements (e.g. Hancock et al. 2021). Indeed, some bee lineages such as the orchid bees (Euglossini) appear to have undergone massive genomic expansions as a result of increased repetitive content (Brand et al. 2017). However, the reverse situation in which repeat sequences are purged from the genome does not appear to be the case in *H. calliopsidis*, which shows comparable or even higher levels of repetitive content to other species as a fraction of the genome. The large fraction of "unclassified" repetitive DNA which could not be matched to existing databases is likely due to the fact that this *de novo* genome assembly has no close relatives which have been previously analyzed for repetitive content. A similar phenomenon has been reported for at least one other phylogenetically distinct genome, that of *Colletes gigas*, which also consisted of >10% "unclassified" repetitive elements (Zhou et al. 2020).

Similarly, the reduced genome size in *H. calliopsidis* does not appear to be a result of net loss in genic content. Though many genes found in other species did not have orthologues annotated in *H. calliopsidis*, the overall number of putative genes identified was similar to other known bees at approximately 12,000. The average number and length of introns is also similar between *H. calliopsidis* and *A. mellifera*, which would seem to rule out any substantial purging of non-coding sequences within genes. However, this certainly does not mean that genic content is static or identical to that of other species. Many gene families are identified as having experienced rapid evolution in *H. calliopsidis*, though with the caveat that differing methodologies for gene annotation used across the compared species may artificially influence this analysis. It may be the case that *H. calliopsidis* does in fact exhibit the loss of a large number of genes associated with non-parasitic life histories, but that these losses are compensated to some extent by a number of rapidly evolving gene families. This finding may in fact be related to the large amount of "unclassified" repetitive elements discussed above, since transposable elements containing functional genes may be included among the most rapidly expanding gene families. It remains unclear exactly which types of material account for the reduction in *H. calliopsidis*' genome size compared to the ancestral bee genome.

Other Features:

Our genome assembly for *H. calliopsidis* is noticeably more GC-rich (42.5%) than most other bee genomes. Of those available on

GenBank, the mean GC-content is 37.6%, and only two species (*Andrena dorsata* and *haemorrhoa*) surpass *H. calliopsidis* in this regard (Sayers et al. 2020). It is unclear what relation if any this may have to the evolution of brood parasitism, however. Proportion of GC bases is not generally correlated with genome size in animals (Li & Du 2014), though a positive correlation between GC-content and recombination rates has been noted in honey bees (Beye et al. 2006).

The orthology detection analysis with OrthoFinder identified fewer unique orthogroups containing only *H. calliopsidis* genes than for any other included species besides *Habropoda laboriosa*. This may initially seem to suggest that *H. calliopsidis* has experienced less evolution of novel genes than most other Hymenoptera. However, over 2,000 putative *H. calliopsidis* genes (a higher fraction than any other species) were not able to be assigned to any orthogroup, and some of these are surely also unique (but single-copy) genes that arose sometime after the split between Nomadinae and its sister group. Unsurprisingly though, the lack of interspecific orthologs for these unassigned genes also meant that the vast majority of them received no functional annotations from blast2GO, and so their identity remains unclear.

The large number of orthogroups identified by CAFE as rapidly evolving in *H. calliopsidis* may in part be a result of different gene annotation pipelines used for different genomes. For example, some gene families that were identified as contractions based on the

presence of multiple copies in other species vs. a single copy in our assembly may simply represent cases where some paralogs were missed during annotation. However, the smaller number of gene family expansions identified in *H. calliopsidis* seem less likely to be the result of methodological differences, since in these cases low copy numbers were independently identified across many other genomes.

Future Directions:

The annotated draft genome of *Holcopasites calliopsidis* presented herein is a major first step towards understanding the genomic basis of brood parasitism in the Nomadinae and in bees more broadly. However, as already addressed throughout this study, the amount of information which can be gained from a single genome is inherently limited. From this dataset alone, it is possible to identify many interesting features of genomic content in *H. calliopsidis*, but not their age or phylogenetic distribution as orthologs. Some such features may in fact be ancient signatures of the initial transition to brood parasitism in the subfamily Nomadinae approximately 100 million years ago, while others could be much more recent innovations related to *H. calliopsidis*' status as a narrow host specialist.

Future investigation into the genomics of brood parasitism should focus on sampling other species representing a wider phylogenetic distribution of parasites, including other members of the Nomadinae as well as bees representing independent origins of parasitism. Comparative genomic investigation has already yielded

interesting results in studying the genomic basis of social behavior in bees (e.g. Kapheim et al. 2015; Shell et al. 2021), and it is our hope that the *H. calliopsidis* genome will serve as a starting point for the parallel exploration of the equally fascinating life history innovation of brood parasitism.

REFERENCES

Ardila-Garcia AM, Umphrey GJ, Gregory TR. 2010. An expansion of the genome size dataset for the insect order Hymenoptera, with a first test of parasitism and eusociality as possible constraints. Insect Mol. Biol. 19(3):337-346. DOI: 10.1111/j.1365-2583.2010.00992.x

Ascher JS, Pickering J. 2020. Discover life bee species guide and world checklist (Hymenoptera: Apoidea: Anthophila). Available at http://www.discoverlife.org/mp/20q?guide=Apoidea_species

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19(5):455-477. DOI: 10.1089/cmb.2012.0021

Beye M, Gattermeier I, Hasselmann M, Gempe T, Schioett M, Baines JF, Schlipalius D, Mougel F, Emore C, Rueppell O, Sirviö A, Guzmán-Novoa E, Hunt G, Solignac M, Page Jr. RE. 2006. Exceptionally high levels of recombination across the honey bee genome. Genome Res. 16:1339-1344. DOI: 10.1101/gr.5680406

Brand P, Saleh N, Pan H, Li C, Kapheim KM, & Ramírez S. 2017. The nuclear and mitochondrial genomes of the facultatively eusocial orchid bee *Euglossa dilemma*. G3: Genes Genom. Genet. 7(9):2891-2898. DOI: 10.1534/g3.117.043687

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom. Bioinform. 3(1):lqaa108. DOI: 10.1093/nargab/lqaa108

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods. 12:59-60.

Danforth B, Minckley R, Neff J. 2019. The Solitary Bees: Biology, Evolution, Conservation. Princeton; Oxford: Princeton University Press.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20:238 DOI:

10.1186/s13059-019-1832-y

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Nat. Acad. Sci. U.S.A. 117(17):9451-9457. DOI: 10.1073/pnas/1921046117

Götz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36(10):3420-35. DOI: 10.1093/nar/gkn176

Gurevich A, Saveliev S, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29(8):1072-1075. DOI: 10.1093/bioinformatics/btt086

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9(7). DOI: 10.1186/gb-2008-9-1-r7

Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol. Biol. Evol. 30(8), 1987-1997. DOI: 10.1093/molbev/mst100

Hancock ZB, Hardin FO, Murthy A, Hillhouse A, Johnston JS. 2021. Rapid genomic expansion and purging associated with habitat transitions in a clade of beach crustaceans (Amphipoda: Haustoriidae). J. Crust. Biol. 41(3):ruab042. DOI: 10.1093/jcbiol/ruab042

Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Charkhchi M, Cummins C, Fioretto LDR, Davidson C, Dodiya K, El Houdaigui B, Fatima R, Gall A, Giron CG, Grego T, Guijarro-Clarke C, Haggerty L, Hemrom A, Hourlier T, Izuogu OG, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Martinez JG, Marugán JC, Maurel T, McMahon AC, Mohanan S, Moore B, Muffato M, Oheh DN, Paraschas D, Parker A, Parton A, Prosovetskaia I, Sakthivel MP, Salam AIA, Schmitt BM, Schuilenburg H, Sheppard D, Steed E, Szpak M, Szuba M,

Taylor K, Thormann A, Threadgold G, Walts B, Winterbottom A, Chakiachvili M, Chaubal A, De Silva N, Flint B, Frankish A, Hunt SE, IIsley GR, Langridge N, Loveland JE, Martin FJ, Mudge JM, Morales J, Perry E, Ruffier M, Tate J, Thybert D, Trevanion SJ, Cunningham F, Yates AD, Zerbino DR, Flicek P. 2021. Ensembl 2021. Nucleic Acids Res. 49(1):884–891. DOI: 10.1093/nar/gkaa942

Jones P, Binns D, Chang H, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30(9):1236-1240. DOI: 10.1093/bioinformatics/btu031

Kapheim M, Pan H, Li C, Salzberg SL, Puiu D, Magoc T, Robertson HM, Hudson ME, Venkat A, Fischman BJ, Hernandez A, Yandell M, Ence D, Holt C, Yocum GD, Kemp WP, Bosch J, Waterhouse RM, Zdobnov EM, Stolle E, Kraus FB, Helbing S, Moritz RFA, Glastad KM, Hunt BG, Goodisman MAD, Hauser F, Grimmelikhuijzen CJP, Pinheiro DG, Nunes FMF, Soares MPM, Tanaka ED, Simoes ZLP, Hartfelder K, Evans JD, Barribeau SM, Johnson RM, Massey JH, Southey BR, Hasselmann M, Hamacher D, Biewer M, Kent CF, Zayed A, Blatti III C, Sinha S, Johnston JS, Hanrahan SJ, Kocher SD, Wang J, Robinson GE, Zhang G. 2015. Genomic signatures of evolutionary transitions from solitary to group living. Science. 348(6239), 1139–1144. DOI: 10.1126/science.aaa4788

Li X-Q, Du D. 2014. Variation, evolution, and correlation analysis of C+G content and genome or chromosome size in different kingdoms and phyla. PLoS One. 9(2):e88339. DOI: 10.1371/journal.pone.0088339

Maere S, Heymans K, Kuiper M. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. Bioinformatics. 21:3448-3449. DOI: 10.1093/bioinformatics/bti551

Michener CD. 2007. The Bees of the World. 2nd Edition. Baltimore: John Hopkins University Press

Pertea G, Pertea M. 2020. GFF Utilities: GffRead and GffCompare. F1000research. 9:304. DOI: 10.12688/f1000research.23297.2

Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A, Podsiadlowski L, Petersen M, Lanfear R, Diez PA, Heraty J, Kjer KM, Klopfstein S, Meier R, Polidori C, Schmitt T, Liu S, Zhou X, Wappler T, Rust J, Misof B, Niehuis O. 2017. Evolutionary history of the Hymenoptera. Curr. Biol. 27(7), 1013-1018. DOI: 10.1013/j.cub.2017.01.027

Poulin R, Randhawa, HS. 2013. Evolution of parasitism along convergent lines: from ecology to genomics. Parasitology. 142:S6-S15. DOI: 10.1017/S0031182013001674

Sadd BM, Barribeau SM, Bloch G, de Graaf DC, Dearden P, Elsik CG, Gadau J, Grimmelikhuijzen CJP, Hasselmann M, Lozier JD, Robertson HM, Smagghe G, Stolle E, Van Vaerenbergh M, Waterhouse RM, Bornberg-Bauer E, Klasberg S, Bennett AK, Câmara F, Guigó R, Hoff K, Mariotti M, Munoz-Torres M, Murphy T, Santesmasses D, Amdam GV, Beckers M, Beye M, Biewer M, Bitondi MMG, Blaxter ML, Bourke AFG, Brown MJF, Buechel SD, Cameron R, Cappelle K, Carolan JC, Christiaens O, Ciborowski KL, Clarke DF, Colgan TJ, Collins DH, Cridge AG, Dalmay T, Dreier S, du Plessis L, Duncan E, Erler S, Evans J, Falcon T, Flores K, Freitas FCP, Fuchikawa T, Gempe T, Hartfelder K, Hauser F, Helbing S, Humann FC, Irvine F, Jermiin LS, Johnson CE, Johnson RM, Jones AK, Kadowaki T, Kidner JH, Koch V, Köhler A, Kraus FB, Lattorff HMG, Leask M, Lockett GA, Mallon EB, Marco Antonio DS, Marxer M, Meeus I, Moritze RFA, Nair A, Näpflin K, Nissen I, Niu J, Nunes FMF, Oakeshott JG, Osborne A, Otte M, Pinheiro DG, Rossié N, Rueppell O, Santos CG, Schmid-Hempel R, Schmitt BD, Schulte C, Simões ZLP, Soares MPM, Swevers L, Winnebeck EC, Wolschin F, Yu N, Zdobnov EM, Aqrawi PK, Blankenburg KP, Coyle M, Francisco L, Hernandez AG, Holder M, Hudson ME, Jackson L, Jayaseelan J, Joshi V, Kovar C, Lee SL, Mata R, Mathew T, Newsham IF, Ngo R, Okwuonu G, Pham C, Pu L-L, Saada N, Santibanez J, Simmons D, Thornton R, Venkat A, Walden KKO, Wu Y-Q, Debyser G, Devreese B, Asher C, Blommaert J, Chipman AD, Chittka L, Fouks B, Liu J, O'Neill MP, Sumner S, Puiu D, Qu J, Salzberg SL, Scherer SE, Muzny DM, Richards S, Robinson GE, Gibbs RA, Schmid-Hempel R, Worley KC. 2015. The genomes of two key bumblebee species with primitive eusocial organization. Genome Biol. 16(76). DOI: 10.1186/s13059-015-0623-3

Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. 2020. GenBank. Nucleic Acids Res. 48(D1):D84-D86. DOI: 10.1093/nar/gkz956

Schrader L, Pan H, Bollazzi M, Schiøtt M, Larabee FJ, Bi X, Deng Y, Zhang G, Boomsma JJ, Rabeling C. 2021. Relaxed selection underlies genome erosion in socially parasitic ant species. Nat. Commun. 12(2918). DOI: 10.1038/s41467-021-23178-w

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. Methods Mol. Biol. 1962:227-245. DOI: 10.1007/978-1-4939-9173-0_14

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13(11):2498-504. DOI: 10.1101/gr.1239303

Shell WA, Steffen MA, Pare HK, Seetharam AS, Severin AJ, Toth AL, Rehan SM. 2021. Sociality sculpts similar patterns of molecular evolution in two independently evolved lineages of eusocial bees. Commun. Biol. 4(253). DOI: 10.1038/s42003-021-01770-6

Smit AFA, Hubley R, Green P. 2013-2015. RepeatMasker Open-4.0. Available from http://www.repeatmasker.org.

Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. 2021. The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob. DNA 12(2). DOI: 10.1186/s13100-020-00230-y

Sundberg L-R, Pulkkinen K. 2015. Genome size evolution in macroparasites. Int. J. Parsitol. 45(5): 285-288. DOI: 10.1016/j.ijpara.2014.12.007

Wallberg A, Bunikis I, Pettersson OV, Mosbech M-B, Childers AK, Evans JD, Mikheyev AS, Robertson HM, Robinson GE, Webster MT. 2019. A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. BMC Genom. 20(275). DOI: 10.1186/s12864-019-5642-0

Wellcome Sanger Institute. 2021. Tree of Life QC: *Nomada fabriciana*. Available from https://tolqc.cog.sanger.ac.uk/darwin/insects/Nomada_fabriciana

Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 31(20): 3350-3352. DOI: 10.1093/bioinformatics/btv383

Zhou Q-S, Luo A, Zhang F, Niu Z-Q, Wu Q-T, Xiong M, Orr MC, Zhu C-D. 2020. The first draft genome of the plasterer bee *Colletes gigas* (Hymenoptera: Colletidae: *Colletes*). Genome Biol. Evol. 12(6): 860-866. DOI: 10.1093/gbe/evaa090

APPENDIX I

Chapter 1 Supplementary Tables (see associated supplementary file):

Table Captions:

Supp. Table 1.1: List of all identified brood parasitic taxa separated by independent origins of parasitism and arranged by major taxon groups.

Supp. Table 1.2: List of example taxa excluded as brood parasites based on criteria in main text, arranged by major taxon groups.

Supp. Table 1.3: Identity and species richness of sister groups for each brood parasitic clade. Topology "A" indicates chosen hypothesis for sister clade; other letters, when present, correspond to alternative hypotheses, with question marks indicating ambiguous studies which could support multiple topologies.

Supp. Table 1.4: Crown ages and divergence times for brood parasitic clades and their sister groups

Supp. Table 1.5: Taxon subsets used for species richness and diversification analyses. "Y" indicates that a clade was included in a given dataset; other entries indicate reason for exclusion (e.g. monotypic taxa, uncertainty in identity or species count of sister group, uncertainty in monophyly of brood parasitic clade, lack of phylogenetic data or estimated dates). Dashes indicate absence of taxa from "ancestral parental care" or "no ancestral parental care" sets respectively, since these are mutually exclusive groups.

Supp. Table 1.6: Data used for species richness comparisons, split into six taxon sets (see Supp. Table 1.5)Supp. Table 1.6 (continued):

Supp. Table 1.7: Data used for species richness comparisons, split into six taxon sets (see Supp. Table 1.5)

Supp. Table 1.8: Results of species richness and diversity rate comparisons between brood parasitic clades and their sister groups. Asterisks indicate statistically significant results (* for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$)

Supp. Table 1.9: Summary of known host associations for each brood parasitic clade. Ancestral background indicates inferred life history strategy prior to evolution of brood parasitism, with particular reference to presence/absence of parental care. Evaluation against Emery's rule considers "strict sense" to mean that parasites use immediate sister group as their primary hosts, while the "loose sense" implies generally close, but non-sister relationship between hosts and parasites

APPENDIX II

Chapter 2 Supplementary Tables (see associated supplementary file):

Table Captions:

Supp. Table 2.1: List of all specimens used in phylogenetic analyses, listed in the same vertical order as they appear in Figure 2.1. Where applicable, classifications follow Bossert et al. 2019 and Bossert et al. 2020 in addition to changes recommended within the present study. For some samples, it was unclear which of two voucher specimens the sequenced material was taken from, so collection data for both are listed where they differed, and relevant species are indicated with an asterisk (*).

Supp. Table 2.2: List of brood parasitic species included in analyses with known host associations from literature.

Supp. Table 2.3: List of brood parasitic species included in analyses with known or suspected mode of parasitism based on Litman et al. 2013 and other sources as indicated.

APPENDIX III

Chapter 2 Supplementary Figures (see associated supplementary file):

Figure Captions:

Supp. Figure 2.1: Unpartitioned phylogeny constructed from 95% taxon-completeness matrix of UCE dataset, resulting in 1247 total loci. SH-aLRT and ultra-fast bootstrap values are indicated for some nodes; all unlabeled nodes have 100/100 support. Scale bar represents number of nucleotide substitutions per site.

Supp. Figure 2.2: Unpartitioned phylogeny constructed from 85% taxon-completeness matrix of UCE dataset, resulting in 1861 total loci. SH-aLRT and ultra-fast bootstrap values are indicated for some nodes; all unlabeled nodes have 100/100 support. Scale bar represents number of nucleotide substitutions per site.

Supp. Figure 2.3: Unpartitioned phylogeny constructed from 75% taxon-completeness matrix of UCE dataset, resulting in 2048 total loci. SH-aLRT and ultra-fast bootstrap values are indicated for some nodes; all unlabeled nodes have 100/100 support. Scale bar represents number of nucleotide substitutions per site.

Supp. Figure 2.4: Locus-partitioned phylogeny (with partitions merged) based on 1247 95% taxon-completeness matrix loci generated in IQ-TREE2. SH-aLRT and ultra-fast bootstrap values are indicated for some nodes; all unlabeled nodes have 100/100 support. Scale bar represents number of nucleotide substitutions per site.

Supp. Figure 2.5: Coalescent species phylogeny based on set of 1247 loci used in 95% taxon-completeness matrix, as generated by ASTRAL. Node labels are local posterior probabilities; all unlabeled nodes have a score of 1. Scale bar represents coalescent units.

Supp. Figure 2.6: SWSC-EN partitioned phylogeny based on 95% taxon-completeness matrix as in Figure 1. Gene and site concordance factors respectively are indicated at all nodes, as calculated by IQ-TREE using the same set of 1247 loci. Scale bar represents expected number of nucleotide substitutions per site.