Univerzita Karlova
Přírodovědecká fakulta

Studijní program: Biologie
Studijní obor: Zoologie

Bc. Tereza Fraňková

Assembly a anotace genomu včely *Ceratina nigrolabiata* s obourodičovskou péčí
Genome Assembly and Annotation of Biparental Bee *Ceratina nigrolabiata*

Diplomová práce

Školitel: Mgr. Jakub Straka, PhD.

Praha, 2020

Poděkování

Chtěla bych poděkovat svému školiteli Jakubovi Strakovi za jeho vstřícnost a skvělé vedení na mé cestě. Děkuji Míšovi Mikátovi za jeho rady a nadšení, které mě celou dobu doprovázejí. Dále patří můj vděk Hlávkově nadaci a Fondu mobility za jejich finanční podporu při cestě na univerzitu v Princetonu do laboratoře Sarah Kocher, kde jsem měla možnost spolupracovat s doktorkou Sarah Kocher a doktorem Andrew Webbem, který mě vedl jako svého padawana na cestě analýzy genomu. Děkuji své rodině za veškerou podporu, které se mi dostává a bez níž bych nemohla dnes psát tuto práci. Děkuji MetaCentru a výpočetním zdrojům Univerzity v Princetonu za možnost využívat jejich výpočetní zdroje při analýze. Děkuji svému příteli, který je mi každý den po boku a plně mě podporuje přestože nesdílí mé nadšení a lásku k hmyzu a trpí tak kvůli mým sběrům, které okupují náš mrazák. Hlavně děkuji naší univerzitě a fakultě a všem, kteří jsou součástí tohoto institutu. Bez Vás bychom o tomhle životě mohli jenom snít.

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 7.8.2020 Bc. Tereza Fraňková

Abstrakt

Obourodičovská péče je u blanokřídlého hmyzu málo prozkoumaný fenomén. Tento typ péče byl objeven a popsán z etologického hlediska u včely druhu *Ceratina nigrolabiata* v České republice. Jelikož ale obourodičovská péče není u hmyzu častým jevem, nedostatek genomických studií tohoto chování komplikuje porozumnění vzniku obourodičovské péče a její podstaty na genomické a fyziologické úrovni. Tato diplomová práce se zabývá analýzou genomu včely *C. nigrolabiata*. Obsahuje shrnutí známého rodičovského chování napříč hmyzem, charakteristiku včel rodu *Ceratina*, popis využitých genomických metod studia genomu a vytyčuje kandidátní geny pro připravovanou transkriptomickou studii obourodičovské péče této včely.

Klíčová slova: obourodičovská péče, genomika, Hymenoptera, Insecta, *Ceratina*, kyjorožky

Abstract

Biparental care in Hymenoptera is a little studied behaviour. This kind of parental care was discovered and the ethological aspect described in the bee *Ceratina nigrolabiata* from the Czech Republic and is well understood on the ethological level. However, biparental care is not a common behaviour and the lack of genomic studies of this behaviour complicates the understanding of the origin of the biparental care and its underlying pathways on the genomic and physiological level. This master's thesis presents the genome analyses of a biparental bee *Ceratina nigrolabiata*. It consists of a brief summary of the known parental behaviour across insects, characteristics of the genus *Ceratina*, summary of the used genomic methods and presents a candidate genes for the transcriptomic study of the biparental behaviour of *Ceratina nigrolabiata*.

Key words: biparental care, genomics, Hymenoptera, Insecta, Ceratina, small carpenter bees

Content

Introduction

The bees are one of the many interesting creatures inhabiting the world and they never cease to amaze us with their variability and behaviour. The foundation of all research is to classify and understand the manners of those living beings so we can probe into the secrets that cannot be seen with our bare eyes. Nowadays, genomics is a widely available field in biology. Although its usage is limited to those laboratories that focus on genetic and genomic information, many laboratories focusing on the ecology and ethology have been incorporating genomic, population genetics and other molecular ways of analysing data into their research.

It has been almost 20 years since the unbelievable expensive and demanding Human Genome Project was published (Venter et al. 2001) and in those 20 years, we have been actively using genomics and other omics in projects focusing on many widely different and even largely unknown species. From the secrets of endosymbiosis (Beinart et al. 2018) to mammoths (Rogers and Slatkin 2017), it seems genomics is a tool that help us broaden our knowledge about things we could only have dreamt about in the past. In viruses, genomic analyses can help us find new prevalent forms which we did not know about before and therefore can be problematic for developing vaccines (Pachetti et al. 2020; Khailany, Safdar, and Ozaslan 2020), sometimes, genomics can make us scratch our heads with new findings (Borowiec et al. 2015) and it can also change our perception of relatedness between different taxa we did not think are related (Boore and Staton 2002). However, genomics is not just a friend but can be an enemy, as we know from the 1991 tree of life (Sogin 1991), where many clades were located at the base of the tree because of LBA (Long Branch Attraction) (Stiller and Hall 1999; Hirt et al. 1999). This confused the scientific community for some time. We were, however, able to clear previous mistakes and for instance find a possible right position of Acoelomorpha (Bourlat et al. 2003; Mwinyi et al. 2010) previously thought to be flatworms. Presenting all the positives and negatives would take us many weeks and maybe months but it is always good to look at both in any field of interest.

Apart from establishing the communities in certain habitats (Evans et al. 2016) or the understanding of diseases (Timmers et al. 2019; Gagliano et al. 2016; Ding, Frantzeskos, and Orozco 2020), we can incorporate the understanding of certain genomic components and how they are related to the behaviour of organisms (Johnsson et al. 2018). Much of the studies in hymenoptera focus on sociality only (Weitekamp, Libbrecht, and Keller 2017; Sirviö et al. 2011; Shell and Rehan 2017; Greenberg et al. 2012; Yanega 1989; Jones et al. 2017; Rehan and

Toth 2015; Berens, Hunt, and Toth 2014; Rehan et al. 2018), however, the question still remains to be answered. Biparentality in hymenoptera is much less studied since not so long ago (Brockmann and Grafen 1989), the presence of biparentality has not been known. The only genomic (transcriptomic) evaluation of biparental care in insects has been done on *Nicrophorus vespilloides* (Silphidae, Coleoptera) (Roy-Zokan et al. 2015). In Hymenoptera only *Trypoxylon* species were known to exhibit a biparental behaviour (Brockmann and Grafen 1989; Alcock 1975), more so a male guarding behaviour, until not so long ago. The ongoing field research in the Czech Republic presented another species with biparental care, *Ceratina nigrolabiata* (Mikát et al. 2019) and possibly also some other species (Mikát et al., unpublished). Because such a limited data was available, biparental care has been studied on the molecular and behavioural level only in mammals and birds where the paternal care is explained as environmental influences being responsible for either caring or not caring (Vincze et al. 2013; Gleason and Marler 2013). Therefore, we are now facing a challenge of puzzling out the genomic pathways underlying biparental care in Hymenoptera and maybe even in insects.

Objectives of the thesis

Understanding of programs and software needed for successful analysis of the genome of *Ceratina nigrolabiata*. The whole genome analyses is demanding and thus it is one of the main objectives of my thesis.

Understanding genes and pathways in the individuals exhibiting traits similar to our species to understand the probable pathways underlying the parental behaviour in insects is needed.

Identifying regions in the genome which could be tied to the genomic pathways underlying the parental (biparental) care in *Ceratina nigrolabiata* need to be targeted for future transcriptomic analyses that will tell us more about these traits.

Biparental care in Hymenoptera and other animals

The concept of parental care is not unknown to the human kind. However common this behavioural trait is, the reasons behind it is actually quite a complicated process. Caring for offspring is an advantageous behaviour, that can improve the fitness of offspring (Tallamy and Wood 1986; Sargent, Taylor, and Gross 1987; Halffter, Huerta, and Lopez-Portillo 1996) either by nest guarding (Winkelman 1996; Vestergaard and Magnhagen 1993; Slack 1976) or prolonged parental care even in adult stages (Beekman, Thompson, and Jusup 2019; Edwards and Aschenborn 1989). Parents are however tied to the current brood and cannot reproduce again, therefore the cost is high and has to be outweighted by the benefits that rise from the care for the current clutch (Hamilton 1964). When the advantages of parental care are outweighed by the disadvantages, parents can leave their offspring and reproduce again (Williams 1966; Székely et al. 1996).

Biparental care is quite a common trait in mammals (Mammalia) (Stockley and Hobson 2016) and common in birds (Aves, Archosauria, Reptilia) (Lack 1968). Less so can be seen in fishes (Actinopterygii) (Goodwin, Balshine-Earn and Reynolds 1998; DeWoody et al. 2000), amphibians (Lissamphibia) (Brown, Morales, and Summers 2010; Jungfer and Weygoldt 1999) and in some crocodiles (Archosauria, Reptilia) (Tullberg, Ah-King, and Temrin 2002).

In insects, cosidering parental care, usually maternal care is present and paternal is not. Maternal care is rather common in insects of taxa Dermaptera (Staerkle Michael and Kölliker Mathias 2008; Radl and Linsenmair 1991), Hemiptera (Stegmann and Linsenmair 2002; Zink 2003; 2005; Hanelová and Vilímová 2013; Guershon and Gerling 2001; Taylor 2017), Embioptera (Edgerly 1987, Ross 2011), Psocoptera (New 1973), Coleoptera (Jordal, Sequeira, and Cognato 2011; Kirkendall 1997; Kavanaugh 1998) and Orthoptera (West and Alexander 1963). In Blattodea, maternal care is highly tied to the care for an ootheca (Baaren et al. 2003, Nalepa and Bell 1997). Strictly paternal care is present in Hemiptera: Belostomatidae (Estévez and Ribeiro 2011), Cydnidae (Filippi et al. 2009; Filippi-Tsukamoto et al. 1995), Acanthosomatidae (Kaitala and Mappes 1997) and Tingidae (Loeb and Bell 2006).

Biparental care is rare in insects. It can be found in Blattodea (Maekawa, Matsumoto, and Nalepa 2008; Park, Grandcolas, and Choe 2002; Seelinger and Seelinger 1983; Nalepa 1990), Thripidae (Tallamy 2001), Coleoptera: Passalidae (Schuster and Schuster 1997), Silphidae: *Nicrophorus spp.* (Fetherston, Scott, and Traniello 1990; Trumbo 1990), Scarabeidae (Hunt and Simmons 2002; Sowig 1996) and Scolitydae (Kirkendall 1997).

Until the end of the 20th century, the concept of biparental care was unknown in Hymenoptera. Maternal care is, however, not an uncommon trait (Matthews and Matthews 2009; Matthews 1968; Lucas et al. 2011; Linksvayer and Wade 2005; Rehan and Richards 2010; Boraschi, Peruquetti, and Del Lama 2005). Biparental care recorded in hymenoptera so far is present in *Trypoxylon* (Brockmann 1992; Brockmann and Grafen 1989), *Ceratina nigrolabiata* (Mikát et al. 2019) and other unpublished species (Mikát et al., unpublished).

Apart from maternal and biparental care, there are male specific behaviours tied to nesting. In *Microstigmus nigrophthalmus* (Crabronidae, Hymenoptera), males exhibit a guarding behaviour by chasing potential intruders away from the nest and therefore indirectly protect the colony, since the behaviour is probably tied to defence (Lucas and Field 2011). Not only in crabronids but also in bumblebee *Bombus griseocollis*, males partake in the nest cycle by incubating pupae the same way the females do and can rise the temperature up to 6 degrees of Celsius (Cameron 1985). We also know of male feeding the larvae in *Polistes fuscatus* and *P. metricus* (Vespidae, Hymenoptera), where males malaxate the food before feeding it to larvae (Hunt and Noonan 1979). This feeding behaviour can also be seen in *Rhopalidia marginata* (Vespidae, Hymenoptera) where the males feed larvae as well (Sen and Gadagkar 2006). In all these cases males are rather siblings than parents. Unfortunately, no further genomic studies that could help us evaluate the behaviour on genomic level were done on these species.

Genus *Ceratina*

Ceratina is a genus of small carpenter solitary bees occasionally demonstrating subsocial behaviour. They are a part of the subfamily Xylocopinae together with carpenter bees such as Xylocopini as a sister group, then Manuellini and two sister taxa Ceratinini and Allodapini (Shell and Rehan 2017) (Fig. 1), all in the family Apidae, Apoidea, Hymenoptera.

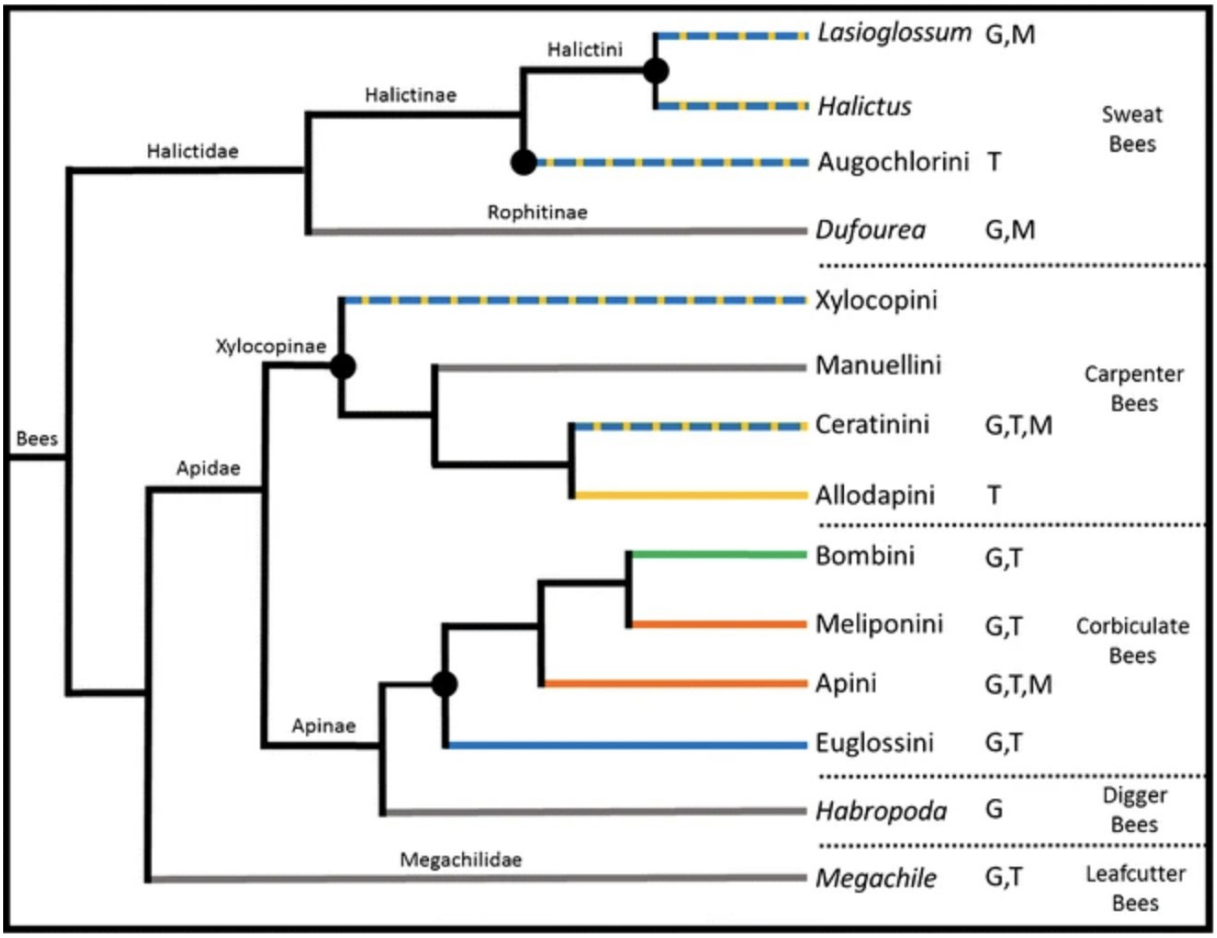

Fig. 1: The phylogenetic tree of Apoidea (Hymenoptera). Illustration of a phylogenetic relationship between sweat bees (**Halictidae**): sister taxa to the rest; Leafcutter Bees (**Megachilidae**): sister taxa to Apidae, which divide into Corbiculate and Digger Bees (**Apinae**) and Carpenter Bees (**Xylocopinae**).
Xylocopinae: sister group **Xylocopini** to Manuellini, Ceratinini and Allodapini, **Manuellini** sister group to Ceratinini and Allodapini, and **Ceratinini** with **Allodapini** creating a sister lineage. *(Shell a Rehan 2017)*

Ceratina nest in broken twigs with pith which they dig using their mandibules. Each cell is divided from each other with a partition made from the pith particles (Michener 2007). *Ceratina* provision their young with pollen (Rehan, 2020). On each pollen ball, the female lays an egg and closes the cell with a partition (Danforth et al. 2019). Each species has a specific pattern of cells in the nest and usually prefers to nest in different species of plants (due to the body size

of the *Ceratina* species) but can be often found in many twig habitats suitable to its needs (Rehan and Richards 2010). In nature, we can quite easily find *Ceratina* in places where there is a high possibility of disturbance (roads, landscape with big animals, pathways) (Fraňková, personal observation). For research, these disturbances are usually created artificially either by breaking the twigs on purpose or by collecting twigs and making sheaves (Mikát et al. 2019). *Ceratina* are small bees with solitary and social behaviour (Maeta a Sakagami 1995; Groom a Rehan 2018) ranging from North America (Rehan and Richards 2010) to Australia (Dew, Rehan, and Schwarz 2016). The research in past was mainly tied to Japanese *Ceratina* bees (Maeta, Saito, and Hyudo 1993; Sakagami 1985; Okazaki 1987) but recently the most studied species is *Ceratina calcarata* (Rehan and Richards 2010; Durant et al. 2016; Lawson, Helmreich, and Rehan 2017) and *Ceratina australensis* (Oppenheimer, Shell, and Rehan 2018; Rehan, Bulova, and O'Donnell 2015; Dew, Rehan, and Schwarz 2016; 2016; Rehan, Richards, and Schwarz 2010). Studying *Ceratina* bees is not demanding, the observations need to be done in their natural habitat because laboratory observations are not yet developed. However, since a twig represents a nest, the only limiting factor is the ability to analyse and dissect the nest, which needs practice and knowledge about the life history of the species.

In the Czech Republic, we have four species of *Ceratina*. *Ceratina cyanea*, *C. chalybea*, *C. cucurbitina* and *C. nigrolabiata* (Straka and Bogusch 2007). *Ceratina cyanea* is usually found across the Czech Republic and is the only species present in Bohemia. *Ceratina cucurbitina*, *C. chalybea* and *C. nigrolabiata* have a high abundance in Moravia, mainly in the South Moravian Region where most of the observations were conducted.

Behavioural characteristics of *Ceratina* bees

Species of the genus *Ceratinas* are variable, exhibiting many different behavioural characteristics. They are mostly solitary, but *Ceratina chalybea* (Mikát, Benda, and Straka 2020) is able to nest socially (where multiple females and males are present) and solitarily. Mother guards the nest or closes the nest with a thick partition and leaves it. *Ceratina chalybea* is also the only species with an alternation of empty and filled cells, this behaviour has antiparasitic properties. Other species capable of producing social nests are Cypriatic *C. bispinosa*, *C. dallatoreana*, *C. cypriaca* (endemic to Cyprus) where males are visitors or helpers from the previous brood, *C. parvula*, *C. mandibularis*, *C. chrysomala* and *C. moricei* (Mikát et al., unpublished). There are usually females and sometimes also male helpers in the nest. Social nests can have more cells provisioned than solitary nests. Social nests are usually longer and

more cells are present, this is true for *C. parvula*, and *C. mandibularis*. This is however different for *C. chalybea* social nests, where there is a lower number of cells. On the other hand, *Ceratina cucurbitina* is a strictly solitary species with a mother who exhibits subsocial behaviour by actively crawling through the nest and cleaning her offspring from feces (Mikát et al., unpublished). She guards and protects it against parasites and usurpation. Subsocial behaviour is also present in *C. cyanea*, where the mother is in contact with her offspring in the outermost cell, which is not closed. Other solitary species is *C. dallatoreana*, where the mother guards offspring until adulthood (Mikát et al., unpublished). Nests can be also left without a mother, however, knowing whether the nests are deserted or orphaned (in the case mother died) is diffuclt to establish (Mikát et al. n.d.). The only known species with parthenogenesis known is *Ceratina dallatoreana* (Daly 1966).

Ceratina calcarata is a well studied species with available transcriptome, genome and methylome. *Ceratina calcarata* is a solitary subsocial North American species, and founds a new nest each spring. The female actively controls the offspring by opening brood cells and cleaning her young which is the same behaviour as in *C. cucurbitina*. If a dead individual is present, the cell is cleaned and the material reused for creating new cell partitions. This subsocial behaviour provides an opportunity for mother and the brood to interact (Rehan and Richards 2010). Another interesting thing about *C. calcarata* is the production of the dwarf eldest daughter, which is the difference between *C. cucurbitina* and *C. calcarata*. This individual is fed less than her siblings (Lawson, Ciaccio, and Rehan 2016; Mikát, Franchino, and Rehan 2017) and therefore smaller. The dwarf eldest daughter partakes in offspring care by foraging for food to feed her siblings.

Ceratina australensis is an Australian endemic species exhibiting an insteresting socially polymorphic behaviour. *Ceratina australensis* is able to nest solitarily and socially. Throughout the winter, young adults overwinter and in the early spring, young females provision their nests. Nest reuse is also present, but newly constructed nests are more common. *Ceratina chalybea* produces social nests while reusing old nests which has not been observed in *C. australensis*. During early summer, almost all first brood nests are already constructed and the newly constructed nests probably hosts a second brood (Rehan, Richards, and Schwarz 2010). Since many second brood females have highly worn margins of wings, it is assumed that the same female produces two broods a year. This phenomenon could be present in Cypriac ceratinas and other *Ceratinas* inhabiting warmer climates. However, no data is available to support the hypothesis. The overwintering individuals are either worn (with worn margins of wings) or new, which are likely the second brood. As well as *C. calcarata* and *C. cucurbitina*, mothers

crawl through the nest to inspect the brood or guard the nest entrance with their abdomen (Rehan, Richards, and Schwarz 2010).

The species of our interest, *Ceratina nigrolabiata* (Mikát et al. 2019, Mikát et al., unpublished), is the only biparental species of *Ceratina* occurring also in the Czech Republic. Male and female are present in the nest in the majority of active brood nests observed. The function of male is to guard the nest while females are departed from the nest on a foraging trip. The behaviour of the male prevents a threat from natural enemies and parastitation of the nest and also a usurpation by another female. Males are long-living and can either guard the nest for a couple of days (about 7) and then depart or they can guard the nest up to more than 30 days. The abscence of male results in lower offspring production and shorter foraging trips of the female. *C. nigrolabiata* mates multiple times, therefore the male is guarding offspring of other males and his own as well. However, the average number of offspring related to the guarding male is only 0.638 (Mikát et al. 2019). *Ceratina nigrolabiata* nests are on avearage 14 cm long and have about 9 cells, however, there is a high variability observed in the number of provisioned cells. Empty cells are often present and are usually longer then provisioned cells. The pattern is rather random. All cells are enclosed and the thick partition at the entrance can also be present. Mothers do not crawl through partitions to check for parasites and to clean their offspring, although feces can be found in the next cell thanks to the perforated partitions between cells (Mikát et al., unpublished).

Hymenoptera genomics

The Reproductive Ground Plan hypothesis (Amdam et al. 2004) proposes the correlation between the behaviour of a foraging individual that is tied to the sensitivity to sucrose and reproductive pathways, and the division of labour in the colony (Scheiner, Page, and Erber 2001; Page, Erber, and Fondrk 1998). Worker bees are facultatively sterile and perform tasks related to their age based on the fluctuations of hormones (Scheiner, Page, and Erber 2001) The main drive in the behaviour of worker is the titres of *vitellogenin* in the hemolymph influencing behaviour driven by the titres of juvenile hormon (JH) in the bee. When the titres of JH are low, the production of vitellogenin is prominent (though still less than in the queen) and the hypopharyngeal glands, where the Major Royal Jelly Protein is made, are active (Rutz et al. 1976). In Honey bees, the transcription level and titter of vitellogenin differs between the young and old worker bees (Amdam et al. 2004). When the bee gets older and switches to a forager, the titre of JH is high and the production of vitellogenin is low, hypopharyngeal glands degenerate (Rutz et al. 1976) together with ovaries. The same findings while implementing a JH analog pyriproxyfen were confirmed by (Pinto, Bitondi, and Simões 2000).

To put it in complete perspective, young nursing bee has a high level of vitellogenin and partially active ovaries (and is able to lay trophic eggs), the forager, however, is reprived of vitellogenin and its ovaries degenerate together with hypopharyngeal glands (Engels and Fahrenhorst 1974; Pinto, Bitondi, and Simões 2000). This is however different in solitary bees, where both egg laying and foraging exists (Michener and Engels 1990). The situation in solitary bees could be explained by the oogenesis-flight syndrome, where the titres of vitellogenin lower in the flight phase and result in the reproduction diapause (Park and Stanley 2015).

Vitellogenin in insects is not naturally produced only in female but also in males – honey bee drones (Engels et al. 1990) or *Nicrophorus vespilloides* (Roy-Zokan et al. 2015). The only biparental insect where the research was done is *N. vespilloides*. Both male and female express vitellogenin, in higher doses while mated and in lower doses while caring for larvae. Vitellogenin receptor is also expressed in males and females with corresponding doses to the production of vitellogenin (Roy-Zokan et al. 2015). Therefore, changes in vitellogenin production are tied to the social changes in both male and female. This corresponding trend could mean, that the same pathways are tied in social insects, however, it does not explain the origin of biparental care or eusociality! It only tells us the current situation that drives the care in female and in male.

To further understand the genetic differences between the queen and the worker, we have to look at the differences occurring throughout the development of nurse and forager castes in *A. mellifera*. In foragers and nurses, the difference in mblk-1 transcription factor is high since it has only been recorded in the nurse brain (Hernández et al. 2012). Mblk-1 transcription factor influences the neural activity of the mushroom body in the brain (Park, Kunieda, and Kubo 2003). KOG proteins (Eukaryotic Orthologous Groups), such as proteins influencing the structure of ribosome, biogenesis of ribosome and translation of genes were differentially expressed in foragers than in other castes of honeybee (namely glucose oxidase, α-amylase, α-glucosidase glucose dehydrogenase and glycerol-3-phosphate dehydrogenase), Hernández et al. (2012) implies its possible connection to the increased brain activity in the forager as a learning process and memorization which are essential for foraging.

The guard worker bee could be analogous to our guarding male in the nest. However, in honeybees, the guarding behaviour is driven by environmental shifts (Downs and Ratnieks 2000) and we cannot establish the reason behind this behaviour of the *C. nigrolabiata* males yet. Gene 14-3-3 and Cyp6a20, which influences aggression, is up-regulated in guard workers. Aggressive bees also have a lower brain mitochondrial activity and the down-regulation of oxidative phosphorylation pathways for Complex I, IV, and V, a major brain metabolism pathway, was largely present (Alaux et al. 2009). We know that the differences in gene expression in the brains of guard vs. nurse and forager are quite prominent (Cash et al. 2005) (Fig. 2), sadly, noone properly studies guards in the Honeybee because this behaviour is very short – term (Trumbo, Huang, and Robinson 1997).

Aggression in *Drosophila melanogaster* is driven by epistatic mechanisms and specific genes such as the sex determination factor Fruitless where its elimination lowers the aggression (Certel et al. 2010). Serotonin (5-HT) and neuropeptide F also influence the aggression, moreover, the silencing of neuropeptide F increases aggression in *Drosophila* (Dierick and Greenspan 2007). The overall difference between low aggression and high aggression lineages is in about 1539 transcripts, 1480 of those specific for both male and female aggression (Edwards et al. 2006).

However, not many studies regarding aggression and guarding behaviour on the genomic level have been done in Hymenopteran lineages, but we assume, that many analogous or orthologous genes underlay this behaviour in insect species. Although one interesting finding regarding aggression was discovered in ant *Harpegnathos saltator* (Gospocic et al. 2017). They presents a neuropeptide corazonin which high levels inhibit the production of vitellogenin in females that express aggression in order to gain the gammergate position (Gospocic et al. 2017).

Vitellogenin therefore supports aggressive behaviour. Looking at male-male interactions where aggression is prominent could also give us some insight. Although linked to sociality, male-male interactions between hymenopterans are common and may facilitate the potential of male for guarding the nest as does the male of *C. nigrolabiata* (O'Neill 1983; Cheng et al. 2003; Eickwort 1977; Barthell and Baird 2004; Mikát et al. 2019).

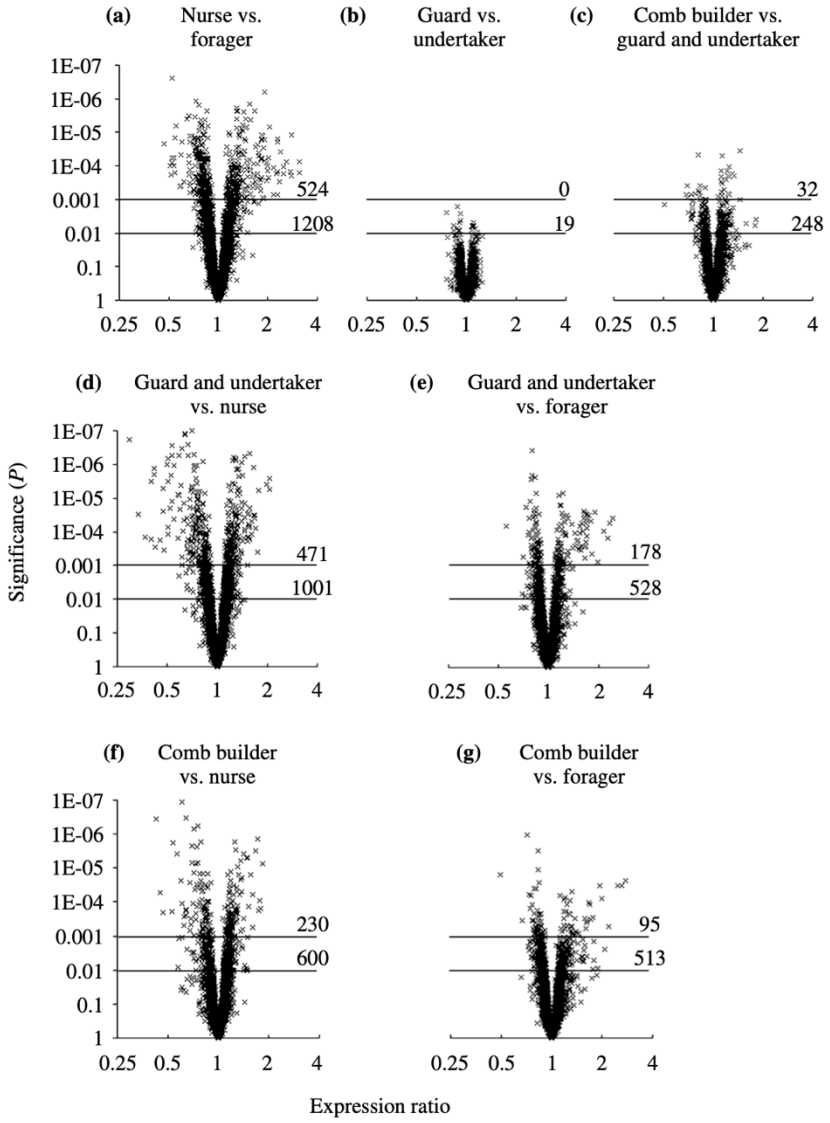

Fig 2. Gene expression in the brains of workers of different behavioural phenotypes. No significance in Guard vs. Undertaker and no difference for Comb builder vs. Guard and Undertaker. Significance found in Nurse vs. Forager and Guard and Undertaker vs. Nurse. Less significant difference found in Guard and Undertaker vs. Forager, Com builder vs. Forager and Comb builder vs. Nurse . *(Cash et al. 2005)*

An epigenetic mechanism widely influencing the differential expression in Hymenoptera is methylation (Kronforst et al. 2008). A genome-wide methylation requires several specific DNA methyl transferases (DNMTs) (Yi and Goodisman 2009). Those DNMTs required for a genome-wide methylation are DNMT1, DNMT2 and DNMT3. Some species, however, do not have active methylation present (Standage et al. 2016). This process can influence the behaviour of individuals from a colony (Amarasinghe, Clayton, and Mallon 2014; Sadd et al. 2015). The transcription factor activity is not only tied to DNMTs, but a huge number of Zinc fingers is present in eukaryotic organisms (Fedotova et al. 2017) and influence many proceses from DNA binding (Wolfe, Nekludova, and Pabo 2000) protein-protein interactions (Simpson et al. 2004) and are also able to recognise mRNA (Plambeck et al. 2003). Their function is tied to chromatin regulation (Peña et al. 2006), transcription regulation (Liew et al. 2005) and ubiquitination (Li et al. 2006).

Lipid metabolism is an important part of insect life in terms of reproduction, growth and functions as an energetic reserve for the unfavourable times (Arrese and Soulages 2010). The most common sugar found in insects is trehalose present in hemolymph in huge quantities (80 %) (Wyatt and Kalf 1957; Becker et al. 1996). Trehalose is not only an energy storage for insects and the mechanism guarding an intake of nutritions, but also functions as a stabilising protein during the thermal and osmotic stress (Thompson 2003). The alocation of energy resources is essential, because the quantities in the hemolymph differ during different phases of an insect life (Stein and Fell 1992). Insulin signalling pathway are influencing metabolism, reproduction, aging and metabolism (Wu and Brown 2006) and inositol monophosphatase mediates a synthesis of myo-inositol which is thought to be involved in lipid metabolism (Michell 2008). All those components of metabolism could be essential for the divergence of behaviour in males as well as other species of *Ceratina* since many of the metabolic tied genes are tied to caste differenciation and sociality (Hunt et al. 2010; Woodard et al. 2011) and to different activities (Harrison 1986). Odorant receptors are a part of the evolution of eusociality (Zhou et al. 2015) and are an essential drive of comunication (Tegoni, Campanacci, and Cambillau 2004; Pask et al. 2017). Odorant receptors are specific to the ecological demands of species. Such differences can be found in parasitoids, that need to locate their host in the environment (Sun et al. 2019; Robertson, Gadau, and Wanner 2010). The communication in *C. nigrolabiata* could seem to be in need of more complex structure because of the prolonged male-female interaction.

Materials and methods

Specimen collection

The individual was collected in 2018 in Havraníky Heathlands, Czech Republic, during a research of *Ceratina* species in NP Podyjí (NPP0188/2017).

DNA extraction and sequencing

The isolation was done in our laboratory with the Blood and Tissue Kit from the Qiagen company. We followed the instructions from Qiagen. The extracted specimen was enclosed in an Eppendorf tube and sent to the GENEWIZ company for sequencing. The DNA was sequenced on the next generation Illumina Platform NovaSeq 6000 for double stranded reads and 2*151 read length. According to approximate genome size of *C. calacrata* which is 199 Mbp (*Ceratina calcarata* BioProject accession number PRJNA299559) we targetted our sequencing data for 100x coverage. Amount of raw data from *C. nigrolabiata* count 33 Gbp.

Genome assembly

The reads were first controlled for quality with FastQC (Andrew 2010). FastQC is a program that estimates the quality of acquired reads based on the encoded quality value for each base. These values are automatically generated while sequencing the requested genome or genes and expresses the confidence of the presence of a particular nucleotide base during sequencing. Warning or failed reads are estimated based on a usual expectance of the quality of a normalised dataset.

For trimming, we used Trimmomatic (Bolger, Lohse and Usadel 2014) to trim the low quality reads and adapters in our fastq dataset. We used the pair-end approach with adapter clipping, and set the quality for the start and end of the read to 5. The minimal length used was 25 bp. After trimming, the reads were controlled again with FastQC (Andrew 2010) to check the quality of reads after the first trimming. Because some adapters usually remain in the dataset, we ran the second round of trimming with Trimmomatic to clean the dataset of the remaining adapters and trim other previously undetected low quality reads. The quality was once again checked with FastQC and estimated to be satisfactory for the next steps.

For error-correcting and extending reads, the Tadpole (Bushnell 2014) was implemented. Tadpole takes k-mers, which can be set up as needed. Both corrections, „pincer“ for the middle of reads and „tail“ for the ends, were used.

The normalisation of the genome was performed with BBNorm (BBMAP) (Bushnell 2014). The error-corrected tadpole output file was run with target 150 (an average depth of 150x) and min 0 to include all unique k-mers. The second run with target 30 and min 0. The normalised genome was checked for quality with FastQC (Andrews 2010).

MIRA (Chevreux 2007; Biswas, Ranjan, and Zubair 2011) is an assembling and mapping software for the whole genome annotation projects. Apart from Illumina sequencing, it is suitable for Sanger, GS20, FLX, Titanium, Ion Torrent and Pacific Biosciences sequences. Reads assembled into contigs by MIRA can be used in annotation projects. Even though MIRA is an old assembler, it works accurately and efficiently only for small data. Assembly to 30-40 coverage was produced with MIRA on COMMON_SETTINGS [output in oetg, 16 CPUs, sep (Skim each part) set to yes (= skim algorihm is called between each pass), cmrnl (check_maxreadnamelength) and cac (check average coverage) to no] and on SOLEXA_SETTINGS (msr (merge_short_reads) was set on no (default is yes), we also saved single simplets in project (sssip) and tagged singlets in project (stsip) producing a high coverage contigs. These contigs were manualy cleand up for non-nucleotide letters and replaced by Ns. The names of contigs remained unchanged. Second assembler, SPAdes (Bankevich et al. 2012) was loaded with MIRA contigs with k-mers 21, 33, 55, 77 and 99 on --only-assembler mode. The –careful mode was enabled to reduce the amount of short indels and mismatches. We used SPAdes to perfect our assembly and when compared, the combination of MIRA and SPAdes proved to be more efficient (even though slightly) than using only a single assembler. Missing BUSCOs (Seppey, Manni and Zdobnov 2019) for *C. nigrolabiata* to hymenoptera_odb9 BUSCO database was only 132 out of 4415 total BUSCOs searched which is 2,99% of BUSCOs missing.

Contigs were renamed and contigs with coverage lower than 2 were removed with GENEIOUS (Kearse et al. 2012) as it could be a possible contamination. Reads were mapped to contigs with MINIMAP2 (H. Li 2018) to produce SAM format and converted to BAM with SAMTOOLS (H. Li et al. 2009) and then sorted and indexed. Sorted and indexed files were concatenated.

Scaffolding was produced with BESST (Sahlin et al. 2014). Scaffolding is a technique that links contigs into a scaffold where contigs are separated by gaps which usually correspond to the real length of a missing sequences in between. This bridges the gaps between contigs and creates a somewhat continuous sequence and creates a so called scaffold (M. D. Cao et al. 2017). BESST was loaded with a contig file and a BAM file. Gaps in the scaffold which are produced during the scaffolding of assembled contigs were filled with GapCloser module of SOAPdenovo (Luo et al. 2012). All the remaining contigs below 200 bp were deleted.

RepeatModeler

To prepare the genome for annotation, the identification of transposable elements in our genome is necessary. First, the database using our genome was created for RepeatModeler (Smit and Hubley 2008) to use. RepeatModeler was then run with ncbi engine to create a file with identified transposable elements in the genome. The output was cleaned from redundant and higly similar sequences with cd-hit (Cluster Database at High Identity with Tolerance). This nonredundant sequence was blasted using blastx as a query towards the uniprot sprot blast protein database and *D. mellanogaster* blast database. All the output files were run through a custom Python script to clean up all additional repeats found with BLAST.

RepeatMasker

With cleaned genome from transposable elements, we had to tackle the interspersed repeats and low complexity DNA regions. These redundant elements can cause problems with gene prediction and often lead to false findings (Alkan, Coe, and Eichler 2011; Tørresen et al. 2019). The nonredundant genome was concatenated with arthropoda RepeatMasker library. RepeatMasker (Smit, Hubley and Green 2013-2015) was run on this library and produced two outputs, one with edited query sequence with masked (replaced by Ns) elements and a file with annotated repeats.

BRAKER

HISAT2-2.0.5. (Kim, Langmead and Salzberg 2015), a program for mapping next-generation sequencing reads, either DNA or RNA, to the reference genome, was used for creating an indexes for our masked genome with hisat2-build. This splits the genome into several smaller files with which the program works. After splitting the genome, hisat2 was run on our data with reference transcriptomes from two individuals obtained from NCBI. At this point, we did not have transcriptomic data from our *Ceratina nigrolabiata*, therefore *Ceratina calcarata* transcriptomes from newly enclosed two individuals were used (accession number: SRX3776755, SRR6819908). The output file from HISAT2 was converted from bam to sam using samtools and sorted.

Using BRAKER (Hoff et al. 2019) is a bit challenging. The installation proces requires many steps and adjustments. However, once installed correctly, the genome analyses is not that

complicated. At this point, BRAKER_v2.1.0. was implemented. BRAKER (Hoff et al. 2019) is a gene prediction pipeline implementing GeneMark (Ter-Hovhannisyan et al. 2008) and Augustus (Stanke et al. 2006). It uses protein homology information or RNA sequence to automatically train the GeneMark and Augustus for gene prediction given the RNA sequence data which does not have to be in high quality.

GeneMark is a program for gene prediction which uses all available information on input data such as the size of RNA sequence dataset, quality of both datasets and the amount of fragmentation in the assembly (GeneMark Manual). GeneMark uses Hidden Markov Model (HMM) for eukaryotic gene prediction, which takes the training set (in our case the provided sorted file from hisat2 run and masked file from RepeatMasker) and focuses on the gene structure elements such as introns and coding regions. These regions are analysed for the nucleotide content and assumed pairing of nucleotides in the training set. The base frequencies usually occurring in the genomes of eukaryotes are taken into consideration in predicting coding regions in our query dataset. The prediction dataset is taken and decoded. The model parameters are then set based on the gene structures from our RNA dataset and declared a parameter which is set by the Maximum Likelihood of the current parameter in the RNA dataset (Munch and Krogh 2006).

Another program for gene prediction, AUGUSTUS 3.3.3. (Stanke et al. 2006), is used in BRAKER_v2.1.0. Augustus is pretrained on many organisms ranging from prokaryotes to eukaryotes. However, it can also be used for ab initio prediction. Augustus uses Generalized Hidden Markov Model (GHMM) which unlike HMM considers the output parameter not as one but as a string of finite length for a current parameter (generalized interval probability) (Xie et al. 2016; Kulp et al. 1996; Stanke, 2003). Augustus also makes the alternative-transcipts option available and thus providing a more successful training. BRAKER_v2.1.0 produces a gff3 files with predictions from augustus and genemark-ex.

Transcriptome preparation

Trinity

Using Trinity (Grabherr et al. 2011), we conductet a *de novo* transcriptome assembly from downloaded *Ceratina calcarata* transcriptomic data from 6 individuals (NCBI SRA accession number: SRR6819907, SRR6819908, SRR6819909, SRR6819910, SRR6819911, SRR6819912). The data was used to construct a transcriptome used for Trinity-GG run (Genome Guided).

Before Trinity-GG run, we ran hisat2-build (Kim, Langmead and Salzberg 2015) to create indexdes for the unmasked assembled genome of *Ceratina nigrolabiata*, hisat2 was run with the two *Ceratina calcarata* individuals obtained from NCBI (NCBI SRA accession number SRX3776755, SRR6819908). The output file was again converted from sam to bam and sorted. The sorted bam files were used for the trinity genome guided assembly of transcriptomes. In the next step, hisat2 was run again on different set of data (NCBI SRA accession number: SRR1284947) of a complete transcriptome of individual of *Ceratina calcarata*. We are using multiple input data to ensure the transcriptome output is as close to a completion as possible. Trinity GG was run again, this time on the sorted data from the second hisat2 run. The output file is later used in MAKER (Holt and Yandell 2011) gene prediction run.

PASA pipeline

In its full name, Program to Assemble Spliced Alignment (Haas et al. 2003; Campbell et al. 2006; Haas et al. 2008), is a program for modelling gene structures from spliced alignments. It can model either partial gene structures or complete gene structures. Apart from this, PASA also annotates deletions and additions of exons, models alternative splicing variants, splitting and merging genes, untranslated regions (UTRs) and models novel genes. Its advantage lies in the ability to identify antisense transcripts, identify and also classify splicing variations found in the genome, map polyadenylation sites to the query sequence and to track down all partial and fill-length protein-coding genes with ab initio prediction.

PASA also incorporates other software into its pipeline, such as seqclean which finds the poly-A ends in the sequence and trimms the sequences which quality is below set threshold. Two engines for mapping aligning transcripts to genome, GMAP (Genome Mapping and Alignment Program) (T. D. Wu and Watanabe 2005) and BLAT (BLAST-Like Alignment Tool) (Kent 2002) is a part of the pipeline as well. With the usage of GMAP (Wu and Watanabe 2005), the alignment is fast and memory efficient. Therefore highly improving the run of a pipeline, which can be critical and challenging. The main advantage of GMAP lies in its memory efficiency with only as much as 128 MB of RAM needed. Apart from that, GMAP can align cDNAs (either mRNA or ESTs) with alternative splicing without the probabilistic splice site models and therefore its usage is not tied to a specific organism but can be used across the tree of life. Small exons (microexons) which would otherwise be discarded will be included in the mapping if they are statistically significant (T. D. Wu et al. 2016).

BLAT (Kent 2002) is another mapping engine used in PASA for the alignment of closely related species. It pre-indexes the genome and a protein database and does not store the genome itself during each run. It makes BLAT highly efficient and precise with finding matches which have up to 95% similarity and needs just 2 GB of RAM, therefore it can be run on an average machine. To align the sequence to the genome, BLAT uses 11-mers to quickly find candidate regions for the alignment. These homologous regions are taken and worked with separately. This requires less computation resources and less time and the K-mers can be reduced to a desired number for a more precise alignment (Bhagwat, Young and Robison 2012).

To further analyse transcriptomes and create a functional annotation, we modified PASA alignAssembly configuration file to suit our needs. Since we had problems with an online MySQL (MySQL 2001) application we chose an alternative approach with SQLite (Bhosale, Patil, and Patil 2015) which is a local library and does not need an access to the world wide web. Database from our *C. nigrolabiata* genome using SQLite was created.

All transcriptomes used were checked with a custom python script to make sure the names of transcripts were unique and there are no overlaps. Transcriptomes of *C. calcarata* constructed with Trinity Genome Guided assembly and the genome of *C. nigrolabiata* were concatenated into one file and the file was extracted from transcript accessions with PASA perl script accession_extractor.pl. Poly-A ends were cleaned from the file with seqclean which is included in the PASA pipeline. The cleaned file was loaded into Launch_PASA_pipeline.pl script together with the *C. nigrolabiata* genome and run with the BLAT (Kent 2002) aligner on one CPU due to our usage of SQLite which is not able to parallelise the process. This slows the alignment therefore figuring out the usage of MySQL is needed for future larger datasets.

At this point, we had an available training dataset for the last step. The training dataset consists of transcriptome assembly which is analysed with TransDecoder (Haas and Papanicolaou 2018). TransDecoder identifies coding sequences in the transcriptome by a minimum length of open reading frame (ORF) present in a transcript and computes its log-likelihood score. The sequence is considered to be coding if the ORF is present in the first reading frame. The last step produces a gff3 file with aligned sequences and provides a sequence ID, type of the sequence (exon, intron, mRNA rRNA etc.), start and end of the sequence, e-value, + or – strand and phase (the position of the ORF (0 = first base, 1 = second base, 2 = third base), which indicates whether some bases before the start of the coding regions need to be removed).

MAKER2

MAKER (Cantarel et al. 2008) is an annotation pipeline for prokaryotic and eukaryotic data invented by Yandell lab. MAKER2 (Holt and Yandell 2011) is an upgraded version of MAKER (which is capable of *de novo* annotation), however MAKER2 includes other *ab initio* prediction tools, which perfects the annotation. MAKER2 needs only a small cluster of training data for gene prediction and is able to analyse genomes of various sizes thanks to the support of MPI parallelisation which is needed for large genomes for example such as those of amphibians (Organ et al. 2011). It is an easy to use software with built in installation executables and therefore is suitable even for beginners in bioinformatics. Maker uses SNAP (Korf 2004), Augustus (Stanke et al. 2006), BLAST (Altschul et al. 1990), RepeatMasker (Smit, Hubley and Green 2013-2015) and other software included in its pipeline. For our run, only BLAST, RepeatMasker, Exonerate (Slater and Birney 2005), interproscan, SNAP and Augustus was used.

First we reformated the gff3 file with a custom python script and looked for matches in IDs which makes it easier to sort and read. We created a new sorted file.

This sorted file was loaded to maker with -fix_nucleotides parameter to fix any letters that are not ACTG and replace them with N's. Maker was run again on paralelisation to complete round one of coding regions prediction. Indexes found were merged with the genome file, gff3 prediction file for all found genes and proteins, and gff3 gene prediction file. The protein file was checked with BUSCO towards the hymenoptera_odb9 database. We had 389 missing buscos out of 4415 total searched. Therefore we aimed for a second maker run to perfect the prediction.

For identifying the protein domains we used InterProScan (Quevillon et al. 2005). We used the *C. nigrolabiata* protein file to find the GOterms (Gene Ontology) for our genome. Results shown in the result section.

From the InterProScan output file of proteins, we took only the unique finds and sorted them. The file with gff3 prediction file for coding regions prediction was concatenated with the unique finds from InterProScan so we have a complete file of found proteins and protein families found.

Round two of maker predictions was run on the InterProScan prediction file and indexes found were merged again with the genome file, gff3 prediction file for all found genes and proteins and gff3 gene prediction file. IDs and GFF IDs were mapped.

BLAST

Gene prediction file was BLASTed (Kent 2002) towards the nt (nucleotide) database downloaded with update_blastdb.pl perl script, which is incorporated into BLAST. The output file was filtered to remove low e-value genes found (e-value < 0.00009). We selected top hits from the filtered file for the purpose of the thesis.

Results

Assembly

After the assembly of the genome, BUSCO (Seppey, Manni and Zdobnov 2019) was implemented to determine the number of missing BUSCOs. For *C. nigrolabiata* to hymenoptera_odb9 BUSCO database there was only 132 missing out of 4 415 total BUSCOs searched which is 2.99% of BUSCOs missing (Table 1). Missing BUSCOs for *Ceratina calcarata* (BioSample accession number: SAMN04252592) were higher, 173, however only 99 frangmented BUSCOs and complete and single-copy BUSCOs were present (Table 1) which is much less than in our assembly. Therefore our genome has more fragmented BUSCOs and less complete single-copy BUSCOs. The single copy BUSCOs towards the hymenoptera BUSCO database in *C. calcarata* was 92.7%, in *C. nigrolabiata* 85.4%. Our genome is lower in quality and this loss is probably tied to this fact.

	C. nigrolabiata	*C. calcarata*
Complete BUSCOs (C)	3 782	4 143
Complete and single-copy BUSCOs (S)	3 770	4 096
Complete and duplicated BUSCOs (D)	12	47
Fragmented BUSCOs (F)	501	99
Missing BUSCOs (M)	132	173
Total BUSCO groups searched	4 415	4 415

Table 1: BUSCO table summary of complete BUSCOs found, complete and single-copy BUSCOs found, complete and duplicated BUSCOs, fragmented buscos, missing buscos and the total BUSCO groups searched. The first half is a summary for *Ceratina nigrolabiata* towards hymenoptera_odb9 BUSCO database, second half is a summary for *Ceratina calcarata* towards hymenoptera_odb9 BUSCO database.

Our genome has 6 539 /scaffolds after the assembly in comparison to 46 222 scafolds in *C. calcarata* (Rehan et al. 2016) which is considerably less! The longest scaffold however is longer in our species, *C. nigrolabiata*, which is 66 3120 bp in comparison to *C. calcarata*, where the longest scaffold is 52 351 bp. The N50 of *C. calcarata* was 73 643, in our species the N50 is 70 145.

RepeatModeler and RepeatMasker redundant elements

All 15 953 redundancies found with RepeatModeler against *D. melanogaster* and 5 016 redundancies against uniprot database were deleted and stored separately. All repetetive elements found with RepeatMasker are shown in the Table 2 where the most of repetitive elements are unclassified as expected.

File name: Cnig.fasta
Sequences: 6 539
Total length 154 475 013
CG level: 42.41%
Bases masked: 11 443 744

	Number of elements	Length occupied	Percentage of sequence
SINEs:	24	1 520 bp	0.00 %
ALUs	0	0 bp	0.00%
MIRs	0	0 bp	0.00 %
LINEs	1 734	211 855 bp	0.14 %
LINE1	44	2 226 bp	0.00 %
LINE2	329	21 263 bp	0.01 %
L3/CR1	82	4 515 bp	0.00 %
LTR elements:	2 337	657 312 bp	0.43 %
ERVL	0	0 bp	0.00 %
ERVL-MaLRs	0	0 bp	0.00 %
ERV_classI	0	0 bp	0.00 %
ERV_classII	0	0 bp	0.00 %
DNA elements:	3 657	509 230 bp	0.33 %
hAT-Charlie	88	8 096 bp	0.01 %
TcMar-Trigger	15	1 115 bp	0.00 %
Unclassified:	44 556	8 295 541 bp	5.37 %
Small RNA:	131	23 223 bp	0.02 %
Satellites:	36	3 053 bo	0.00 %
Simple repeats:	35 5šě	1 469 648 bp	0.95 %
Low complexity:	5 767	288 732 bp	0.19 %

Table 2: Summary of found repetetive elements for *Ceratina nigrolabiata* genome with RepeatMasker. The number of sequences searched is 6539, total length of the sequences together is 154 475 013 bp with the level of GC content being 42.41 %. We masked 11443744 bp of sequences, which is 8.41% of the initial length. SINEs (Short Interspersed Nuclear Elements), LINEs (Long Interspersed Nuclear Elements), LTR elements (Long Terminal Repeat elements), ERVL (Endogenous Retro-virus Related), MaLRs (Mammalian Apparent LTR-retrotransposon). Highest number was in Unclassified elements.

The total masked bases is 7.41% (Table 2) and although these elements are essential for the organism (Shapiro and Sternberg 2005), it is preferable to strip the genome off of them for the annotation because they can cause false findings (Alkan, Coe, and Eichler 2011). Also, the data is stored in a separate file which can be used for the analyses of repetitive elements found in the genome.

BRAKER gene prediction

With BRAKER, we can check the structure of our genome with braker-produced files. This can give a general idea of the number of genes present in the genome with the number of start and stop codons, introns and exons. However, this is more for a general idea of the number of expected genes and the general length of intronic and exonic regions in the genome. The problem is that there are many duplicities since there can be found many variations in the intronic and exonic regions. Therefore the genes and transcripts found are the outcome of the duplicities as well. This table then serves a purpose of a control in between steps. If the numbers were really low, in our case, we would be looking for a mistake in our scripts. In our genome, we had 300 970 genes (taking into consideration the different variations found) (Table 3). This also informed that the analyses is running correctly.

Stop-codon	13 561
CDS	92 525
Intron	77 078
Exon	92 525
Start-codon	13 412
Gene	300 970
Transcript	302 505

Tab 3: BRAKER output summary file. Found elements of CDS (coding sequence), introns, exons, start-codons, stop-codons, genes and transcripts.

MAKER and BLAST gene predictions

Unique orthogroups

From MAKER, we got protein predictions from an InterProScan run for the significant proteins in our genome. In this file, we got 3 299 unique proteins found many of which are transcription

factors (TF). The majority (15/23) of Pfam reported unique orthogroups in *C. calcarata* were found also in our genome (Table 4).

PF00001	7 transmembrane receptor (rhodopsin family)
PF00046	Homeodomain
PF00078	Reverse transcriptase (RNA-dependent DNA polymerase)
PF00089	Trypsin
PF00096	Zinc finger, C2H2 type
PF00098	Zinc knuckle
PF00226	DnaJ domain
PF00341	PDGF/VEGF domain
PF00929	Exonuclease
PF01607	Chitin binding Peritrophin-A domain
PF05485	THAP domain
PF05970	PIF1-like helicase
PF07679	Immunoglobulin I-set domain
PF07690	Major Facilitator Superfamily
PF10545	Alcohol dehydrogenase transcription factor Myb/SANT-like

Table 4: Unique orthogroups found in *Ceratina nigrolabiata*. First column are Pfam accession numbers, second column specific orthogroups found.

With BLAST, we found 1 049 261 genes. A lot of them were of low value and were filtered out (e-value < 0.00009). Only 8 881 top-hits were taken out of the filtered file and analysed for the purpose of the thesis. We assume that our genome has a lower quality and therefore some genes might be missing. It is also possible that the gene prediction algorhythms were set to be too sensitive and therefore we missed a number of genes. Several candidate genes will be listed.

Yellow and Major Royal Jelly Proteins

Four copies of Major Royal Jelly Proteins/Yellow proteins (IPR017996) were detected in our species. Yellow-like proteins found from our tophits BLAST prediction were: *Eufriesea mexicana* protein yellow-like (LOC108552730), mRNA, *Ceratina calcarata* protein yellow (LOC108629389), three copies of yellow-like (LOC108626509) and yellow-like (LOC108629401). From the MRJP, *Ceratina calcarata* major royal jelly protein 1 (LOC108629418), transcript variant X2, mRNA; major royal jelly protein 1-like (LOC108626615), mRNA and major royal jelly protein 1-like (LOC108629504).

Vitellogenin and vitellogenin related genes

Vitellogenin and vitellogenin related genes are present in our genome: vitellogenin receptor (LOC108623179), mRNA; vitellogenin-like (LOC108623035), mRNA; vitellogenin-like (LOC108625909), mRNA.

Transcription modification mechanisms

DNA methyl transferases (DNMTs) found are DNMT1 (LOC108629831) and DNMT3A (LOC108627172). As expected, many E3 ubiquitin-protein ligases were found in the genome, some of them with transcription factor properties such as E3 SUMO-protein ligase PIAS1 (LOC108627000), E3 ubiquitin-protein ligase Bre1 (LOC108625380), E3 ubiquitin-protein ligase RING1 (LOC108632232), enhancer of polycomb homolog 1 (LOC108622233), KAT8 regulatory NSL complex subunit 2 (LOC108627782) and KAT8 regulatory NSL complex subunit 3 (LOC108630754).

Most of the TFs are the usual zinc finger C2H2 type (IPR013087), zinc finger RING type (IPR001841) and THAP type zinc finger (IPR006612). Other zinc fingers in our genome were C4H2 domain-containing protein (LOC108622924), zinc finger CCCH domain-containing protein 10-like (LOC108631488), zinc finger CCCH domain-containing protein 11A-like

(LOC108623833), zinc finger CCCH domain-containing protein 13 (LOC108627873), zinc finger CCCH domain-containing protein 13-like (LOC108628954, LOC108632776), zinc finger CCCH domain-containing protein 14 (LOC108625667), zinc finger CCCH domain-containing protein 15 homolog (LOC108629546), zinc finger CCCH domain-containing protein 18-like (LOC108626140, LOC108626140, LOC108626140), zinc finger CCCH domain-containing protein 3 (LOC108630137), zinc finger CCCH-type with G patch domain-containing protein (LOC108627347), zinc finger CCHC domain-containing protein 10-like (LOC108625928), zinc finger CCHC domain-containing protein 2 (LOC108625428), zinc finger CCHC domain-containing protein 24-like (LOC108627709), zinc finger CCHC domain-containing protein 4 (LOC108623856), zinc finger CCHC domain-containing protein 8 homolog (LOC108628608), zinc finger CCHC-type and RNA-binding motif-containing protein 1-like (LOC108625119) Zinc finger RING type (IPR001841)

Metabolism related genes

Trehalose and trehalose-related genes found in our genome were trehalose transporter Tret1 (LOC108622238, LOC108630318) several copies of trehalose transporter Tret1-2 homolog(LOC108622481, LOC108622840, LOC108627635, LOC108627874, LOC113464173), several copies of facilitated trehalose transporter Tret1-like (LOC108622488, LOC108622655, LOC108623328, LOC108623915, LOC108624163, LOC108627678, LOC108627882, LOC108628450, LOC108628450, LOC108629180, LOC108631903, LOC113464927), gustatory receptor 5a for trehalose-like (LOC108624801)

Inositol monophosphatase genes and related genes found were putative inositol monophosphatase 3 (LOC108629481), inositol monophosphatase 3-like (LOC117213019), inositol monophosphatase-like (LOC108632425), inositol monophosphatase ttx-7-like (LOC108632419) and inositol monophosphatase derivate myo-inositol 2-dehydrogenase-like (LOC108623353).

Insulin interaction was present in duplications of insulin gene enhancer protein ISL-1 (LOC108626854), insulin receptor substrate 1 (LOC108623285), insulin receptor-like (LOC108622672), insulin-like growth factor 2 mRNA-binding protein 1 (LOC108628931).

Odorant receptors

We found only 11 odorant receptors which in comparison to 16 odorant receptor families *C. calcarata*. Odorant receptor 13a-like (LOC108625197, LOC108626998, LOC108627089,

LOC108627093, LOC108627129, LOC108627605, LOC108631820), odorant receptor 4-like (LOC108624242, LOC108624267, LOC108625415) odorant receptor 46a-like (LOC113464013), odorant receptor 49a-like (LOC113464848), odorant receptor 49b-like (LOC108630531, LOC108630531), odorant receptor 67c-like (LOC113465063), odorant receptor 82a-like (LOC108622449, LOC108624260, LOC113464235), odorant receptor 85b-like (LOC108624245), odorant receptor Or2-like (LOC108624238, LOC108631828), odorant receptor 19b (LOC113463995), odorant receptor 7tm (PF02949) were found.

Conclusion and discussion

The genome was sequenced with the next generation approach. These techniques are becoming slowly overcome by the third generation sequencing techniques such as Oxford Nanopore technologies (Cao et al. 2017). For our future genomic studies, we will sequence our genomes with newer technologies since they are being more affordable and the assembly from long reads is much easier and accurate. Also, in Hymenoptera, males are usually sequenced because their haplodiploidity makes it easier for assembly (Lu, Giordano, and Ning 2016). We opted for female instead which could result in false duplicities in case our assembly was not done properly. We made sure our genomes are males for future studies and we will possibly sequence another *C. nigrolabiata* to get a better insight into the differences of assembling the diploid and haploid genome.

The genome of *C. nigrolabiata* was successfully assembled and annotated. Our genome was of a lower coverage, which may have resulted in an incomplete annotation of our genome. The comparison with another *Ceratina* species, *C. calcarata* (Rehan et al. 2016) showed the incompletness of our genome where the number of complete and single copy BUSCOs in our genome is lower and the number of fragmented BUSCOs is much higher than in *C. calcarata*. Therefore, trying a more suitable and newer assembly software and approaches is needed to achieve a better genome assembly where we will have less fragmented elements.

We cleaned the genome of low-complexity elements. Since we did not have RNA information (transcriptome) from our species, we opted to use a well sequenced species *C. calcarata*. We trained the gene prediction program Augustus with our genome and downloaded transcriptome data and successfully predicted coding regions and their structure. To predict other coding regions that could occur in the genome, we assembled downloaded transcriptomes from *C. calcarata* individuals and used for further detection of coding regions in our genome. With the training dataset, we analysed the genome and created a file with found and predicted coding sequences. These sequences were BLASTed towards the NCBI nucleotide database. The gene file from BLAST with identified genes was filtered from low e-values and from the duplicities, only top hits were taken. Those top-hits were analysed and several candidate genes were taken for further analyses. We found 15 out of 23 unique orthogroups previously discovered in *C. calcarata*. We presume our genome contains more shared unique orthogroups, however, they were not identified in our analyses.

Our gene dataset from BLAST is smaller than that of *C. calcarata*. There are several explanations, either the genome contains less genes than *C. calcarata*, the genome was not

properly sequenced, the assembly was not precise or the annotation came across a problem because we do not have transcriptome from our species and used a north american species, *Ceratina calcarata*.

From the genes analysed, we focused mainly on transcription factors – zinc fingers, which are essential for a differential expression which we will be analysing in the future. We expect significant differences taking part in the different parts of the nesting season and between *C. nigrolabiata* males that guard the nest for a short time and males that guard a long time. Apart from transcription factor activity, we expect significant differences in the metabolism in individuals in different phases of the nesting season, therefore, we focused also on the genes that take part in mediating and influencing the lipid metabolism, such as trehalosa, inositol monophosphatase and insulin pathways. Insulin takes part in many different processes in insects, from a stimulatory effect on the oogenesis (Brown et al. 2008) to the metabolic uptake driven by insulin signaling pathways which influence the proliferation of cells (Drummond-Barbosa and Spradling 2001). The vitellogenesis is also stimulated by insulin observed in Diptera (Roy, Hansen, and Raikhel 2007). With insulin influencing the pathways in many different processes, we expect insulin to be tied to vitellogenesis in females and males in case of vitellogenin presence in males. Inositol monophosphatase as mentioned above mediates a synthesis of myo-inositol. This product being involved in lipid metabolism (Michell 2008) is present in our species and we expect to see differences in its occurence in our species during different phases of the season and between different behavioural cohorts mentioned. The enrichment of inositol monophosphatase gene family in *L. albipes* is suggested to be tied to the diapause in adult foundresses (Kocher et al. 2013) therefore it is an active proces mediating the usage of energy resources. Many duplications of trehalosa transporters are present which is expected, since trehalosa facilitates the membrane comunication, energy and carbon uptake and even has a signaling function (Kanamori et al. 2010). We presume there will be differences between different metabolic levels in our cohorts for the future transcriptional analyses and looking for the levels of different sugars and transporters might give us an insight into the energy demand each cohort has. Because males guarding the nest allows the female longer foraging trips, we expect to see a higher metabolism activity in those females. We also think there might be a higher metabolism activity in males guarding the nest for a short time since they are more active regarding the flight activity when migrating from the nest unlike the long term guarding males, that stay in the nest and have to invest into saving the energy. We however do not know, whether males feed in the nest or not.

DNA methyl transferase (DNMT) in studied genome found are DNMT1 and DNMT 3A. The DNMT1 is responsible for the methylation of cytosine and is a part of gene silencing processes in mammals (Tate and Bird 1993). The processes are involved in reproduction and therefore an organism with DNMT1 knocked out posttranscriptionally is not able to reproduce (Bewick et al. 2019). We assume that active methylation process is present in our species since DNMT1 can be missing in insect species and those species do not posses methylation processes (Lyko and Maleszka 2011; Standage et al. 2016). The DNMT3A, on the other hand, is a methyltransferase regulating learning, synaptic plasticity and memory in the brain (Feng et al. 2010; 2005). Because our species forages for food and actively navigates through the habitat, learning and memory abilities are essential.

The E3 SUMO-protein ligase uses a process called sumoylation which is similar in terms of effects to ubiquitination (Geiss-Friedlander and Melchior 2007) which flags desired element which is afterwards targeted and disassembled. The E3 ubiquitin-protein ligase Bre1 plays an important role in H2B histone tagging for epigenetic transcriptional activation, telomeric silencing, elongation by RNA polymerase II and the modulation of the formation of double strand brakes during meiosis (Xiao et al. 2005; Yamashita, Shinohara, and Shinohara 2004; Wood et al. 2003). E3 ubiquitin-protein ligase RING1 acts as an epigenetic tag on H2A histone therefore influencing the transcription by repression in mammals (Cao, Tsukada, and Zhang 2005). An opposite process, acetylation, which enhances or enables the transcription of a certain gene (Verdone, Caserta, and Di Mauro 2005) is mediated by the enhancer of polycomb homolog 1, which acts in the acetylation of histon H4 and H2A, therefore regulating transcription of specific genes present in organisms from yeast to human (Doyon et al. 2004). This enhancer is a critical structure for gene regulation and is also associated with some diseases (Searle and Pillus 2018).

Found KAT8 regulatory NSL complex subunit 2 is another acetylation mediator in histone H4. It is an essential process regulating house-keeping genes expression (Pavlova et al. 2019) and its absence can cause lethal effects.

Apart from the unique orthogroups, we found 47 zinc finger coppies, in comparison to *C. calcarata* (Rehan et al. 2016) with 50 zinc finger copies, 16 copies in *L. albipes* and 2 in *A. mellifera* (Kocher et al. 2013). The zinc finger C2H2 TF is abundant in many eukaryotic organisms, however, the principle of zinc finger C2H2 is still poorly understood (Fedotova et al. 2017). This type of zinc finger was reported to be unique in *C. calcarata* genome Rehan et al. 2016) in comparison to other bee genomes. We report the presence of zinc finger C2H2 in our genome as well. Known influence of the C2H2 type is protein folding and assembly, lipid

binding, RNA packaging, DNA recognition and transcriptional activation, cell apoptosis and proliferation (Laity, Lee, and Wright 2001). The most interesting is not their DNA binding however, but the interaction with proteins (Polekhina et al. 2002). The protein-protein interaction is documented in Zinc finger RING type (IPR001841) as well (Saurin et al. 1996). Second most common zinc finger was the THAP type zinc finger (IPR006612), a sequence-specific DNA-binding zinc finger (Bessière et al. 2008).

The Major Royal Jelly Proteins/Yellow proteins are present in our species as well (IPR017996). Those proteins are present only in insects and some bacteria (Ferguson et al. 2011). These genes are thought to be associated with melanic pigmentation (Wittkopp, Vaccaro, and Carroll 2002; Gompel et al. 2005) and also in the formation of chorion and cuticule of insects (Li and Christensen 2011). In *Apis mellifera*, ten yellow genes were found and clustered as Major Royal Jelly Proteins (MRJPs) influencing eusocial behaviour (Drapeau et al. 2006). The MRJP1 fed to the larvae increases its body mass and supports the ovary proliferation (Kamakura 2011) therefore producing caste-specific morphotypes. Apart from MRJP1, MRJP2 complex acts as an antimicrobial unit and possesses antioxidant properties in the royal jelly produced by workers (Park et al. 2019). In our genome, we found 4 copies of MRJP/yellow protein and we assume it to have properties regarding melanisation and provisioning of larvae mainly. We also expect to see the production of Major Royal Jelly Proteins 1 and 2 in the females. The deposition of pollen balls in the nest requires antimicrobial properties to stay edible for the larvae. The female offspring also mate without exception since no workers in *C. nigrolabiata* are present, therefore the presence of MRJP1 could influence the ovary proliferation.

Odorant receptors were thought to be in a higher abundance in social species (Wilson 1971) and although some research support the thesis (Yan et al. 2017), a parasitoid wasp *Nasonia vitripennis* has a higher number of odorant receptors (Robertson and Wanner 2006) than *A. mellifera* (Robertson, Gadau, and Wanner 2010). We expected to see a similar number of odorant receptors as in *C. calcarata*, which is 16 (Rehan et al. 2016). However, we thought that the occurence of male in the nest could pose a challenge for the communication and therefore, the number of odorant receptors would be higher. This hypothesis seems to be wrong since only 12 odorant receptors were found in *C. nigrolabiata* and more insight into the comunication of the species could be splendid. We assume that the reason might be either the lower quality of our genome or there indeed is a contraction of the odorant receptor gene families tied to the ecology of the species. For comparison,

Thanks to the study of (Roy-Zokan et al. 2015), we focus on *vitellogenin* and vitellogenin-receptor in our species. We found both, *vitellogenin* and vitellogenin-receptor genes in our

genome. We expect to see the activity of *vitellogenin* in the males and females of *C. nigrolabiata*, which will most probably fluctuate throughout the nesting season as well. We did not find any discussed aggression related genes in our species.

While genomics is a tool that helps us to understand processes on the molecular level, we must not forget, that the basis of biological sciences is a complicated and still poorly understood field. It is important to seek answers to our hypothesis not only in one aspect but we have to consider the ecology and evolution of the species as well. I would like to use this opportunity and stress out the importance of the taxonomists, ecologists, ethologists and field scientists for the science. We are now in a period, where these fields, mainly taxonomists and field scientists, are needed and the surge of molecularly oriented scientists is high. Also, as I experienced, the proper use of genomics takes a lot of practice and experience and we have to use it correctly. Learning the genomic methods on our own is a demanding and difficult process and without proper guidance usually leads to incorrect results. There is not enough resources for individuals considering this path in their career and they often have to seek help somewhere else. I hope that we will continue to improve the situation and that genomic techniques and methods will be more available in the future for the students and the scientists likewise.

References:

Alaux, Cédric, Saurabh Sinha, Linda Hasadsri, Greg J. Hunt, Ernesto Guzmán-Novoa, Gloria DeGrandi-Hoffman, José Luis Uribe-Rubio, Bruce R. Southey, Sandra Rodriguez-Zas, and Gene E. Robinson. 2009. 'Honey Bee Aggression Supports a Link between Gene Regulation and Behavioral Evolution'. *Proceedings of the National Academy of Sciences* 106 (36): 15400–405. https://doi.org/10.1073/pnas.0907043106.

Alcock, John. 1975. 'Territorial Behaviour by Males of Philanthus Multimaculatus (Hymenoptera: Sphecidae) with a Review of Territoriality in Male Sphecids'. *Animal Behaviour* 23, Part 4 (November): 889–95. https://doi.org/10.1016/0003-3472(75)90113-X.

Alkan, Can, Bradley P. Coe, and Evan E. Eichler. 2011. 'Genome Structural Variation Discovery and Genotyping'. *Nature Reviews Genetics* 12 (5): 363–76. https://doi.org/10.1038/nrg2958.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. 'Basic Local Alignment Search Tool'. *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Amarasinghe, Harindra E., Crisenthiya I. Clayton, and Eamonn B. Mallon. 2014. 'Methylation and Worker Reproduction in the Bumble-Bee (Bombus Terrestris)'. *Proc. R. Soc. B* 281 (1780): 20132502. https://doi.org/10.1098/rspb.2013.2502.

Amdam, Gro V., Kari Norberg, M. Kim Fondrk, and Robert E. Page. 2004. 'Reproductive Ground Plan May Mediate Colony-Level Selection Effects on Individual Foraging Behavior in Honey Bees'. *Proceedings of the National Academy of Sciences* 101 (31): 11350–55. https://doi.org/10.1073/pnas.0403073101.

Andrews, Simon. 2010. *FastQC: A Quality Control Tool for High Throughput Sequence Data*.

Arrese, Estela L., and Jose L. Soulages. 2010. 'Insect Fat Body: Energy, Metabolism, and Regulation'. *Annual Review of Entomology* 55 (1): 207–25. https://doi.org/10.1146/annurev-ento-112408-085356.

Baaren, J. van, A.-S. Bonhomme, P. Deleporte, and J.-S. Pierre. 2003. 'Behaviours Promoting Grouping or Dispersal of Mother and Neonates in Ovoviviparous Cockroaches'. *Insectes Sociaux* 50 (1): 45–53. https://doi.org/10.1007/s000400300007.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. 'SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing'. *Journal of Computational Biology* 19 (5): 455–77. https://doi.org/10.1089/cmb.2012.0021.

Barthell, John F., and Troy A. Baird. 2004. 'Size Variation and Aggression among Male Xylocopa Virginica (L.) (Hymenoptera: Apidae) at a Nesting Site in Central Oklahoma'. *Journal of the Kansas Entomological Society* 77 (1): 10–20. https://doi.org/10.2317/0301-27.1.

Becker, A., P. Schlöder, J. E. Steele, and G. Wegener. 1996. 'The Regulation of Trehalose Metabolism in Insects'. *Experientia* 52 (5): 433–39. https://doi.org/10.1007/BF01919312.

Beekman, Madeleine, Michael Thompson, and Marko Jusup. 2019. 'Thermodynamic Constraints and the Evolution of Parental Provisioning in Vertebrates'. *Behavioral Ecology* 30 (3): 583–591.

Beinart, R. A., J. Rotterová, I. Čepička, R. J. Gast, and V. P. Edgcomb. 2018. 'The Genome of an Endosymbiotic Methanogen Is Very Similar to Those of Its Free-Living Relatives'. *Environmental Microbiology* 20 (7): 2538–51. https://doi.org/10.1111/1462-2920.14279.

Berens, Ali J., James H. Hunt, and Amy L. Toth. 2014. 'Comparative Transcriptomics of Convergent Evolution: Different Genes but Conserved Pathways Underlie Caste Phenotypes across Lineages of Eusocial Insects'. *Molecular Biology and Evolution* 32 (3): 690–703.

Bessière, Damien, Chrystelle Lacroix, Sébastien Campagne, Vincent Ecochard, Valérie Guillet, Lionel Mourey, Frédéric Lopez, et al. 2008. 'Structure-Function Analysis of the THAP Zinc Finger of THAP1, a Large C2CH DNA-Binding Module Linked to Rb/E2F Pathways'. *Journal of Biological Chemistry* 283 (7): 4352–63. https://doi.org/10.1074/jbc.M707537200.

Bewick, Adam J., Zachary Sanchez, Elizabeth C. Mckinney, Allen J. Moore, Patricia J. Moore, and Robert J. Schmitz. 2019. 'Dnmt1 Is Essential for Egg Production and Embryo Viability in the Large Milkweed Bug, Oncopeltus Fasciatus'. *Epigenetics & Chromatin* 12 (1): 6. https://doi.org/10.1186/s13072-018-0246-5.

Bhagwat, Medha, Lynn Young, and Rex R. Robison. 2012. 'Using BLAT to Find Sequence Similarity in Closely Related Genomes'. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]* 0 10 (March): Unit10.8. https://doi.org/10.1002/0471250953.bi1008s37.

Bhosale, S. T., T. Patil, and P. Patil. 2015. 'Sqlite: Light Database System'. *International Journal of Computer Science and Mobile Computing* 4 (4): 882.

Biswas, Abhishek, Desh Ranjan, and Mohammad Zubair. 2011. 'Parallelization of Mira Whole Genome and Est Sequence Assembler'. In *Workshop on Parallel Algorithms and Software for Analysis of Massive Graphs (ParGraph)*.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data'. *Bioinformatics* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Boore, Jeffrey L., and Joseph L. Staton. 2002. 'The Mitochondrial Genome of the Sipunculid Phascolopsis Gouldii Supports Its Association with Annelida Rather than Mollusca'. *Molecular Biology and Evolution* 19 (2): 127–137.

Boraschi, Daniele, Rui Carlos Peruquetti, and Marco Antônio Del Lama. 2005. 'Biologia, comportamento social e alocação sexual de Digelasinus diversipes (Kirby, 1882) (Hymenoptera, Argidae)'. *Revista Brasileira de Entomologia* 49 (2): 253–63. https://doi.org/10.1590/S0085-56262005000200010.

Borowiec, Marek L., Ernest K. Lee, Joanna C. Chiu, and David C. Plachetzki. 2015. 'Extracting Phylogenetic Signal and Accounting for Bias in Whole-Genome Data Sets Supports the Ctenophora as Sister to Remaining Metazoa'. *BMC Genomics* 16 (1): 1–15.

Bourlat, Sarah J., Claus Nielsen, Anne E. Lockyer, D. Timothy J. Littlewood, and Maximilian J. Telford. 2003. 'Xenoturbella Is a Deuterostome That Eats Molluscs'. *Nature* 424 (6951): 925–28. https://doi.org/10.1038/nature01851.

Brockmann, H. Jane. 1992. 'Male Behavior, Courtship and Nesting in Trypoxylon (Trypargilum) Monteverdeae (Hymenoptera: Sphecidae)'. *Journal of the Kansas Entomological Society*, 66–84.

Brockmann, H.Jane, and Alan Grafen. 1989. 'Mate Conflict and Male Behaviour in a Solitary Wasp, Trypoxylon (Trypargilum) Politum (Hymenoptera: Sphecidae)'. *Animal Behaviour* 37 (2): 232–55. https://doi.org/10.1016/0003-3472(89)90113-9.

Brown, Jason L., Victor Morales, and Kyle Summers. 2010. 'A Key Ecological Trait Drove the Evolution of Biparental Care and Monogamy in an Amphibian'. *The American Naturalist* 175 (4): 436–446.

Brown, Mark R., Kevin D. Clark, Monika Gulia, Zhangwu Zhao, Stephen F. Garczynski, Joe W. Crim, Richard J. Suderman, and Michael R. Strand. 2008. 'An Insulin-like Peptide Regulates Egg Maturation and Metabolism in the Mosquito Aedes Aegypti'. *Proceedings of the National Academy of Sciences* 105 (15): 5716–21. https://doi.org/10.1073/pnas.0800478105.

Cameron, Sydney A. 1985. 'Brood Care by Male Bumble Bees'. *Proceedings of the National Academy of Sciences* 82 (19): 6371–6373.

Campbell, Matthew A., Brian J. Haas, John P. Hamilton, Stephen M. Mount, and C. Robin Buell. 2006. 'Comprehensive Analysis of Alternative Splicing in Rice and Comparative Analyses with Arabidopsis'. *BMC Genomics* 7 (1): 327. https://doi.org/10.1186/1471-2164-7-327.

Cantarel, Brandi L., Ian Korf, Sofia M. C. Robb, Genis Parra, Eric Ross, Barry Moore, Carson Holt, Alejandro Sánchez Alvarado, and Mark Yandell. 2008. 'MAKER: An Easy-to-Use Annotation Pipeline Designed for Emerging Model Organism Genomes'. *Genome Research* 18 (1): 188–96. https://doi.org/10.1101/gr.6743907.

Cao, Minh Duc, Son Hoang Nguyen, Devika Ganesamoorthy, Alysha G. Elliott, Matthew A. Cooper, and Lachlan J. M. Coin. 2017. 'Scaffolding and Completing Genome Assemblies in Real-Time with Nanopore Sequencing'. *Nature Communications* 8 (1): 14515. https://doi.org/10.1038/ncomms14515.

Cao, Ru, Yu-Ichi Tsukada, and Yi Zhang. 2005. 'Role of Bmi-1 and Ring1A in H2A Ubiquitylation and Hox Gene Silencing'. *Molecular Cell* 20 (6): 845–54. https://doi.org/10.1016/j.molcel.2005.12.002.

Cash, A. C., C. W. Whitfield, N. Ismail, and G. E. Robinson. 2005. 'Behavior and the Limits of Genomic Plasticity: Power and Replicability in Microarray Analysis of Honeybee Brains'. *Genes, Brain and Behavior* 4 (4): 267–71. https://doi.org/10.1111/j.1601-183X.2005.00131.x.

Certel, Sarah J., Adelaine Leung, Chih-Yung Lin, Philip Perez, Ann-Shyn Chiang, and Edward A. Kravitz. 2010. 'Octopamine Neuromodulatory Effects on a Social Behavior Decision-Making Network in Drosophila Males'. *PLOS ONE* 5 (10): e13248. https://doi.org/10.1371/journal.pone.0013248.

Cheng, Ling-lan, Ralph W. Howard, James F. Campbell, Ralph E. Charlton, James R. Nechols, and Sonny Ramaswamy. 2003. 'Behavioral Interaction Between Males of Cephalonomia Tarsalis (Ashmead) (Hymenoptera: Bethylidae) Competing for Females'. *Journal of Insect Behavior* 16 (5): 625–45. https://doi.org/10.1023/B:JOIR.0000007700.10615.35.

Chevreux, Bastien. 2007. 'MIRA: An Automated Genome and EST Assembler'.

Daly, Howell V. 1966. 'Biological Studies on Ceratina Dallatorreana, an Alien Bee in California Which Reproduces by Parthenogenesis (Hymenoptera: Apoidea)'. *Annals of the Entomological Society of America* 59 (6): 1138–54.

Danforth, Bryan N., Robert L. Minckley, John L. Neff, and Frances Fawcett. 2019. *The Solitary Bees: Biology, Evolution, Conservation*. Princeton University Press.

Dew, Rebecca M., Sandra M. Rehan, and Michael P. Schwarz. 2016. 'Biogeography and Demography of an Australian Native Bee Ceratina Australensis (Hymenoptera, Apidae) since the Last Glacial Maximum'. *Journal of Hymenoptera Research* 49 (April): 25–41. https://doi.org/10.3897/JHR.49.8066.

DeWoody, J. Andrew, Dean E. Fletcher, S. David Wilkins, William S. Nelson, and John C. Avise. 2000. 'Genetic Monogamy and Biparental Care in an Externally Fertilizing Fish, the Largemouth Bass (Micropterus Salmoides)'. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267 (1460): 2431–37. https://doi.org/10.1098/rspb.2000.1302.

Dierick, Herman A., and Ralph J. Greenspan. 2007. 'Serotonin and Neuropeptide F Have Opposite Modulatory Effects on Fly Aggression'. *Nature Genetics* 39 (5): 678–82. https://doi.org/10.1038/ng2029.

Ding, James, Antonios Frantzeskos, and Gisela Orozco. 2020. 'Functional Genomics in Autoimmune Diseases'. *Human Molecular Genetics*.

Downs, Stephen G., and Francis LW Ratnieks. 2000. 'Adaptive Shifts in Honey Bee (Apis Mellifera L.) Guarding Behavior Support Predictions of the Acceptance Threshold Model'. *Behavioral Ecology* 11 (3): 326–333.

Doyon, Yannick, William Selleck, William S. Lane, Song Tan, and Jacques Côté. 2004. 'Structural and Functional Conservation of the NuA4 Histone Acetyltransferase Complex from Yeast to Humans'. *Molecular and Cellular Biology* 24 (5): 1884–96. https://doi.org/10.1128/mcb.24.5.1884-1896.2004.

Drapeau, Mark David, Stefan Albert, Robert Kucharski, Carsten Prusko, and Ryszard Maleszka. 2006. 'Evolution of the Yellow/Major Royal Jelly Protein Family and the Emergence of Social Behavior in Honey Bees'. *Genome Research* 16 (11): 1385–94. https://doi.org/10.1101/gr.5012006.

Drummond-Barbosa, Daniela, and Allan C. Spradling. 2001. 'Stem Cells and Their Progeny Respond to Nutritional Changes during Drosophila Oogenesis'. *Developmental Biology* 231 (1): 265–78. https://doi.org/10.1006/dbio.2000.0135.

Durant, Dusty R., Ali J. Berens, Amy L. Toth, and Sandra M. Rehan. 2016. 'Transcriptional Profiling of Overwintering Gene Expression in the Small Carpenter Bee, Ceratina Calcarata'. *Apidologie* 47 (4): 572–582.

Edgerly, Janice S. 1987. 'Maternal Behaviour of a Webspinner (Order Embiidina)'. *Ecological Entomology* 12 (1): 1–11. https://doi.org/10.1111/j.1365-2311.1987.tb00979.x.

Edwards, Alexis C., Stephanie M. Rollmann, Theodore J. Morgan, and Trudy FC Mackay. 2006. 'Quantitative Genomics of Aggressive Behavior in Drosophila Melanogaster'. *PLoS Genet* 2 (9): e154.

Edwards, Penelope B., and H. H. Aschenborn. 1989. 'Maternal Care of a Single Offspring in the Dung Beetle Kheper Nigroaeneus: The Consequences of Extreme Parental Investment'. *Journal of Natural History* 23 (1): 17–27. https://doi.org/10.1080/00222938900770021.

Eickwort, George C. 1977. 'Male Territorial Behaviour in the Mason Bee Hoplitis Anthocopoides (Hymenoptera: Megachilidae)'. *Animal Behaviour* 25: 542–554.

Engels, W., H. Kaatz, A. Zillikens, Z. L. P. Simões, A. Trube, R. Braun, and F. Dittrich. 1990. 'Honey Bee Reproduction: Vitellogenin and Caste-Specific Regulation of Fertility.' *Honey Bee Reproduction: Vitellogenin and Caste-Specific Regulation of Fertility.*, 495–502.

Engels, W. and Fahrenhorst H. 1974. '[Age- and Caste-Dependent Changes in the Haemolymph Protein Patterns OfApis Mellifica].' *Wilhelm Roux' Archiv Fur Entwicklungsmechanik Der Organismen* 174 (3): 285–96. https://doi.org/10.1007/bf00573233.

Estévez, Ana Lía, and José Ricardo Inacio Ribeiro. 2011. 'Weberiella De Carlo, 1966 (Insecta: Heteroptera: Belostomatidae) Revisited: Redescription with a Key to the Genera of Belostomatidae and Considerations on Back-Brooding Behaviour'. *Zoologischer Anzeiger - A Journal of Comparative Zoology* 250 (1): 46–54. https://doi.org/10.1016/j.jcz.2010.11.002.

Evans, Nathan T., Brett P. Olds, Mark A. Renshaw, Cameron R. Turner, Yiyuan Li, Christopher L. Jerde, Andrew R. Mahon, Michael E. Pfrender, Gary A. Lamberti, and David M. Lodge. 2016. 'Quantification of Mesocosm Fish and Amphibian Species Diversity via Environmental DNA Metabarcoding'. *Molecular Ecology Resources* 16 (1): 29–41. https://doi.org/10.1111/1755-0998.12433.

Fedotova, A. A., A. N. Bonchuk, V. A. Mogila, and P. G. Georgiev. 2017. 'C2H2 Zinc Finger Proteins: The Largest but Poorly Explored Family of Higher Eukaryotic Transcription Factors'. *Acta Naturae* 9 (2): 47–58.

Feng, Jian, Hua Chang, En Li, and Guoping Fan. 2005. 'Dynamic Expression of de Novo DNA Methyltransferases Dnmt3a and Dnmt3b in the Central Nervous System'. *Journal of Neuroscience Research* 79 (6): 734–46. https://doi.org/10.1002/jnr.20404.

Feng, Jian, Yu Zhou, Susan L. Campbell, Thuc Le, En Li, J. David Sweatt, Alcino J. Silva, and Guoping Fan. 2010. 'Dnmt1 and Dnmt3a Maintain DNA Methylation and Regulate Synaptic Function in Adult Forebrain Neurons'. *Nature Neuroscience* 13 (4): 423–30. https://doi.org/10.1038/nn.2514.

Ferguson, Laura C., Jack Green, Alison Surridge, and Chris D. Jiggins. 2011. 'Evolution of the Insect Yellow Gene Family'. *Molecular Biology and Evolution* 28 (1): 257–72. https://doi.org/10.1093/molbev/msq192.

Fetherston, Isabelle A., Michelle Pellissier Scott, and James F. A. Traniello. 1990. 'Parental Care in Burying Beetles: The Organization of Male and Female Brood-Care Behavior'. *Ethology* 85 (3): 177–90. https://doi.org/10.1111/j.1439-0310.1990.tb00398.x.

Filippi, Lisa, Narumi Baba, Koichi Inadomi, Takao Yanagi, Mantaro Hironaka, and Shintaro Nomakuchi. 2009. 'Pre- and Post-Hatch Trophic Egg Production in the Subsocial Burrower Bug, Canthophorus Niveimarginatus (Heteroptera: Cydnidae)'. *Naturwissenschaften* 96 (2): 201–11. https://doi.org/10.1007/s00114-008-0463-z.

Filippi-Tsukamoto, Lisa, Shintaro Nomakuchi, Kanae Kuki, and Sumio Tojo. 1995. 'Adaptiveness of Parental Care in Parastrachia Japonensis (Hemiptera: Cydnidae)'. *Annals of the Entomological Society of America* 88 (3): 374–83. https://doi.org/10.1093/aesa/88.3.374.

Gagliano, Sarah A., Jennie G. Pouget, John Hardy, Jo Knight, Michael R. Barnes, Mina Ryten, and Michael E. Weale. 2016. 'Genomics Implicates Adaptive and Innate Immunity in Alzheimer's and Parkinson's Diseases'. *Annals of Clinical and Translational Neurology* 3 (12): 924–933.

Geiss-Friedlander, Ruth, and Frauke Melchior. 2007. 'Concepts in Sumoylation: A Decade On'. *Nature Reviews Molecular Cell Biology* 8 (12): 947–56. https://doi.org/10.1038/nrm2293.

Gleason, Erin D., and Catherine A. Marler. 2013. 'Non-Genomic Transmission of Paternal Behaviour between Fathers and Sons in the Monogamous and Biparental California Mouse'. *Proceedings of the Royal Society B: Biological Sciences* 280 (1763): 20130824. https://doi.org/10.1098/rspb.2013.0824.

Gompel, Nicolas, Benjamin Prud'homme, Patricia J. Wittkopp, Victoria A. Kassner, and Sean B. Carroll. 2005. 'Chance Caught on the Wing: Cis -Regulatory Evolution and the Origin of Pigment Patterns in Drosophila'. *Nature* 433 (7025): 481–87. https://doi.org/10.1038/nature03235.

Goodwin, Nicholas B., Sigal Balshine-Earn, and John D. Reynolds. 1998. 'Evolutionary Transitions in Parental Care in Cichlid Fish'. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265 (1412): 2265–2272.

Gospocic, Janko, Emily J. Shields, Karl M. Glastad, Yanping Lin, Clint A. Penick, Hua Yan, Alexander S. Mikheyev, et al. 2017. 'The Neuropeptide Corazonin Controls Social Behavior and Caste Identity in Ants'. *Cell* 170 (4): 748-759.e12. https://doi.org/10.1016/j.cell.2017.07.014.

Grabherr, Manfred G., Brian J. Haas, Moran Yassour, Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, et al. 2011. 'Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome'. *Nature Biotechnology* 29 (7): 644–52. https://doi.org/10.1038/nbt.1883.

Greenberg J. K., Xia J., Zhou X., Thatcher S. R., Gu X., Ament S. A., Newman T. C., et al. 2012. 'Behavioral Plasticity in Honey Bees Is Associated with Differences in Brain MicroRNA Transcriptome'. *Genes, Brain and Behavior* 11 (6): 660–70. https://doi.org/10.1111/j.1601-183X.2012.00782.x.

Groom, Scott V. C., and Sandra M. Rehan. n.d. 'Climate-Mediated Behavioural Variability in Facultatively Social Bees'. *Biological Journal of the Linnean Society*. Accessed 28 July 2018. https://doi.org/10.1093/biolinnean/bly101.

Guershon, Moshe, and Dan Gerling. 2001. 'Parental Care in the Whitefly Aleyrodes Singularis'. *Ecological Entomology* 26 (5): 467–472.

Haas, Brian J., Arthur L. Delcher, Stephen M. Mount, Jennifer R. Wortman, Roger K. Smith, Linda I. Hannick, Rama Maiti, et al. 2003. 'Improving the Arabidopsis Genome Annotation Using Maximal Transcript Alignment Assemblies'. *Nucleic Acids Research* 31 (19): 5654–66. https://doi.org/10.1093/nar/gkg770.

Haas, Brian J., Steven L. Salzberg, Wei Zhu, Mihaela Pertea, Jonathan E. Allen, Joshua Orvis, Owen White, C. Robin Buell, and Jennifer R. Wortman. 2008. 'Automated Eukaryotic Gene Structure Annotation Using EVidenceModeler and the Program to Assemble Spliced Alignments'. *Genome Biology* 9 (1): R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Haas, B., and A. Papanicolaou. "TransDecoder (find coding regions within transcripts)(2018)."

Halffter, GONZALO, CARMEN Huerta, and JORGE Lopez-Portillo. 1996. 'Parental Care and Offspring Survival in Copris Incertus Say, a Sub-Social Beetle'. *Animal Behaviour* 52 (1): 133–139.

Hamilton, W.D. 1964. 'The Genetical Evolution of Social Behaviour. I'. *Journal of Theoretical Biology* 7 (1): 1–16. https://doi.org/10.1016/0022-5193(64)90038-4.

Hanelová, J., and J. Vilímová. 2013. 'Behaviour of the Central European Acanthosomatidae (Hemiptera: Heteroptera: Pentatomoidea) during Oviposition and Parental Care'. *Acta Musei Moraviae, Scientiae Biologicae (Brno)* 98 (2): 433–457.

Harrison, Jon M. 1986. 'Caste-Specific Changes in Honeybee Flight Capacity'. *Physiological Zoology* 59 (2): 175–87. https://doi.org/10.1086/physzool.59.2.30156031.

Hernández, Liudy G., Bingwen Lu, Gabriel C. N. da Cruz, Luciana K. Calábria, Natalia F. Martins, Roberto Togawa, Foued S. Espindola, John R. Yates, Ricardo B. Cunha, and Marcelo V. de Sousa. 2012. 'The Worker Honeybee Brain Proteome'. *Journal of Proteome Research* 11 (3): 1485–93. https://doi.org/10.1021/pr2007818.

Hirt, R. P., J. M. Logsdon, B. Healy, M. W. Dorey, W. F. Doolittle, and T. M. Embley. 1999. 'Microsporidia Are Related to Fungi: Evidence from the Largest Subunit of RNA Polymerase II and Other Proteins'. *Proceedings of the National Academy of Sciences* 96 (2): 580–85. https://doi.org/10.1073/pnas.96.2.580.

Hoff, Katharina J., Alexandre Lomsadze, Mark Borodovsky, and Mario Stanke. 2019. 'Whole-Genome Annotation with BRAKER'. *Methods in Molecular Biology (Clifton, N.J.)* 1962: 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5.

Holt, Carson, and Mark Yandell. 2011. 'MAKER2: An Annotation Pipeline and Genome-Database Management Tool for Second-Generation Genome Projects'. *BMC Bioinformatics* 12 (1): 491. https://doi.org/10.1186/1471-2105-12-491.

Hunt, J. H., and K. C. Noonan. 1979. 'Larval Feeding by MalePolistes Fuscatus AndPolistes Metricus (Hymenoptera: Vespidae)'. *Insectes Sociaux* 26 (3): 247–251.

Hunt, James H., Florian Wolschin, Michael T. Henshaw, Thomas C. Newman, Amy L. Toth, and Gro V. Amdam. 2010. 'Differential Gene Expression and Protein Abundance Evince Ontogenetic Bias toward Castes in a Primitively Eusocial Wasp'. *PLOS ONE* 5 (5): e10674. https://doi.org/10.1371/journal.pone.0010674.

Hunt, and L. W. Simmons. 2002. 'Behavioural Dynamics of Biparental Care in the Dung Beetle Onthophagus Taurus'. *Animal Behaviour* 64 (1): 65–75. https://doi.org/10.1006/anbe.2002.3036.

Johnsson, Martin, Rie Henriksen, Jesper Fogelholm, Andrey Höglund, Per Jensen, and Dominic Wright. 2018. 'Genetics and Genomics of Social Behavior in a Chicken Model'. *Genetics* 209 (1): 209–21. https://doi.org/10.1534/genetics.118.300810.

Jones, Beryl M., Callum J. Kingwell, William T. Wcislo, and Gene E. Robinson. 2017. 'Caste-Biased Gene Expression in a Facultatively Eusocial Bee Suggests a Role for Genetic Accommodation in the Evolution of Eusociality'. *Proc. R. Soc. B* 284 (1846): 20162228. https://doi.org/10.1098/rspb.2016.2228.

Jordal, Bjarte H., Andrea S. Sequeira, and Anthony I. Cognato. 2011. 'The Age and Phylogeny of Wood Boring Weevils and the Origin of Subsociality'. *Molecular Phylogenetics and Evolution* 59 (3): 708–24. https://doi.org/10.1016/j.ympev.2011.03.016.

Jungfer, Karl-Heinz, and Peter Weygoldt. 1999. 'Biparental Care in the Tadpole-Feeding Amazonian Treefrog Osteocephalus Oophagus'. *Amphibia-Reptilia* 20 (3): 235–49. https://doi.org/10.1163/156853899X00277.

Kaitala, Arja, and Johanna Mappes. 1997. 'Parental Care and Reproductive Investment in Shield Bugs (Acanthosomatidae, Heteroptera)'. *Oikos*, 3–7.

Kamakura, Masaki. 2011. 'Royalactin Induces Queen Differentiation in Honeybees'. *Nature* 473 (7348): 478–83. https://doi.org/10.1038/nature10093.

Kanamori, Yasushi, Ayako Saito, Yuka Hagiwara-Komoda, Daisuke Tanaka, Kanako Mitsumasu, Shingo Kikuta, Masahiko Watanabe, Richard Cornette, Takahiro Kikawada, and Takashi Okuda. 2010. 'The Trehalose Transporter 1 Gene Sequence Is Conserved in Insects and Encodes Proteins with Different Kinetic Properties Involved in Trehalose Import into Peripheral Tissues'. *Insect Biochemistry and Molecular Biology* 40 (1): 30–37. https://doi.org/10.1016/j.ibmb.2009.12.006.

Kavanaugh, David H. 1998. 'Field Observations Confirming Brood Care in Percus Passerinii in the Tuscan Apennines, Italy (Coleoptera Carabidae)'. *Bollettino Della Società Entomologica Italiana* 130 (1): 55–63.

Kearse, Matthew, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, et al. 2012. 'Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data'. *Bioinformatics* 28 (12): 1647–49. https://doi.org/10.1093/bioinformatics/bts199.

Kent, W. James. 2002. 'BLAT—The BLAST-Like Alignment Tool'. *Genome Research* 12 (4): 656–64. https://doi.org/10.1101/gr.229202.

Khailany, Rozhgar A., Muhamad Safdar, and Mehmet Ozaslan. 2020. 'Genomic Characterization of a Novel SARS-CoV-2'. *Gene Reports*, 100682.

Kim, Daehwan, Ben Langmead, and Steven L. Salzberg. 2015. 'HISAT: A Fast Spliced Aligner with Low Memory Requirements'. *Nature Methods* 12 (4): 357–60. https://doi.org/10.1038/nmeth.3317.

Kirkendall, Lawrence R. 1997. 'Interactions among Males, Females and Offspring in Bark and Ambrosia Beetles: The Significance of Living in Tunnels for the Evolution of Social Behavior'. *The Evolution of Social Behavior in Insects and Arachnids*, 181–215.

Kirkendall, L. R., D. S. Kent, and K. F. Raffa. "The Evolution of Social Behavior in Insects and Arachnids, eds Choe JC, Crespi BJ." (1997): 181-215.

Kocher, Sarah D., Cai Li, Wei Yang, Hao Tan, Soojin V. Yi, Xingyu Yang, Hopi E. Hoekstra, Guojie Zhang, Naomi E. Pierce, and Douglas W. Yu. 2013. 'The Draft Genome of a Socially Polymorphic Halictid Bee, Lasioglossum Albipes'. *Genome Biology* 14 (December): R142. https://doi.org/10.1186/gb-2013-14-12-r142.

Korf, Ian. 2004. 'Gene Finding in Novel Genomes'. *BMC Bioinformatics* 5 (1): 59.

Kronforst, Marcus R., David C. Gilley, Joan E. Strassmann, and David C. Queller. 2008. 'DNA Methylation Is Widespread across Social Hymenoptera'. *Current Biology* 18 (7): R287–88. https://doi.org/10.1016/j.cub.2008.02.015.

Kulp, D., D. Haussler, M. G. Reese, and F. H. Eeckman. 1996. 'A Generalized Hidden Markov Model for the Recognition of Human Genes in DNA'. *Proceedings. International Conference on Intelligent Systems for Molecular Biology* 4: 134–42.

Lack, David Lambert. 1968. 'Ecological Adaptations for Breeding in Birds'.

Laity, John H, Brian M Lee, and Peter E Wright. 2001. 'Zinc Finger Proteins: New Insights into Structural and Functional Diversity'. *Current Opinion in Structural Biology* 11 (1): 39–46. https://doi.org/10.1016/S0959-440X(00)00167-6.

Lawson, Sarah P., Krista N. Ciaccio, and Sandra M. Rehan. 2016. 'Maternal Manipulation of Pollen Provisions Affects Worker Production in a Small Carpenter Bee'. *Behavioral Ecology and Sociobiology* 70 (11): 1891–1900.

Lawson, Sarah P., Salena L. Helmreich, and Sandra M. Rehan. 2017. 'Effects of Nutritional Deprivation on Development and Behavior in the Subsocial Bee Ceratina Calcarata (Hymenoptera: Xylocopinae)'.

Li, Heng. 2018. 'Minimap2: Pairwise Alignment for Nucleotide Sequences'. *Bioinformatics* 34 (18): 3094–3100.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. 'The Sequence Alignment/Map Format and SAMtools'. *Bioinformatics* 25 (16): 2078–2079.

Li, Jianyong, and Bruce M. Christensen. 2011. 'Biological Function of Insect Yellow Gene Family'. In *Recent Advances in Entomological Research: From Molecular Biology to Pest Management*, edited by Tongxian Liu and Le Kang, 121–31. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-17815-3_7.

Li, Zhizhong, Ru Cao, Ming Wang, Michael P. Myers, Yi Zhang, and Rui-Ming Xu. 2006. 'Structure of a Bmi-1-Ring1B Polycomb Group Ubiquitin Ligase Complex'. *Journal of Biological Chemistry* 281 (29): 20643–49. https://doi.org/10.1074/jbc.M602461200.

Liew, Chu Kong, Raina J. Y. Simpson, Ann H. Y. Kwan, Linda A. Crofts, Fionna E. Loughlin, Jacqueline M. Matthews, Merlin Crossley, and Joel P. Mackay. 2005. 'Zinc Fingers as Protein Recognition Motifs: Structural Basis for the GATA-1/Friend of GATA Interaction'. *Proceedings of the National Academy of Sciences* 102 (3): 583–88. https://doi.org/10.1073/pnas.0407511102.

Linksvayer, Timothy A, and Michael J Wade. 2005. 'The Evolutionary Origin and Elaboration of Sociality in the Aculeate Hymenoptera: Maternal Effects, Sib-social Effects, and Heterochrony'. *The Quarterly Review of Biology* 80 (3): 317–36. https://doi.org/10.1086/432266.

Loeb, Michael LG, and Lisa K. Bell. "Distribution of care-giving effort in a communally breeding lace bug: fair guarding without coercion." *Journal of insect behavior* 19.1 (2006): 19-30.

Lu, Hengyun, Francesca Giordano, and Zemin Ning. 2016. 'Oxford Nanopore MinION Sequencing and Genome Assembly'. *Genomics, Proteomics & Bioinformatics*, SI: Big Data and Precision Medicine, 14 (5): 265–79. https://doi.org/10.1016/j.gpb.2016.05.004.

Lucas, E. R., R. P. Martins, L. R. S. Zanette, and J. Field. 2011. 'Social and Genetic Structure in Colonies of the Social Wasp Microstigmus Nigrophthalmus'. *Insectes Sociaux* 58 (1): 107–14. https://doi.org/10.1007/s00040-010-0123-0.

Lucas, Eric R., and Jeremy Field. 2011. 'Active and Effective Nest Defence by Males in a Social Apoid Wasp'. *Behavioral Ecology and Sociobiology* 65 (8): 1499–1504. https://doi.org/10.1007/s00265-011-1159-5.

Luo, Ruibang, Binghang Liu, Yinlong Xie, Zhenyu Li, Weihua Huang, Jianying Yuan, Guangzhu He, Yanxiang Chen, Qi Pan, and Yunjie Liu. 2012. 'SOAPdenovo2: An Empirically Improved Memory-Efficient Short-Read de Novo Assembler'. *Gigascience* 1 (1): 18.

Lyko, Frank, and Ryszard Maleszka. 2011. 'Insects as Innovative Models for Functional Studies of DNA Methylation'. *Trends in Genetics* 27 (4): 127–31. https://doi.org/10.1016/j.tig.2011.01.003.

Maekawa, K., T. Matsumoto, and C. A. Nalepa. 2008. 'Social Biology of the Wood-Feeding Cockroach Genus Salganea (Dictyoptera, Blaberidae, Panesthiinae): Did Ovoviviparity Prevent the Evolution of Eusociality in the Lineage?' *Insectes Sociaux* 55 (2): 107–114.

Maeta, Y., and Shoichi Sakagami. 1995. 'Oophagy and Egg Replacement in Artificially Induced Colonies of a Basically Solitary Bee, Ceratina (Ceratinidia) Okinawana (Hymenoptera, Anthophoridae, Xylocopinae), with a Comparison of Social Behavior among Ceratina, Xylocopa and the Halictine Bees'. *Japanese Journal of Entomology* 63 (2): 347–75.

Maeta, Yasuo, Kazuhiro Saito, and Kaoru Hyudo. 1993. 'Diapause and Non-Delayed Eusociality in a Univoltine and Basically Solitary Bee, Ceratina Japonica (Hymenoptera, Anthophoridae) : I. Diapause Termination by Cooling and Application of Juvenile Hormone Analog'. *Japanese Journal of Entomology* 61 (2): 203–11.

Maeta, Yasuo, and Shôichi F. Sakagami. 1995. 'Oophagy and Egg Replacement in Artificially Induced Colonies of a Basically Solitary Bee, Ceratina (Ceratinidia) Okinawana (Hymenoptera, Anthophoridae, Xylocopinae), with a Comparison of Social Behavior among Ceratina, Xylocopa and the Halictine Bees'. *Japanese Journal of Entomology* 63 (2): 347–75.

Matthews, Robert W. 1968. 'Microstigmus Comes: Sociality in a Sphecid Wasp'. *Science* 160 (3829): 787–88. https://doi.org/10.1126/science.160.3829.787.

Matthews, Robert W., and Janice R. Matthews. 2009. 'Nesting Behavior of Abispa Ephippium (Fabricius)(Hymenoptera: Vespidae: Eumeninae): Extended Parental Care in an Australian Mason Wasp'. *Psyche: A Journal of Entomology* 2009.

Michell, Robert H. 2008. 'Inositol Derivatives: Evolution and Functions'. *Nature Reviews Molecular Cell Biology* 9 (2): 151–61. https://doi.org/10.1038/nrm2334.

Michener, C., and W. Engels. 1990. *Social Insects: An Evolutionary Approach to Castes and Reproduction*. Heidelberg: Springer Verlag 1990 chap. Reproduction and caste in social ….

Michener, Charles D. 2007. *The Bees of the World*. 2nd ed. The Johns Hopkins University Press.

Mikát, Michael, Daniel Benda, and Jakub Straka. 2020. 'Unrelated Males in Colonies of Facultatively Social Bee'. *BioRxiv*.

Mikát, Michael, Cullen Franchino, and Sandra M. Rehan. 2017. 'Sociodemographic Variation in Foraging Behavior and the Adaptive Significance of Worker Production in the Facultatively Social Small Carpenter Bee, Ceratina Calcarata'. *Behavioral Ecology and Sociobiology* 71 (9): 135. https://doi.org/10.1007/s00265-017-2365-6.

Mikát, Michael, Lukáš Janošík, Kateřina Černá, Eva Matoušková, Jiří Hadrava, Vít Bureš, and Jakub Straka. 2019. 'Polyandrous Bee Provides Extended Offspring Care Biparentally as an Alternative to Monandry Based Eusociality'. *Proceedings of the National Academy of Sciences* 116 (13): 6238–43. https://doi.org/10.1073/pnas.1810092116.

Mikát, Michael, Jitka Waldhauserová, Tereza Fraňková, Kateřina Čermáková, Vojtěch Brož, Šimon Zeman, Marcela Dokulilová, and Jakub Straka. n.d. 'Only Mothers Feed Mature Offspring in European Ceratina Bees'. *Insect Science* n/a (n/a). Accessed 7 August 2020. https://doi.org/10.1111/1744-7917.12859.

Munch, Kasper, and Anders Krogh. 2006. 'Automatic Generation of Gene Finders for Eukaryotic Species'. *BMC Bioinformatics* 7 (1): 263. https://doi.org/10.1186/1471-2105-7-263.

Mwinyi, A., X. Bailly, S. Bourlat, U. Jondelius, D. T. Littlewood, and L. Podsiadlowski. 2010. 'The Phylogenetic Position of Acoela as Revealed by the Complete Mitochondrial Genome of Symsagittifera Roscoffensis'. *BMC Evolutionary Biology* 10 (1): 309.
MySQL, A. B. 2001. *MySQL*.

Nalepa, Christine A. 1990. 'Early Development of Nymphs and Establishment of Hindgut Symbiosis in Cryptocercus Punctulatus (Dictyoptera: Cryptocercidae)'. *Annals of the Entomological Society of America* 83 (4): 786–89. https://doi.org/10.1093/aesa/83.4.786.

Nalepa, Christine A., and William J. Bell. "The Evolution of Social Behavior in Insects and Arachnids: Postovulation parental investment and parental care in cockroaches." (1997).

New, T. R. 1973. 'The Archipsocidae of South America (Psocoptera)'. *Transactions of the Royal Entomological Society of London* 125 (1): 57–103. https://doi.org/10.1111/j.1365-2311.1973.tb00538.x.

Okazaki, Katsunori. 1987. 'Life Cycle of a Subtropical Xylocopine Bee, Ceratina Okinawana, with Some Related Problems'. *Kontyu* 55 (1): 1–8.

O'Neill, Kevin M. 1983. 'The Significance of Body Size in Territorial Interactions of Male Beewolves (Hymenoptera: Sphecidae, Philanthus)'. *Animal Behaviour* 31 (2): 404–411.

Oppenheimer, Robert L., Wyatt A. Shell, and Sandra M. Rehan. 2018. 'Phylogeography and Population Genetics of the Australian Small Carpenter Bee, Ceratina Australensis'. *Biological Journal of the Linnean Society* 124 (4): 747–55. https://doi.org/10.1093/biolinnean/bly070.

Organ, C. L., A. Canoville, R. R. Reisz, and M. Laurin. 2011. 'Paleogenomic Data Suggest Mammal-like Genome Size in the Ancestral Amniote and Derived Large Genome Size in Amphibians'. *Journal of Evolutionary Biology* 24 (2): 372–80. https://doi.org/10.1111/j.1420-9101.2010.02176.x.

Pachetti, Maria, Bruna Marini, Francesca Benedetti, Fabiola Giudici, Elisabetta Mauro, Paola Storici, Claudio Masciovecchio, Silvia Angeletti, Massimo Ciccozzi, and Robert C. Gallo. 2020. 'Emerging SARS-CoV-2 Mutation Hot Spots Include a Novel RNA-Dependent-RNA Polymerase Variant'. *Journal of Translational Medicine* 18: 1–9.

Page Jr, R. E., J. Erber, and M. K. Fondrk. 1998. 'The Effect of Genotype on Response Thresholds to Sucrose and Foraging Behavior of Honey Bees (Apis Mellifera L.)'. *Journal of Comparative Physiology A* 182 (4): 489–500.

Park, Jung-Min, Takekazu Kunieda, and Takeo Kubo. 2003. 'The Activity of Mblk-1, a Mushroom Body-Selective Transcription Factor from the Honeybee, Is Modulated by the Ras/MAPK Pathway'. *Journal of Biological Chemistry* 278 (20): 18689–94. https://doi.org/10.1074/jbc.M300486200.

Park, Min Ji, Bo Yeon Kim, Hee Geun Park, Yijie Deng, Hyung Joo Yoon, Yong Soo Choi, Kwang Sik Lee, and Byung Rae Jin. 2019. 'Major Royal Jelly Protein 2 Acts as an Antimicrobial Agent and Antioxidant in Royal Jelly'. *Journal of Asia-Pacific Entomology* 22 (3): 684–89. https://doi.org/10.1016/j.aspen.2019.05.003.

Park, Youngjin, and David Stanley. 2015. 'Physiological Trade-off between Cellular Immunity and Flight Capability in the Wing-Dimorphic Sand Cricket, Gryllus Firmus'. *Journal of Asia-Pacific Entomology* 18 (3): 553–559.

Park, Yung Chul, Philippe Grandcolas, and Jae Chun Choe. 2002. 'Colony Composition, Social Behavior and Some Ecological Characteristics of the Korean Wood-Feeding Cockroach (Cryptocercus Kyebangensis)'. *Zoological Science* 19 (10): 1133–39. https://doi.org/10.2108/zsj.19.1133.

Pask, Gregory M., Jesse D. Slone, Jocelyn G. Millar, Prithwiraj Das, Jardel A. Moreira, Xiaofan Zhou, Jan Bello, et al. 2017. 'Specialized Odorant Receptors in Social Insects That Detect Cuticular Hydrocarbon Cues and Candidate Pheromones'. *Nature Communications* 8 (1): 297. https://doi.org/10.1038/s41467-017-00099-1.

Pavlova, Gera A., Julia V. Popova, Evgeniya N. Andreyeva, Lyubov A. Yarinich, Mikhail O. Lebedev, Alyona V. Razuvaeva, Tatiana D. Dubatolova, et al. 2019. 'RNAi-Mediated Depletion of the NSL Complex Subunits Leads to Abnormal Chromosome Segregation and Defective Centrosome Duplication in Drosophila Mitosis'. *PLOS Genetics* 15 (9): e1008371. https://doi.org/10.1371/journal.pgen.1008371.

Peña, Pedro V., Foteini Davrazou, Xiaobing Shi, Kay L. Walter, Vladislav V. Verkhusha, Or Gozani, Rui Zhao, and Tatiana G. Kutateladze. 2006. 'Molecular Mechanism of Histone H3K4me3 Recognition by Plant Homeodomain of ING2'. *Nature* 442 (7098): 100–103. https://doi.org/10.1038/nature04814.

Pinto, Lucimara Z, Márcia M. G Bitondi, and Zilá L. P Simões. 2000. 'Inhibition of Vitellogenin Synthesis in Apis Mellifera Workers by a Juvenile Hormone Analogue, Pyriproxyfen'. *Journal of Insect Physiology* 46 (2): 153–60. https://doi.org/10.1016/S0022-1910(99)00111-0.

Plambeck, Craig A., Ann H. Y. Kwan, David J. Adams, Belinda J. Westman, Louise van der Weyden, Robert L. Medcalf, Brian J. Morris, and Joel P. Mackay. 2003. 'The Structure of the Zinc Finger Domain from Human Splicing Factor ZNF265 Fold'. *Journal of Biological Chemistry* 278 (25): 22805–11. https://doi.org/10.1074/jbc.M301896200.

Polekhina, Galina, Colin M. House, Nadia Traficante, Joel P. Mackay, Frédéric Relaix, David A. Sassoon, Michael W. Parker, and David DL Bowtell. 2002. 'Siah Ubiquitin Ligase Is Structurally Related to TRAF and Modulates TNF-α Signaling'. *Nature Structural Biology* 9 (1): 68–75.

Quevillon, Emmanuel, Ville Silventoinen, Sharmila Pillai, Nicola Harte, N. Mulder, Rolf Apweiler, and Rodrigo Lopez. 2005. 'InterProScan: Protein Domains Identifier'. *Nucleic Acids Research* 33 (suppl_2): W116–W120.

Radl, Ruth C., and K. Eduard Linsenmair. 1991. 'Maternal Behaviour and Nest Recognition in the Subsocial Earwig Labidura Riparia Pallas (Dermaptera: Labiduridae)'. *Ethology* 89 (4): 287–96. https://doi.org/10.1111/j.1439-0310.1991.tb00374.x.

Rehan, M. H. Richards, and M. P. Schwarz. 2010. 'Social Polymorphism in the Australian Small Carpenter Bee, Ceratina (Neoceratina) Australensis'. *Insectes Sociaux* 57 (4): 403–12. https://doi.org/10.1007/s00040-010-0097-y.

Rehan, S. M., K. M. Glastad, M. A. Steffen, C. R. Fay, B. G. Hunt, and A. L. Toth. 2018. 'Conserved Genes Underlie Phenotypic Plasticity in an Incipiently Social Bee'. *Genome Biology and Evolution*, September. https://doi.org/10.1093/gbe/evy212.

Rehan, Sandra M. 2020 'Small Carpenter Bees'.

Rehan, Sandra M., Susan J. Bulova, and Sean O''Donnell. 2015. 'Cumulative Effects of Foraging Behavior and Social Dominance on Brain Development in a Facultatively Social Bee (Ceratina Australensis)'. *Brain, Behavior and Evolution* 85 (2): 117–24. https://doi.org/10.1159/000381414.

Rehan, Sandra M., Karl M. Glastad, Sarah P. Lawson, and Brendan G. Hunt. 2016. 'The Genome and Methylome of a Subsocial Small Carpenter Bee, Ceratina Calcarata'. *Genome Biology and Evolution* 8 (5): 1401–10. https://doi.org/10.1093/gbe/evw079.

Rehan, Sandra M., and Miriam H. Richards. 2010. 'Nesting Biology and Subsociality in Ceratina Calcarata (Hymenoptera: Apidae)'. *The Canadian Entomologist* 142 (1): 65–74. https://doi.org/10.4039/n09-056.

Rehan, Sandra M., and Amy L. Toth. 2015. 'Climbing the Social Ladder: The Molecular Evolution of Sociality'. *Trends in Ecology & Evolution* 30 (7): 426–433.

Robertson, H. M., J. Gadau, and K. W. Wanner. 2010. 'The Insect Chemoreceptor Superfamily of the Parasitoid Jewel Wasp Nasonia Vitripennis'. *Insect Molecular Biology* 19 (February): 121–36. https://doi.org/10.1111/j.1365-2583.2009.00979.x.

Robertson, Hugh M., and Kevin W. Wanner. 2006. 'The Chemoreceptor Superfamily in the Honey Bee, Apis Mellifera: Expansion of the Odorant, but Not Gustatory, Receptor Family'. *Genome Research* 16 (11): 1395–1403. https://doi.org/10.1101/gr.5057506.

Rogers, Rebekah L., and Montgomery Slatkin. 2017. 'Excess of Genomic Defects in a Woolly Mammoth on Wrangel Island'. *PLOS Genetics* 13 (3): e1006601. https://doi.org/10.1371/journal.pgen.1006601.

Ross, Edward S. "Embia: Contributions to the biosystematics of the insect order Embiidina. Part 1, Origin, relationships and integumental anatomy of the insect order Embiidina. Part 2. A review of the biology of Embiidina." *Occasional papers of the California Academy of Sciences* 149 (2000): 1-53.

Roy, Saurabh G., Immo A. Hansen, and Alexander S. Raikhel. 2007. 'Effect of Insulin and 20-Hydroxyecdysone in the Fat Body of the Yellow Fever Mosquito, Aedes Aegypti'. *Insect Biochemistry and Molecular Biology* 37 (12): 1317–26. https://doi.org/10.1016/j.ibmb.2007.08.004.

Roy-Zokan, Eileen M., Christopher B. Cunningham, Lauren E. Hebb, Elizabeth C. McKinney, and Allen J. Moore. 2015. 'Vitellogenin and Vitellogenin Receptor Gene Expression Is Associated with Male and Female Parenting in a Subsocial Insect'. *Proceedings. Biological Sciences* 282 (1809): 20150787. https://doi.org/10.1098/rspb.2015.0787.

Rutz, W., L. Gerig, H. Wille, and M. Lüscher. 1976. 'The Function of Juvenile Hormone in Adult Worker Honeybees, Apis Mellifera'. *Journal of Insect Physiology* 22 (11): 1485–91. https://doi.org/10.1016/0022-1910(76)90214-6.

Sadd, Ben M., Seth M. Barribeau, Guy Bloch, Dirk C. De Graaf, Peter Dearden, Christine G. Elsik, Jürgen Gadau, et al. 2015. 'The Genomes of Two Key Bumblebee Species with Primitive Eusocial Organization'. *Genome Biology* 16 (1): 76.

Sahlin, Kristoffer, Francesco Vezzi, Björn Nystedt, Joakim Lundeberg, and Lars Arvestad. 2014. 'BESST-Efficient Scaffolding of Large Fragmented Assemblies'. *BMC Bioinformatics* 15 (1): 281.

Sakagami, S. F. 1985. 'Multifemale Nests and Rudimentary Castes, in the Normally Solitary Bee Ceratina Japonica (Hymenoptera : Xylocopinae)'. *Journal of the Kansas Entomological Society* 57: 639–56.

Sargent, Robert Craig, Peter D. Taylor, and Mart R. Gross. 1987. 'Parental Care and the Evolution of Egg Size in Fishes'. *American Naturalist*, 32–46.

Saurin, Andrew J., Katherine L. B. Borden, Michael N. Boddy, and Paul S. Freemont. 1996. 'Does This Have a Familiar RING?' *Trends in Biochemical Sciences* 21 (6): 208–14. https://doi.org/10.1016/S0968-0004(96)80017-X.

Scheiner, Ricarda, Robert E. Page, and Joachim Erber. 2001. 'The Effects of Genotype, Foraging Role, and Sucrose Responsiveness on the Tactile Learning Performance of Honey Bees (Apis Mellifera L.)'. *Neurobiology of Learning and Memory* 76 (2): 138–50. https://doi.org/10.1006/nlme.2000.3996.

Schuster, Jack C., and Laura B. Schuster. 1997. 'The Evolution of Social Behavior in Passalidae (Coleoptera)'. The Evolution of Social Behaviour in Insects and Arachnids. March 1997. https://doi.org/10.1017/CBO9780511721953.013.

Searle, Naomi E., and Lorraine Pillus. 2018. 'Critical Genomic Regulation Mediated by Enhancer of Polycomb'. *Current Genetics* 64 (1): 147–54. https://doi.org/10.1007/s00294-017-0742-3.

Seelinger, Günter, and Ute Seelinger. 1983. 'On the Social Organisation, Alarm and Fighting in the Primitive Cockroach Cryptocercus Punctulatus Scudder'. *Zeitschrift Für Tierpsychologie* 61 (4): 315–33. https://doi.org/10.1111/j.1439-0310.1983.tb01347.x.

Sen, Ruchira, and Raghavendra Gadagkar. 2006. 'Males of the Social Wasp Ropalidia Marginata Can Feed Larvae, given an Opportunity'. *Animal Behaviour* 71 (2): 345–350.

Seppey, Mathieu, Mosè Manni, and Evgeny M. Zdobnov. 2019. 'BUSCO: Assessing Genome Assembly and Annotation Completeness'. In *Gene Prediction: Methods and Protocols*, edited by Martin Kollmar, 227–45. Methods in Molecular Biology. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-9173-0_14.

Shapiro, James A., and Richard von Sternberg. 2005. 'Why Repetitive DNA Is Essential to Genome Function'. *Biological Reviews* 80 (2): 227–50. https://doi.org/10.1017/S1464793104006657.

Shell, Wyatt A., and Sandra M. Rehan. 2017. 'Behavioral and Genetic Mechanisms of Social Evolution: Insights from Incipiently and Facultatively Social Bees'. *Apidologie*, July, 1–18. https://doi.org/10.1007/s13592-017-0527-1.

Simpson, Raina J. Y., Stella Hoi Yi Lee, Natalie Bartle, Eleanor Y. Sum, Jane E. Visvader, Jacqueline M. Matthews, Joel P. Mackay, and Merlin Crossley. 2004. 'A Classic Zinc Finger from Friend of GATA Mediates an Interaction with the Coiled-Coil of Transforming Acidic Coiled-Coil 3'. *Journal of Biological Chemistry* 279 (38): 39789–97. https://doi.org/10.1074/jbc.M404130200.

Sirviö, Anu, J. Spencer Johnston, Tom Wenseleers, and Pekka Pamilo. 2011. 'A High Recombination Rate in Eusocial Hymenoptera: Evidence from the Common Wasp Vespula Vulgaris'. *BMC Genetics* 12 (November): 95. https://doi.org/10.1186/1471-2156-12-95.

Slack, R. Douglas. 1976. 'Nest Guarding Behavior by Male Gray Catbirds'. *The Auk* 93 (2): 292–300.

Slater, Guy St C., and Ewan Birney. 2005. 'Automated Generation of Heuristics for Biological Sequence Comparison'. *BMC Bioinformatics* 6 (1): 31. https://doi.org/10.1186/1471-2105-6-31.

Smit, AFA, Hubley, R. *RepeatModeler Open-1.0*. 2008-2015 <http://www.repeatmasker.org>.

Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-4.0*. 2013-2015 <http://www.repeatmasker.org>.

Sogin, Mitchell L. 1991. 'Early Evolution and the Origin of Eukaryotes'. *Current Opinion in Genetics & Development* 1 (4): 457–463.

Sowig, Peter. 1996. 'Duration and Benefits of Biparental Brood Care in the Dung Beetle Onthophagus Vacca (Coleoptera: Scarabaeidae)'. *Ecological Entomology* 21 (1): 81–86. https://doi.org/10.1111/j.1365-2311.1996.tb00269.x.

Staerkle Michael, and Kölliker Mathias. 2008. 'Maternal Food Regurgitation to Nymphs in Earwigs (Forficula Auricularia)'. *Ethology* 114 (9): 844–50. https://doi.org/10.1111/j.1439-0310.2008.01526.x.

Standage, Daniel S., Ali J. Berens, Karl M. Glastad, Andrew J. Severin, Volker P. Brendel, and Amy L. Toth. 2016. 'Genome, Transcriptome and Methylome Sequencing of a Primitively Eusocial Wasp Reveal a Greatly Reduced DNA Methylation System in a Social Insect'. *Molecular Ecology* 25 (8): 1769–84. https://doi.org/10.1111/mec.13578.

Stanke, Mario. n.d. 'Gene Prediction with a Hidden Markov Model', 104.

Stanke, Mario, Oliver Keller, Irfan Gunduz, Alec Hayes, Stephan Waack, and Burkhard Morgenstern. 2006. 'AUGUSTUS: Ab Initio Prediction of Alternative Transcripts'. *Nucleic Acids Research* 34 (suppl_2): W435–39. https://doi.org/10.1093/nar/gkl200.

Stegmann, Ulrich E., and K. Eduard Linsenmair. 2002. 'Assessing the Semelparity Hypothesis: Egg-Guarding and Fecundity in the Malaysian Treehopper Pyrgauchenia Tristaniopsis'. *Ethology* 108 (10): 857–69. https://doi.org/10.1046/j.1439-0310.2002.00818.x.

Stein, Kenneth J., and Richard D. Fell. 1992. 'Seasonal Comparison of Weight, Energy Reserve, and Nitrogen Changes in Queens of the Baldfaced Hornet (Hymenoptera: Vespidae)'. *Environmental Entomology* 21 (1): 148–55. https://doi.org/10.1093/ee/21.1.148.

Stiller, J. W., and B. D. Hall. 1999. 'Long-Branch Attraction and the RDNA Model of Early Eukaryotic Evolution'. *Molecular Biology and Evolution* 16 (9): 1270–79. https://doi.org/10.1093/oxfordjournals.molbev.a026217.

Stockley, Paula, and Liane Hobson. 2016. 'Paternal Care and Litter Size Coevolution in Mammals'. *Proceedings of the Royal Society B: Biological Sciences* 283 (1829): 20160140. https://doi.org/10.1098/rspb.2016.0140.

Straka, Jakub, and Petr Bogusch. 2007. 'Apoidea: Apiformes (včely)'. *Acta Entomologica Musei Nationalis Pragae*, 60.

Sun, Y.-L., J.-F. Dong, C. Ning, P.-P. Ding, L.-Q. Huang, J.-G. Sun, and C.-Z. Wang. 2019. 'An Odorant Receptor Mediates the Attractiveness of Cis-Jasmone to Campoletis Chlorideae, the Endoparasitoid of Helicoverpa Armigera'. *Insect Molecular Biology* 28 (1): 23–34. https://doi.org/10.1111/imb.12523.

Székely, Tamás, James N. Webb, Alasdair I. Houston, and John M. McNamara. 1996. 'An Evolutionary Approach to Offspring Desertion in Birds'. In *Current Ornithology*, 271–330. Current Ornithology. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-5881-1_6.

Tallamy, D W, and T K Wood. 1986. 'Convergence Patterns in Subsocial Insects'. *Annual Review of Entomology* 31 (1): 369–90. https://doi.org/10.1146/annurev.en.31.010186.002101.

Tallamy, Douglas W. 2001. 'Evolution of Exclusive Paternal Care in Arthopods'. *Annual Review of Entomology* 46 (1): 139.

Tate, Peri H., and Adrian P. Bird. 1993. 'Effects of DNA Methylation on DNA-Binding Proteins and Gene Expression'. *Current Opinion in Genetics & Development* 3 (2): 226–31. https://doi.org/10.1016/0959-437X(93)90027-M.

Taylor, Steven J. 2017. 'Observations on Parental Care in the Family Aradidae (Heteroptera).' *The Great Lakes Entomologist* 21 (4): 5.

Tegoni, Mariella, Valérie Campanacci, and Christian Cambillau. 2004. 'Structural Aspects of Sexual Attraction and Chemical Communication in Insects'. *Trends in Biochemical Sciences* 29 (5): 257–64. https://doi.org/10.1016/j.tibs.2004.03.003.

Ter-Hovhannisyan, Vardges, Alexandre Lomsadze, Yury O. Chernoff, and Mark Borodovsky. 2008. 'Gene Prediction in Novel Fungal Genomes Using an Ab Initio Algorithm with Unsupervised Training'. *Genome Research* 18 (12): 1979–90. https://doi.org/10.1101/gr.081612.108.

Thompson, S. Nelson. 2003. 'Trehalose—the Insect 'Blood'Sugar'. *Advances in Insect Physiology* 31 (Supplement C): 205–285.

Timmers, Paul RHJ, Ninon Mounier, Kristi Lall, Krista Fischer, Zheng Ning, Xiao Feng, Andrew D. Bretherick, David W. Clark, Xia Shen, and Tonu Esko. 2019. 'Genomics of 1 Million Parent Lifespans Implicates Novel Pathways and Common Diseases and Distinguishes Survival Chances'. *Elife* 8: e39856.

Tørresen, Ole K., Bastiaan Star, Pablo Mier, Miguel A. Andrade-Navarro, Alex Bateman, Patryk Jarnot, Aleksandra Gruca, et al. 2019. 'Tandem Repeats Lead to Sequence Assembly Errors and Impose Multi-Level Challenges for Genome and Protein Databases'. *Nucleic Acids Research* 47 (21): 10994–6. https://doi.org/10.1093/nar/gkz841.

Trumbo, Stephen T. 1990. 'Reproductive Benefits of Infanticide in a Biparental Burying Beetle Nicrophorus Orbicollis'. *Behavioral Ecology and Sociobiology* 27 (4): 269–73. https://doi.org/10.1007/BF00164899.

Trumbo, Stephen T., Zhi-Yong Huang, and Gene E. Robinson. 1997. 'Division of Labor between Undertaker Specialists and Other Middle-Aged Workers in Honey Bee Colonies'. *Behavioral Ecology and Sociobiology* 41 (3): 151–63. https://doi.org/10.1007/s002650050374.

Tullberg, Birgitta S., Malin Ah-King, and Hans Temrin. 2002 'Phylogenetic Reconstruction of Parental-Care Systems in the Ancestors of Birds'. Accessed 11 August 2014. http://www.zoologi.su.se/research/tullberg/papers/02-PTRSB-357.pdf.

Venter, J. Craig, Mark D. Adams, Eugene W. Myers, Peter W. Li, Richard J. Mural, Granger G. Sutton, Hamilton O. Smith, et al. 2001. 'The Sequence of the Human Genome'. *Science* 291 (5507): 1304–51. https://doi.org/10.1126/science.1058040.

Verdone, Loredana, Micaela Caserta, and Ernesto Di Mauro. 2005. 'Role of Histone Acetylation in the Control of Gene Expression'. *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire* 83 (3): 344–53. https://doi.org/10.1139/o05-041.

Vestergaard, Klaus, and Carin Magnhagen. 1993. 'Brood Size and Offspring Age Affect Risk-Taking and Aggression in Nest-Guarding Common Gobies'. *Behaviour* 125 (3–4): 233–43. https://doi.org/10.1163/156853993X00263.

Vincze, Orsolya, Tamás Székely, Clemens Küpper, Monif AlRashidi, Juan A. Amat, Araceli Argüelles Ticó, Daniel Burgas, et al. 2013. 'Local Environment but Not Genetic Differentiation Influences Biparental Care in Ten Plover Populations'. *PLOS ONE* 8 (4): e60998. https://doi.org/10.1371/journal.pone.0060998.

Weitekamp, Chelsea A., Romain Libbrecht, and Laurent Keller. 2017. 'Genetics and Evolution of Social Behavior in Insects'. *Annual Review of Genetics* 51 (1): 219–39. https://doi.org/10.1146/annurev-genet-120116-024515.

West, Mary Jane, and Richard D. Alexander. 1963. 'Sub-Social Behavior in a Burrowing Cricket Anurogryllus Muticus (De Geer)'. *Ohio Journal of Science*.

Williams, George C. 1966. 'Natural Selection, the Costs of Reproduction, and a Refinement of Lack's Principle'. *American Naturalist* 100 (916): 687–690.

Wilson, Edward O. 1971. *The Insect Societies.* Cambridge, Massachusetts, USA, Harvard University Press

Winkelman, Dana L. 1996. 'Reproduction under Predatory Threat: Trade-Offs between Nest Guarding and Predator Avoidance in Male Dollar Sunfish (Lepomis Marginatus)'. *Copeia* 1996 (4): 845–51. https://doi.org/10.2307/1447646.

Wittkopp, Patricia J., Kathy Vaccaro, and Sean B. Carroll. 2002. 'Evolution of Yellow Gene Regulation and Pigmentation in Drosophila'. *Current Biology* 12 (18): 1547–56. https://doi.org/10.1016/S0960-9822(02)01113-2.

Wolfe, Scot A., Lena Nekludova, and Carl O. Pabo. 2000. 'DNA Recognition by Cys2His2 Zinc Finger Proteins'. *Annual Review of Biophysics and Biomolecular Structure* 29 (1): 183–212. https://doi.org/10.1146/annurev.biophys.29.1.183.

Wood, Adam, Nevan J. Krogan, Jim Dover, Jessica Schneider, Jonathan Heidt, Marry Ann Boateng, Kimberly Dean, et al. 2003. 'Bre1, an E3 Ubiquitin Ligase Required for Recruitment and Substrate Selection of Rad6 at a Promoter'. *Molecular Cell* 11 (1): 267–74. https://doi.org/10.1016/s1097-2765(02)00802-x.

Woodard, S. Hollis, Brielle J. Fischman, Aarti Venkat, Matt E. Hudson, Kranthi Varala, Sydney A. Cameron, Andrew G. Clark, and Gene E. Robinson. 2011. 'Genes Involved in Convergent Evolution of Eusociality in Bees'. *Proceedings of the National Academy of Sciences* 108 (18): 7472–77. https://doi.org/10.1073/pnas.1103457108.

Wu, Qi, and Mark R. Brown. 2006. 'Signaling and Function of Insulin-Like Peptides in Insects'. *Annual Review of Entomology* 51 (1): 1–24. https://doi.org/10.1146/annurev.ento.51.110104.151011.

Wu, Thomas D., Jens Reeder, Michael Lawrence, Gabe Becker, and Matthew J. Brauer. 2016. 'GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality'. In *Statistical Genomics: Methods and Protocols*, edited by Ewy Mathé and Sean Davis, 283–334. Methods in Molecular Biology. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3578-9_15.

Wu, Thomas D., and Colin K. Watanabe. 2005. 'GMAP: A Genomic Mapping and Alignment Program for MRNA and EST Sequences'. *Bioinformatics* 21 (9): 1859–75. https://doi.org/10.1093/bioinformatics/bti310.

Wyatt, G. R., and G. F. Kalf. 1957. 'The chemistry of insect hemolymph II. Trehalose and other carbohydrates'. *Journal of General Physiology* 40 (6): 833–47. https://doi.org/10.1085/jgp.40.6.833.

Xiao, Tiaojiang, Cheng-Fu Kao, Nevan J. Krogan, Zu-Wen Sun, Jack F. Greenblatt, Mary Ann Osley, and Brian D. Strahl. 2005. 'Histone H2B Ubiquitylation Is Associated with Elongating RNA Polymerase II'. *Molecular and Cellular Biology* 25 (2): 637–51. https://doi.org/10.1128/MCB.25.2.637-651.2005.

Xie, Feng-Yun, You-Min Hu, Bo Wu, and Yan Wang. 2016. 'A Generalized Hidden Markov Model and Its Applications in Recognition of Cutting States'. *International Journal of Precision Engineering and Manufacturing* 17 (11): 1471–82. https://doi.org/10.1007/s12541-016-0173-y.

Yamashita, Kentaro, Miki Shinohara, and Akira Shinohara. 2004. 'Rad6-Bre1-Mediated Histone H2B Ubiquitylation Modulates the Formation of Double-Strand Breaks during Meiosis'. *Proceedings of the National Academy of Sciences of the United States of America* 101 (31): 11380–85. https://doi.org/10.1073/pnas.0400078101.

Yan, Hua, Comzit Opachaloemphan, Giacomo Mancini, Huan Yang, Matthew Gallitto, Jakub Mlejnek, Kevin Haight, et al. 2017. 'Olfactory Receptors Are Required For Social Behavior And Neural Plasticity In Ants, As Evidenced By CRISPR-Mediated Gene Knockout'. *BioRxiv*, May, 142232. https://doi.org/10.1101/142232.

Yanega, D. 1989. 'Caste Determination and Differential Diapause within the First Brood of *Halictus Rubicundus* in New York (Hymenoptera: Halictidae)'. *Behavioral Ecology and Sociobiology* 24 (2): 97–107. https://doi.org/10.1007/BF00299641.

Yi, Soojin V., and Michael A. D. Goodisman. 2009. 'Computational Approaches for Understanding the Evolution of DNA Methylation in Animals'. *Epigenetics* 4 (8): 551–56. https://doi.org/10.4161/epi.4.8.10345.

Zhou, Xiaofan, Antonis Rokas, Shelley L. Berger, Jürgen Liebig, Anandasankar Ray, and Laurence J. Zwiebel. 2015. 'Chemoreceptor Evolution in Hymenoptera and Its Implications for the Evolution of Eusociality'. *Genome Biology and Evolution* 7 (8): 2407–2416.

Zink, Andrew G. 2003. 'Quantifying the Costs and Benefits of Parental Care in Female Treehoppers'. *Behavioral Ecology* 14 (5): 687–93. https://doi.org/10.1093/beheco/arg044.

———. 2005. 'The Dynamics of Brood Desertion among Communally Breeding Females in the Treehopper, Publilia Concava'. *Behavioral Ecology and Sociobiology* 58 (5): 466–73. https://doi.org/10.1007/s00265-005-0952-4.