

# Endosymbiosis, Co-option and Recombination: Mechanisms to generate Evolutionary Novelty in the Ant *Cardiocondyla obscurior*



DISSERTATION ZUR ERLANGUNG DES  
DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)  
DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN  
DER UNIVERSITÄT REGENSBURG

vorgelegt von

**Antonia Klein**

aus

Mallersdorf-Pfaffenberg

im Jahr

2015



# Endosymbiosis, Co-option and Recombination: Mechanisms to generate Evolutionary Novelty in the Ant *Cardiocondyla obscurior*



DISSERTATION ZUR ERLANGUNG DES  
DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)  
DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN  
DER UNIVERSITÄT REGENSBURG

vorgelegt von

**Antonia Klein**

aus

Mallersdorf-Pfaffenberg

im Jahr

2015





Das Promotionsgesuch wurde eingereicht am: 11. 09. 2015

Die Arbeit wurde angeleitet von: Dr. rer. nat. Jan Oettler

Unterschrift:



*“To myself I am only a child playing on the beach, while vast oceans of truth lie  
undiscovered before me.”*

*Isaac Newton*



# Manuscripts that contributed to this Thesis and Author's contribution

## Chapter 2:

Klein A., Schrader L., Gil R., Manzano-Marín A., Flórez L., Wheeler D., Werren J. H., Latorre A., Heinze J., Kaltenpoth M., Moya A. & Oettler, J. (2015) **A novel intracellular mutualistic bacterium in the invasive ant *Cardiocondyla obscurior*.** *The ISME Journal*.

Author's contribution: Designed the study: AK, LS, JO; Wrote the manuscript: AK, LS, JH, JO; Performed the experiments: AK, JO; Analysed the data: AK, LS; Prokaryote sequence annotation and HGT detection: DW, JW; Westeberhardia genome assembly, functional annotation and phylogenomic analysis: AL, AM, AMM, RG; Westeberhardia localization (FISH): LF, MK; All authors read, reviewed and accepted the final version of the manuscript.

## Chapter 3:

Klein A., Schultner E., Lowak H., Schrader L., Heinze J. & Oettler J. **Co-option of sex differentiation pathways facilitates the evolution of social insect polyphenism.** *Manuscript*.

Author's contribution: JO and AK designed the study. AK, HL and ES performed the experiments. AK and ES analysed the data. AK, ES, LS, JH and JO wrote the manuscript. All authors read and accepted the final version of the manuscript.

## Chapter 4:

Klein A., Rueppell O., Heinze J. & Oettler J. **Local recombination rate is positively correlated with genetic diversity and negatively linked to transposable element content in an inbred ant.** *Manuscript*.

Author's contribution: JO and AK designed the study. AK performed the experiments. AK and OR analysed the data. AK, OR, JH and JO wrote the manuscript. All authors read and accepted the final version of the manuscript.

In addition to the three manuscripts that are the basis of this thesis, I contributed to the following papers:

Schrader L., Kim J. W., Ence D., Zimin A., Klein A., Wyschetzki K., Weichselgartner T., Kemena C., Stökl J., Schultner E., Wurm Y., Smith C. D., Yandell M., Heinze J., Gadau J. & Oettler J. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nat. Comms.* **5**: 5495

Schrempf A., Wyschetzki K. von, Klein A., Schrader L., Oettler J. & Heinze J. (2015). Mating with an allopatric male triggers immune response and decreases longevity of ant queens. *Mol. Ecol.* **24(14)**: 3618–3627.

# Contents

<b>1</b>	<b>General Introduction</b>	<b>1</b>
1.1	Evolutionary Novelties . . . . .	1
1.2	The holobiont concept . . . . .	3
1.3	Endosymbionts as source for evolutionary novelties . . . . .	5
1.4	The hourglass model of insect sex determination . . . . .	7
1.5	The evolutionary potential of the master switch gene <i>doublesex</i> . . . . .	9
1.6	Homologous recombination accelerates evolution . . . . .	11
1.6.1	Genetic linkage mapping . . . . .	12
1.6.2	Eusociality as driver of recombination? . . . . .	13
1.7	Ergatoid males as evolutionary novelties in the genus <i>Cardiocondyla</i> . . . . .	15
1.8	The model organism <i>Cardiocondyla obscurior</i> . . . . .	16
1.9	Aims of this thesis . . . . .	17
<b>2</b>	<b>A novel intracellular mutualistic bacterium in the invasive ant <i>Cardiocondyla obscurior</i></b>	<b>18</b>
2.1	Abstract . . . . .	19
2.2	Introduction . . . . .	19
2.3	Material and Methods . . . . .	20
2.4	Results . . . . .	24
2.5	Discussion . . . . .	32
2.6	Acknowledgements . . . . .	37
<b>3</b>	<b>Co-option of sex differentiation pathways</b>	<b>39</b>
3.1	Abstract . . . . .	40
3.2	Introduction . . . . .	40
3.3	Results . . . . .	42
3.4	Discussion . . . . .	51
3.5	Material and Methods . . . . .	53
<b>4</b>	<b>Local recombination rate in <i>Cardiocondyla obscurior</i></b>	<b>57</b>
4.1	Abstract . . . . .	58
4.2	Introduction . . . . .	58
4.3	Material and methods . . . . .	60
4.4	Results . . . . .	66
4.5	Discussion . . . . .	74
<b>5</b>	<b>General Discussion</b>	<b>79</b>
5.1	Mechanisms to generate evolutionary novelties in <i>C. obscurior</i> . . . . .	79
5.2	Endosymbionts . . . . .	80

5.2.1	<i>Wolbachia</i> - a further player influencing <i>C. obscurior</i> biology? . . .	80
5.2.2	The tripartite interaction between <i>C. obscurior</i> , <i>Westeberhardia</i> and <i>Wolbachia</i> . . . . .	85
5.3	Sex determination in <i>Cardiocondyla obscurior</i> . . . . .	88
5.3.1	Absence of single-locus sex determination in <i>C. obscurior</i> . . . .	88
5.3.2	Evolution of alternative sex determination mechanisms in <i>C. ob-</i> <i>scurior</i> . . . . .	91
5.4	Supressed recombination in supergenes . . . . .	94
5.5	Major challenges for the future of <i>C. obscurior</i> research . . . . .	96
<b>6</b>	<b>Summary</b>	<b>97</b>
<b>7</b>	<b>Zusammenfassung</b>	<b>99</b>
	<b>Supplement Chapter 2</b>	<b>S1</b>
	<b>Supplement Chapter 3</b>	<b>S11</b>
	<b>Supplement Chapter 4</b>	<b>S26</b>
	<b>Bibliography</b>	<b>iii</b>
	<b>Acknowledgements</b>	<b>xxix</b>



# 1 General Introduction

## 1.1 Evolutionary Novelties

Biologists have long recognized the importance of evolution by gradual change and natural selection (Darwin, 1872). However, a major unresolved question is how major evolutionary novelty emerges (Brigandt & Love, 2012; Pigliucci, 2012). When challenged by St. George J. Mivart, already Darwin struggled to explain the origin of evolutionary novelties such as the mammary gland (Wagner & Lynch, 2010). Other examples for evolutionary novelties are beetle horns (Moczek, 2005), flight of vertebrates and invertebrates (Prum, 2005), the evolution of *Hox* genes (Wagner *et al.*, 2003) or the emergence of eusociality (Bourke, 2011). These examples show that the term "*evolutionary novelty*" is inconsistently used for a wide range of morphological, behavioral and genetic traits (Brigandt & Love, 2012; Pigliucci, 2012). While broader definitions include structures that evolved from homologous traits, e.g. the wings of birds that originated from the forelimbs of reptiles, stricter definitions apply the concept only to non-homologous structures such as the feathers of birds (Wagner & Lynch, 2010). However, neither the formation of wings, feathers, nor the evolution of flight as evolutionary novelties can be investigated independently, and without considering the other. The inconsistency of the definitions of evolutionary novelty is not just a conceptual problem, but also affects how we analyze the evolutionary and developmental origin of novelties. For example, if only non-homologous traits are defined as novelties, quantitative change would not suffice to explain the origin of discrete novelties (West-Eberhard, 2003). Consequently a consensus definition may not be desirable after all (Brigandt & Love, 2012).

Pigliucci (2012) defined evolutionary novelties as "*[...] new traits or behaviors, or novel combinations of previously existing traits or behaviors, arising during the evolution of a lineage, and that perform a new function within the ecology of that lineage.*". Other authors also highlight the importance of ecology by defining novelties as traits which enable the transition between adaptive peaks on a fitness landscape (Hallgrímsson *et al.*, 2012; Erwin, 2012). Such evolutionary novelties can be environmentally or mutationally

induced (West-Eberhard, 2005), although the importance of random mutations on evolution as proposed by the New Synthesis has been questioned by recent evidence stressing the importance of environmental factors (West-Eberhard, 2009).

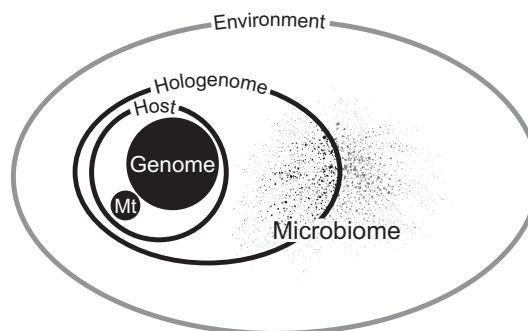
Since the advent of the genomic era it has been shown that the birth of novel genes may contribute to the adaptive origin of evolutionary novelties. Mechanisms giving rise to novel genes go far beyond simple gene duplication and neo-functionalization, but involve retrotransposition, acquisition of genomic parasites, recruitment of ancestrally non-functional sequence and shuffling of existing genetic material via recombination (Kaessmann, 2010; Pigliucci, 2012). Moreover, it became clear that novel traits often involve molecular changes in regulatory networks, leading to the advent of a new holistic field of evolutionary research, aptly termed eco-evo-devo (ecological evolutionary developmental biology). For instance, the lanterns of fireflies, which are a novelty according to the strictest definition, have evolved by rewiring of ancestral regulatory gene networks through acquisition of novel targets (Stansburg & Moczek, 2014). The connection of novel input signals with novel phenotypes via rewiring of existing genes is called "*co-option*" (West-Eberhard, 2005). Especially pleiotropic genes are good candidates due to their interaction with multiple targets (Torday, 2015).

Other sources for evolutionary novelties are transposable elements (TEs) (Stapley *et al.*, 2015). TEs are stretches of DNA that migrate in the genome and are suggested to be more important for character origination than for character modification (Wagner & Lynch, 2010). Their evolutionary potential lies in their frequent interaction with transcription factors. These transcription factor binding sites have the potential to become new promoters, enhancers or insulators when inserted in new locations in the genome together with TEs. Furthermore, exon shuffling through retrotransposons may be a general mechanism to create novel genes (Moran *et al.*, 1999).

A further important mechanism for generating novelty is the shuffling of genomes by recombination, which has been shown to be adaptive in empirical tests (Rice, 2002). In particular, recombination between non-homologous sequence may lead to extensive genomic rearrangements and thus drive genome evolution. Lastly, a less recognized mechanism to generate evolutionary novelty is the acquisition of novel genetic elements via horizontal gene transfers and the acquisition of microbial symbionts from the environment. The combination of preexisting independent genomic units into a single evolutionary unit under selection is verbalized as the concept of the "holobiont".

## 1.2 The holobiont concept

Microbial symbionts are universal in eukaryotes and it is well known that symbiosis with prokaryotes shapes evolution of eukaryotes, probably best exemplified by the endosymbiont-theory addressing the origin of chloroplasts and mitochondria from cyanobacteria and proteobacteria (Margulis, 1993). Symbiosis in a wide sense includes all persistent associations between two species (Douglas, 2010). Depending on the resulting fitness outcome, the association is classified as mutualistic (i.e. positive for both partners), commensalistic (i.e. positive for one and neutral for the other partner) or parasitic (i.e. harmful for one partner). The unit comprising host and associated organisms is referred to as the so-called "*holobiont*", a concept introduced by Margulis (1993). In their hologenome theory of evolution Zilber-Rosenberg *et al.* (2008) suggested that the host plus all its associated microbiota constitutes the unit under selection. Accordingly, the entity of host genomes, i.e. nuclear genome, mitochondrial genome (and optionally chloroplast genome), plus the genomes of all associated microbiota is referred to as the "*hologenome*" (Figure 1.1).



**Figure 1.1:** The hologenome is composed of the nuclear and mitochondrial (Mt) host genome and the genomes of all associated microbiota (from Brucker & Bordenstein, 2012).

A precondition for the hologenome theory of evolution is the stable transmission of symbionts to the next host generation. Although some gut microbiota may be transmitted in a stable manner (Sanders *et al.*, 2014; Abdul Rahman *et al.*, 2015), gut microbiota are in general more affected by the environment and in particular by diet (Hu *et al.*, 2014; Pérez-Cobas *et al.*, 2015). In contrast, intracellular endosymbionts are transmitted vertically to the next generation and some of them may even blur the borders between symbionts and organelles (Tamames *et al.*, 2007). Endosymbionts are known to be important players in insect evolution, being one factor that enabled the enormous diversification of this clade (Moya *et al.*, 2008; Gil *et al.*, 2010). The intensity of insect-bacteria symbiosis may range from facultative to obligate. Whereas

infection with facultative (secondary, S) symbionts is not fixed in the host population and the host is able to survive without the symbiont, infection with obligate (primary, P) symbionts is fixed and the host relies on the symbiont for survival or reproduction. Table 1.1 summarizes characteristics of facultative versus obligate symbionts.

**Table 1.1:** Comparison between facultative, secondary (S) symbionts and obligate, primary (P) symbionts.

characteristics	facultative symbionts	obligate symbionts
infection density	not perfect	perfect
transmission	not perfect; vertical and horizontal	perfect; normally only vertical
phylogenetic congruence	absent	present
tissue specificity	low	high
genome reduction	no	yes
metabolic interaction	weak	strong
symbiosis	beneficial-harmful	beneficial

While obligate symbionts have often evolved mutualistic association with the host, facultative symbionts may be either mutualistic or parasitic. However, in both cases the symbionts may have significant impact on the host. For example tse tse flies (genus *Glossina*) harbor the obligate (primary) symbiont *Wigglesworthia glossinidia*, and the secondary (facultative) symbionts *Sodalis glossinidius* and *Wolbachia*. *Wigglesworthia* provides vitamins to the host and is essential for viability and fertility (Michalkova *et al.*, 2014), while *Sodalis* are suggested to increase the susceptibility of flies to trypanosome infections (Welburn *et al.*, 1993; Dale & Welburn, 2001). Bacteria of the *Sodalis* clade seem to be especially promising for research on host-symbiont interaction, because here the transition from facultative to obligate symbiosis is an ongoing process, which can be directly observed and promises major advances in understanding the principles of host-symbiont-coevolutionary processes (Clayton *et al.*, 2012). The strict classification in mutualistic and parasitic (pathogenic) bacteria has been criticized because it neglects the dynamic aspect of symbiosis (Pérez-Brocal *et al.*, 2013). Thus it is of great importance to evaluate possible fitness benefits of the symbiosis on the host, which may be harmful and beneficial at the same time depending on external conditions (Pérez-Brocal *et al.*, 2013).

## 1.3 Endosymbionts as source for evolutionary novelties

Ivan Wallin hypothesized that new species could arise through the acquisition of bacterial endosymbionts (Wallin, 1927). And indeed symbiosis with prokaryotes may drive eukaryote evolution in many different ways and is thought to enhance speciation events (Brucker & Bordenstein, 2012). Even when reproductive isolation is founded in host-encoded genes, an involvement of endosymbionts can not be ruled out, especially in the case of genes associated with immunity (Brucker & Bordenstein, 2012).

In mutualistic interactions, endosymbionts can confer novel traits to the host, e.g. via upgrading the diet of their hosts (Feldhaar, 2011). Nutritional upgrading has enormous potential for the evolution of novel traits because it allows the host to access new food sources of low quality or unbalanced compounds, e.g. phloem-sap (Baumann, 2005), blood (Michalkova *et al.*, 2014) or wood (Brune, 2014). For example, aphids feeding on protein-poor phloem-sap upgrade their diet with the help of *Buchnera* symbionts, which provide essential amino acids to their hosts (Buchner, 1965; Baumann, 2005; Douglas, 2006). The primary symbiont of blood-feeding tsetse flies *Wigglesworthia* has been shown to supplement the hosts' diet with vitamin B<sub>6</sub> (Michalkova *et al.*, 2014). Likewise, *Blochmannia* residing in bacteriocytes intercalated in the mid-gut tissue of *Camponotus* ants provide the host with essential amino acids and may also help in nitrogen recycling (Feldhaar *et al.*, 2007). In olive flies (*Bactrocera oleae*), which feed on carbohydrate rich but nitrogen poor diets, bacterial symbionts have been shown to metabolize urea into an available nitrogen source for the fly and thus significantly elevate egg production (Ben-Yosef *et al.*, 2014).

Besides nutritional upgrading, endosymbionts can confer defensive mechanisms to the host, thereby increasing the host's resistance against parasites, pathogens or fungi. For example, *Streptomyces* bacteria residing in specialized antennal glands in female bee-wolves (*Philanthus triangulum*) are applied to brood cells by the females where their fungizide secretions protect the larvae from fungal infestation (Kaltenpoth *et al.*, 2005). Secondary symbionts of aphids have been shown to protect aphids against parasitoids (Oliver *et al.*, 2003) and have additionally been suggested to play a role in adaptation to abiotic environments by providing enhanced stress tolerance to increased temperatures (Russell & Moran, 2006). In termites, which harbor a diversity of microbiota allowing for digestion of wood (Brune, 2014), symbionts also protect the host against fungal pathogens (Rosengaus *et al.*, 2014). Gut microbiota have also been shown to confer defensive mechanisms in bumble bees (Koch & Schmid-Hempel, 2011). However,

the tsetse fly secondary endosymbiont *Sodalis* is suggested to increase trypanosome susceptibility of its host (Aksoy, 2000; Aksoy & Rio, 2005; Wang *et al.*, 2013a), emphasizing the importance of evaluating both, positive as well as negative effects of endosymbionts on host immunity.

Another causative agent for the emergence of novel traits through endosymbiosis is horizontal gene transfer (HGT), i.e. the transfer of genes from the symbiont genome into the host genome putatively via retroviral or retrotransposable element activity (Wheeler *et al.*, 2013). HGTs may provide new genetic pathways and metabolic capacities for the host and strengthen the symbiosis by allowing symbiont genome reduction (Hotopp *et al.*, 2007; Alves *et al.*, 2013). It is not necessary that HGTs incorporated into the host genome originate from the present symbiont. Instead HGTs of ancestral symbionts no longer present in the host may take part in holobiont metabolic networks and thus influence holobiont evolution (Husník *et al.*, 2013; Nikoh *et al.*, 2010).

In addition to the transfer of novel traits by mutualistic symbionts, parasitic interactions also bear potential for evolutionary novelties in a wide sense, i.e. speciation events. Here constant coevolution may influence host evolution via so-called "red queen" effects. A prevalent target of facultative symbionts is host reproduction. In general, only females transmit symbionts via the cytoplasm of the oocyte, although more and more examples for paternal transmission are being reported (Moran & Dunbar, 2006; Damiani *et al.*, 2008; Watanabe *et al.*, 2014; de Vooght *et al.*, 2015). Under the premise of purely maternal transmission, bacteria acting as reproductive manipulators will try to increase the proportion of female offspring by different mechanisms, e.g. cytoplasmic incompatibility (CI), parthenogenesis induction, male-killing or feminization of genetic males (Cordaux *et al.*, 2011). The textbook example for a reproductive manipulator is the alpha-proteobacterium *Wolbachia*, which is estimated to infect 66 % of all insect species (Hilgenboecker *et al.*, 2008). It was shown that CI induced by *Wolbachia* preceded the evolution of other post-mating reproductive barriers in *Nasonia* and thus induced speciation events (Bordenstein *et al.*, 2001). Interestingly, in *Nasonia* the gut microbiome was also shown to be a reproductive barrier (Brucker & Bordenstein, 2013). The genus *Wolbachia* is diverse and some *Wolbachia* have evolved mutualistic relationships with their hosts (Hosokawa *et al.*, 2010; Ye *et al.*, 2013; Zug & Hammerstein, 2015) even when coincidentally affecting host reproduction, again illustrating the dynamic and evolutionary potential of symbioses.

It is hypothesized that reproductive manipulators such as *Wolbachia* may have been a driver in the evolution of haplodiploidy (Engelstädter & Hurst, 2006; Kuijper & Pen, 2010). Under the influence of male-killing as well as CI-inducing endosymbionts,

hosts are predicted to evolve increased viability of haploid males, which may have been a precursor for the evolution of haplodiploidy (Engelstädter & Hurst, 2006). This applies especially under high levels of inbreeding, as a male will benefit from being killed and reallocating resources to his sisters if relatedness is high (Kuijper & Pen, 2010; Engelstädter & Hurst, 2014). However, direct evidence is scarce and Kuijper & Pen (2010) suggest that mutualistic endosymbiosis provides a much more promising route to evolution of haplodiploidy than parasitic endosymbiosis.

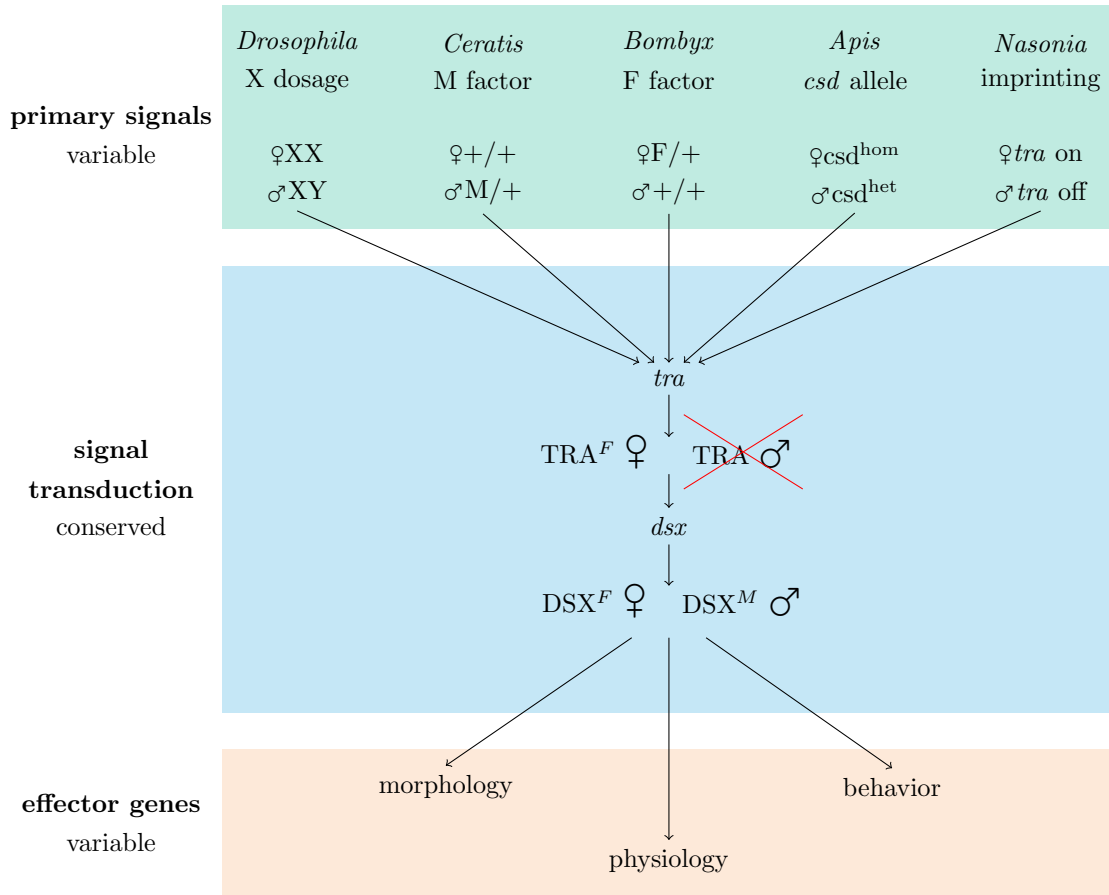
Chapter 2 of this thesis addresses a previously unknown bacterial species (*Candidatus Westeberhardia cardiocondylae*), which lives in symbiosis with the host *Cardiocondyla obscurior* and may provide novel traits to its host. By provisioning with tyrosine-precursors, *Westerberhardia* may upgrade the diet of the ant, allow the exploitation of novel food sources and habitats and play a role in the invasive success of *C. obscurior*. The second major bacterial species residing in *C. obscurior*, *Wolbachia*, will be addressed in detail in the General Discussion (Chapter 5).

## 1.4 The hourglass model of insect sex determination

Sexual reproduction is a major biological phenomenon that provides fitness benefits by generating offspring genetic diversity and removing deleterious mutations. Sexual reproduction in most cases is accompanied by sexual dimorphisms, such as sex-specific morphologies (e.g. different sizes, ornaments), physiologies (e.g. different lifespans), or behaviors (e.g. sex-specific mating or brood care behavior). Although sex-specific traits evolve quickly and thus are often species-specific, their expression is regulated by a single family of master switch genes - the *doublesex/mab-3* related (*Dmrt*) genes - all across the metazoa (Zarkower, 2001; Kopp, 2012). *Dmrt* genes act pleiotropically as transcription factors on a wide range of target genes which in turn induce tissue sex-specific development (Matson & Zarkower, 2012). All genes in this family show a relatively conserved domain, the DM domain, which codes for a zinc-finger-motif binding as a transcription factor in the minor groove of DNA, whereas the residual sequence of *Dmrt* genes shows almost no conservation across species (Kopp, 2012).

Whereas in mammals *Dmrt* genes act as activators or repressors, *doublesex* (*dsx*) genes in insects exhibit alternative splicing as an additional regulatory switch, which leads to male and female-specific DSX proteins that induce sex-specific tissue development, physiology, and behavior. At the onset of the insect sex determination cascade, primary signals that initiate sex determination are diverse across species, e.g. single-locus sex determination (sl-sd) in *Apis* (Beye *et al.*, 2003), maternal imprinting in *Nasonia*

(Verhulst *et al.*, 2010), X chromosome dosage in *Drosophila* (Erickson & Quintero, 2007), male determining M factor in *Musca domestica* (Hamm *et al.*, 2015) and *Ceratitis* (Willhoeft, 1996) or female determining W-chromosomes in lepidopterans (Fujii & Shimada, 2007). These primary signals result in an ON/OFF signal of *transformer* (*tra*). In females the primary signals induce an active TRA protein, whereas it results in a non-functional TRA in males, e.g. by an early in-frame stop codon. Depending on the existence of a functional TRA protein *dsx* is differentially spliced and its sex-specific isoforms act on a variety of target genes to regulate sex-specific development. Accordingly, sex determination in insects is often described as an hourglass model, with variable primary signals, a conserved part containing *tra* and *dsx* and a wide range of downstream effector genes (Bopp *et al.*, 2014, Figure 1.2).



**Figure 1.2: The hourglass model of insect sex determination evolution, adapted from Bopp *et al.* (2014).** The model consists of variable instruction signals at the top of the cascade (green), variable effector genes (orange) at the bottom, but a conserved transduction part in the middle (blue), consisting of *transformer* (*tra*) and *doublesex* (*dsx*). It was suggested that the *tra-dsx* transduction module (blue) constitutes the ancestral core of the pathway, whereas evolutionary diversification took place in primary signals and effector genes (Bopp *et al.*, 2014).



## 1.5 The evolutionary potential of the master switch gene *doublesex*

Although sex-specific splicing of *dsx* is conserved across insects, the observed splicing patterns vary across species: In *Drosophila melanogaster*, *Apis mellifera*, *Bombyx mori* and *Solenopsis invicta* the male isoform is generated by skipping a female-specific exon (Baker & Wolfner, 1988; Ohbayashi *et al.*, 2001; Cho *et al.*, 2007; Nipitwattanaphon *et al.*, 2014), whereas in *Nasonia* the female-specific isoform is generated by excluding an intron from the primary transcript (Oliveira *et al.*, 2009). The sex-specific isoforms DSX<sup>F</sup> and DSX<sup>M</sup> act as transcription factors on a wide range of targets. For example, they have been shown to regulate sex-specific gene expression in the developing *Aedes aegypti* brain (Tomchaney *et al.*, 2014), to regulate genital disc formation in *Drosophila* (Chatterjee *et al.*, 2011) and to regulate female reproductive physiology such as oocyte development or egg production in *Tribolium* (Shukla & Palli, 2012a). Traits are expected to be differentially expressed not only between sexes, but also between female queen and worker castes of eusocial insects. For example brain gene expression patterns have been shown to be associated with caste and reproductive status in *Apis* (Grozingier *et al.*, 2007). Reproductive division of labor is the basis of eusociality, with queens reproducing and workers who refrain from reproduction to gain indirect fitness benefits (Hamilton, 1964). Consequently reproduction is a trait that should be regulated caste-specifically. Thus differential developmental trajectories are required for tissue-specific development not only between females and males, but also between queens and workers. This might be regulated by changes of cis-regulatory sequences of *dsx* target genes on the one hand (e.g. Williams *et al.*, 2008) or by changes in *dsx* expression itself on the other hand.

Studies in horned-beetles of the genus *Onthopagus* demonstrate that *dsx* has the potential to evolve additional regulatory function. *Onthopagus* males possess exaggerated horns and RNAi experiments revealed that *dsx* is linked to horn development (Kijimoto *et al.*, 2012), a sex-specific trait corresponding to the original sex differentiation function of *dsx*. However, development of male horns seems to be a nutrition-dependent trait, as males under optimal feeding conditions develop full-size horns used for fighting, whereas males facing poor feeding conditions during larval development develop much smaller horns and emerge as nonaggressive sneaker males (Moczek & Emlen, 2000). Kijimoto *et al.* (2012) showed that expression of the male *dsx* isoform correlates with horn size, suggesting that *dsx* function has undergone rapid evolution in this species. Further evidence for the evolutionary potential of *dsx* comes from recruitment of *dsx*

into the pathway of sex comb development, a recent evolutionary innovation in the genus *Drosophila* (Tanaka *et al.*, 2011) and *dsx*-mediated control of a supergene shaping wing-patterning among polymorphic females in the genus *Papilio* (Kunte, 2014). Moreover Eirín-López & Sánchez (2015) have demonstrated that evolutionary rates of *dsx* among insects is much higher than previously thought. In *Drosophila*, *dsx* evolves even more rapidly than the master regulator *sex-lethal* (Eirín-López & Sánchez, 2015), questioning the hourglass-model of sex-determination evolution (see section 1.4).

Eusociality has evolved repeatedly in the social Hymenoptera. This repeated independent evolution of morphologically, physiologically and behaviorally distinct castes (queens and workers) suggests that the same regulatory pathways may be differentially regulated to express these phenotypes. Indeed, studies on caste-specific gene expression revealed conserved genes to be differentially regulated among castes, e.g. the TOR pathway (Patel *et al.*, 2007), sphingolipid metabolism (Schrader *et al.*, 2015), or *Epidermal growth factor receptor* (*Egfr*) (Alvarado *et al.*, 2015). An increasing number of studies found caste-specificity on different regulatory levels, e.g. genetic components (Schwander & Keller, 2008; Smith *et al.*, 2008), hormones (Wheeler, 1990), nutrition (Kamakura, 2011), methylation patterns (Simola *et al.*, 2013) and different transcription factor binding site landscapes (Simola *et al.*, 2013; Schrader *et al.*, 2015). However, uncertainty remains about primary signals inducing caste-specific development, as it is difficult to unravel cause from consequence. Advance is hindered by the use of different development stages, tissue types and methods of data analyses. For example Morandin *et al.* (2015) could show that gene expression patterns depend on developmental stage. A further drawback in most species is the missing level of comparison, as gene expression differences between queens and workers may be assigned to several traits including reproductive physiology, morphology (winged/wingless) and brain gene expression (behaviors like brood-care, foraging, mating). The model system *C. obscurior* promises great advances due to the possibility of a four level comparison (Schrader *et al.*, 2015). In Chapter 3 of this thesis, I use the model system *C. obscurior* to test the hypothesis that *dsx*, and more generally a broader range of regulatory sex differentiation genes, have been co-opted from the sex differentiation pathway into the female caste (queen/worker) and male morph (winged/ergatoid) differentiation pathways.

## 1.6 Homologous recombination accelerates evolution

Homologous recombination is a fundamental biological mechanism which is present across all three domains of life. During mitosis, homologous recombination is an important mechanism to repair double strand breaks. During meiosis, homologous recombination both stabilizes the pairing of homologous chromosomes, thereby ensuring proper segregation of homologous chromosomes, and generates offspring genetic diversity. However, meiotic recombination may also be disadvantageous by disrupting well-adapted parental genotypes and by leading to genetic homogenisation of populations via gene conversion events. Recombination rates have been shown to vary both across and within species. For example, mice show sex-specific recombination patterns (Dumont & Payseur, 2011). Even within genomes local recombination rates can show considerable levels of variation, exemplified by so-called recombination hot-spots in humans and mice that have been shown to be associated with rapidly evolving PRDM9 proteins (Baudat *et al.*, 2010).

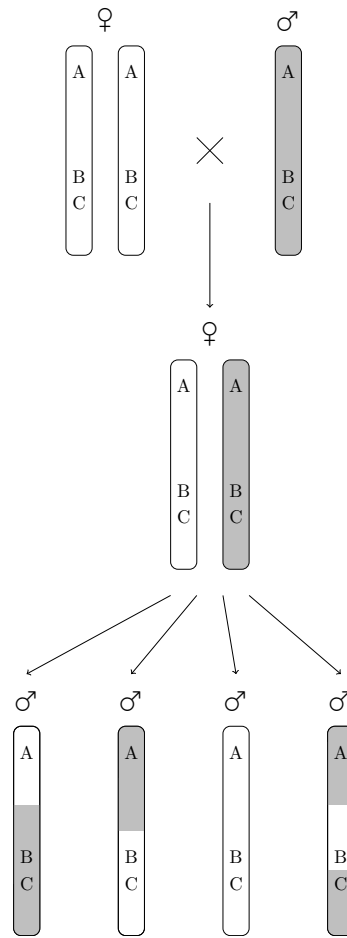
The significance of recombination as a driver of speciation lies in the acceleration of adaptation via Hill-Robertson effects (Hill & Robertson, 1966; Roze & Barton, 2006). By disrupting linkage between harmful and advantageous loci and by linking multiple beneficial loci, recombination increases the fixation probability of advantageous loci in a population compared to stepwise mutations (Hill & Robertson, 1966). Moreover, recombination not only occurs between homologous loci (= allelic recombination), but also between paralogous loci (= ectopic recombination). Ectopic recombination may lead to large chromosomal rearrangements such as inversions, in which recombination is suppressed (e.g. in maize and fire ants; Wang *et al.*, 2013a; Rodgers-Melnick *et al.*, 2015). Thus ectopic recombination can induce speciation (McGaugh & Noor, 2012). Especially inversions excluding the centromere will have drastic effects as they produce acentric and dicentric products, which may not produce viable gametes (McGaugh & Noor, 2012). Ectopic recombination often occurs between paralogous copies of transposable elements (TEs). Class I retrotransposons are predestined for ectopic recombination as they occur in several copies throughout the genome due to their transposition (copy/paste) activity. Ectopic recombination between TEs often results in structural rearrangements, which in turn prevent recombination and lead to reduced levels of recombination in TE-rich regions (Boissinot *et al.*, 2001; Bartolomé *et al.*, 2002; Rizzon *et al.*, 2002; Song & Boissinot, 2007). The *C. obscurior* genome has been found to be enriched in certain classes of TEs, which drives genomic novelty and thereby adaptation (Schrader *et al.*, 2014). The TEs are suggested to be organized in "TE-islands", i.e. regions with enriched TE abundance (Schrader *et al.*, 2014). The construction of a

genetic linkage map for *C. obscurior* which is described in Chapter 4 of this thesis will help to improve the genetic architecture of the *C. obscurior* genome assembly, address the integrity of the TE-islands and test for a relationship between recombination and TE content. The next section describes the process of genetic linkage mapping.

### 1.6.1 Genetic linkage mapping

The construction of genetic linkage maps is based on the tracing of recombination events during meiosis. Therefore, a cross between parental individuals with an appropriate level of genetic divergence is performed. The level of genetic divergence between parental populations is important because too low divergence will lead to too few detectable markers, whereas too high divergence will lead to non-correct pairing of homologous chromosomes during meiosis. Normally, intraspecific crosses are used for mapping, but in highly inbred species the use of sister species may be appropriate, e.g. an interspecific cross between *Nasonia vitripennis* and *Nasonia giraulti* was used to construct a linkage map for *Nasonia* (Niehuis *et al.*, 2010; Desjardins *et al.*, 2013). The degree of genetic diversity and the kind of genetic markers used will determine the number of loci that can be used for linkage mapping and will thus define the resolution of the genetic map. Originally, detectable phenotypes were used for mapping. The father of linkage mapping, Thomas Hunt Morgan, used eye color or wing polyphenisms to map the corresponding loci in *Drosophila* (Morgan *et al.*, 1915). Later microsatellites, RAPD markers or AFLPs were used for linkage mapping (e.g. Gadau *et al.*, 2001; Sirviö *et al.*, 2006), whereas nowadays next-generation-sequencing techniques (e.g. Restriction-site-associated DNA sequencing (RADseq) or microarray analysis, e.g. Miller *et al.*, 2007) are used to detect single nucleotide polymorphisms (SNPs) or insertion-deletion-polymorphisms (InDels) between the parental genomes.

In haplodiploid species (i.e. species where females are diploid and males are haploid), hybrid F2 males directly represent meiotic products of their mothers and thus are ideal for linkage mapping (Figure 1.3). Recombination fractions are calculated by dividing the number of recombinant individuals by the number of all individuals for each pair of markers. Theory predicts that the probability for cross-over events increases with increasing distance between two loci, thus recombination fraction can be translated into the genetic distance on the map in centiMorgan (cM), where 1 cM is defined as 1% recombination. To deal with high marker numbers, double-cross-over events and cross-over interference, mapping functions, e.g. the Kosambi or Haldane mapping function, are used to construct genetic maps.



**Figure 1.3: Linkage Mapping under Haplodiploidy.** In the example two F2 male show one cross-over event, one F2 male shows no cross-over event and one F2 male shows a double-cross-over event. Recombination fractions (rf) are calculated as the number of recombining individuals divided by the total number of individuals for each pair of markers. The lower rf between loci B and C ( $\text{rf}(B,C)=1/4$ ) suggests that loci B and C are probably closer than loci A and B ( $\text{rf}(AB)=3/4$ ).

### 1.6.2 Eusociality as driver of recombination?

Recombination rates are inferred from the comparison between the distance of two loci on a genetic linkage map in cM and the distance of the same loci on the physical genome assembly in basepairs, and are thus given in cM/Mb. Reported recombination rates are higher in eusocial Hymenoptera than in parasitic or solitary insects (Wilfert *et al.*, 2007). It has been suggested that high levels of recombination may have been important for the evolution of distinct castes and/or for the evolution and maintenance of a diverse range of worker phenotypes (Kent *et al.*, 2012). Alternatively, it has been suggested that high recombination rates may be important for the predicted diverse immune repertoire of eusocial insect colonies that live in high densities and

are expected to have a high pathogen load (Fischer & Schmid-Hempel, 2005). These ultimate explanations for high recombination rates under eusociality emphasize the importance of offspring genetic diversity. However in a simulation study Rueppell *et al.* (2012) showed that recombination does not contribute to offspring genetic diversity. Genetic diversity in eusocial insects can also be generated by independent assortment of parental chromosomes, multiple mating (polyandry) or the presence of multiple queens per nest (polygyny). Thus it remains elusive why recombination is higher in eusocial than non-social species and calls for further investigations. The limited availability of recombination rates complicates further clarification (Table 1.2).

**Table 1.2: Recombination rates of eusocial Hymenoptera.**



species	RR in cM/Mb	reference
<i>Apis mellifera</i>	19-37	Beye <i>et al.</i> (2006); Solignac <i>et al.</i> (2007); Liu <i>et al.</i> (2015b); Wallberg <i>et al.</i> (2015)
<i>Bombus impatiens</i>	4.76	Stolle <i>et al.</i> (2011)
<i>Vespula vulgaris</i>	9.7	Sirviö <i>et al.</i> (2011)
<i>Acromyrmex echinator</i>	6.2	Sirviö <i>et al.</i> (2006)
<i>Pogonomyrmex rugosus</i>	14.0	Sirviö <i>et al.</i> (2010)
<i>Cardiocondyla obscurior</i>	10.0	Chapter 4 of this thesis

Instead of linkage maps other measures have been used to infer recombination rates. For example Kent *et al.* (2012) used GC-content to infer recombination rates, as it was shown that recombination rate correlates positively with GC content due to GC-biased gene conversion (Duret & Arndt, 2008). However, this method may be misleading as it is unclear whether this correlation is valid for all species. Ross *et al.* (2015) showed that eusocial species exhibit higher rates of change in chromosome number than non-social species, which may be a consequence of elevated recombination rates, as these may lead to chromosomal rearrangements more often than in low-recombining species. However, Stolle *et al.* (2011) showed that chromosomal architecture across the Apidae is quite conserved despite high rates of recombination. The linkage map provided in Chapter 4 of this thesis allowed for a precise estimation of the genome wide recombination rate for *C. obscurior* and corroborates elevated recombination rates in eusocial species. Furthermore, estimation of local recombination rates and their correlation with genomic features such as GC- and TE-content provides information about the adaptive value and evolutionary consequences of recombination.

## 1.7 Ergatoid males as evolutionary novelties in the genus *Cardiocondyla*

An evolutionary novelty that arose within the Myrmicine ant genus *Cardiocondyla* are so-called "ergatoid" (literally "worker-like") males (Kinomura & Yamauchi, 1987). In contrast to the ancestral winged male morph, which exhibits a morphology typical for Hymenopteran males, ergatoid males exhibit an entire range of novel traits (see Table 1.3 for a comparison of winged and ergatoid males). Contrary to the mating biology typical for Hymenopterans where mating takes place during mating flights, ergatoid *Cardiocondyla* males stay inside their maternal nests and engage in lethal fights with other ergatoid males to monopolize mating with closely related virgin queens (Kinomura & Yamauchi, 1987).

**Table 1.3: Traits differing between winged and ergatoid *Cardiocondyla* males.**

	winged male	ergatoid male
		
wings	yes	no
mandibles	normal	enlarged
eyes	large	small
antennae	long	normal
ocelli	yes	no
lifelong spermatogenesis	no	yes
dispersal	yes	no
fighting behavior	no	yes
lifespan	short	long

Ergatoid males in the genus *Cardiocondyla* exhibit either sickle-shaped or shear-shaped mandibles according to their evolutionary history (Oettler *et al.*, 2010). While the sickle-shaped male types (Clade A) use their mandibles to size rivals and mark them with an unknown hindgut secretion which elicits worker aggression, males with shear-shaped mandibles (Clade B) are additionally able to actively crush the cuticle of rivals with their strong mandibles (Oettler *et al.*, 2010). Interestingly, some *Cardiocondyla* species of Clade B have re-evolved monogyny from ancestral polygyny, which can be associated with co-occurrence of multiple ergatoid males in one nest (Schrempf & Heinze, 2007).

Ergatoid males can be regarded as an evolutionary novelty themselves, as they represent a second fitness optimum besides the winged disperser morph according to Hallgrímsson *et al.* (2012). Male diphenism is associated with other evolutionary novelties in ergatoid males such as life-long spermatogenesis, enlarged mandibles or hindgut secretions (Heinze *et al.*, 1998). On the genomic level, the presence of ergatoid males results in extreme levels of inbreeding (Heinze *et al.*, 2006), to which *Cardiocondyla* is well-adapted (Schrempf *et al.*, 2015). A recent phylogeny of the Myrmicinae assigned a long branch to the genus *Cardiocondyla* whose phylogenetic placement is ambiguous (Ward *et al.*, 2014). This suggests that ergatoid males as an evolutionary novelty of *Cardiocondyla* may have dramatic effects on the genomic level.

## 1.8 The model organism *Cardiocondyla obscurior*

*Cardiocondyla obscurior* (Wheeler 1929), originally described as *C. wroughtonii* from Taiwan (Forel 1890) (Kugler, 1983) is a polygynous species whose ergatoid males have sickle-shaped mandibles. Colonies occur in trees in preformed cavities such as rolled leaves or hollow twigs. Its origin is presumably located in southeast Asia and it has been distributed throughout the tropics and subtropics via human trading activities (Heinze *et al.*, 2006). *C. obscurior* is perfectly suited as an ant model system due to small individual and colony size and short generation time. Unlike most ant species, controlled crosses can be performed year-round in the laboratory. It is possible to monitor the entire lifecycle of a queen in under one year and to record her lifespan and reproductive output, allowing the study of sexual selection or aging (Schrempf *et al.*, 2005, 2008). Moreover *C. obscurior* is highly suited for studies of development and phenotypic plasticity (Schrader *et al.*, 2015, Chapter 3 of this thesis) as both males and females display diphenic development, thus adding one level of comparison to most other eusocial insects.

High levels of inbreeding result in extreme levels of genetic homozygosity. This raises the question how novel genotypes as substrate for selection are generated. The analyses of the *C. obscurior* genome revealed that one mechanism for generating variation is the enrichment of the genome with transposable elements (TEs) (Schrader *et al.*, 2014). Regions enriched with transposons are sources for genetic divergence between two main study populations from Brazil and Japan, which differ in several traits on the phenotypic level, e.g. behavior (aggression), cuticular hydrocarbons or body size (Schrader *et al.*, 2014). Thus, the invasive species *C. obscurior* is highly suited for studying the causative factors of rapid adaptation to new habitats on a genomic level.



## 1.9 Aims of this thesis

This thesis investigates three mechanisms that generate evolutionary novelty in *Cardiocondyla obscurior*. To this end, genomic tools that were recently developed for *C. obscurior* were used and improved.

Sequencing and assembly of the *C. obscurior* hologenome (Schrader *et al.*, 2014) has led to the detection of the main endosymbionts of *C. obscurior*: *Wolbachia*, which is a prevalent symbiont across insects, and the newly described "*Candidatus Westeberhardia cardiocondylae*" (Chapter 2). The aim of this thesis was to provide a first characterization of the relationship between *C. obscurior* and *Westerberhardia* and to evaluate the evolutionary implications of endosymbiosis for *C. obscurior* biology.

Second, the recent availability of genomic resources for non-model organisms has given rise to a wide range of studies investigating a fundamental unresolved question in eusocial insect research: how distinct castes develop from the same genetic background. Eusociality can be regarded as a major evolutionary transition (Bourke, 2011), but has evolved repeatedly in the Hymenopteran lineage, suggesting that different mechanisms regulate caste determination across species. One aim of this thesis was to investigate the potential of co-option as a mechanism to generate distinct castes by recruiting a conserved set of genes, the sex differentiation genes (Chapter 3).

Finally, the resolution of the *C. obscurior* draft genome Cobs 1.4 is far from being perfect. Especially the repeat-rich regions caused problems in genomic assembly. Hence, I established a genetic linkage map to test and improve the resolution of the *C. obscurior* genome (Chapter 4). In addition, Chapter 4 addresses the hypothesis that high rates of homologous recombination compensate for high inbreeding levels. Lastly, the potential of recombination for generating novel genotypes, and thus also novel phenotypes, was evaluated.

## 2 A novel intracellular mutualistic bacterium in the invasive ant *Cardiocondyla obscurior*

Antonia Klein<sup>1,8</sup>, Lukas Schrader<sup>1,8</sup>, Rosario Gil<sup>2</sup>, Alejandro Manzano-Marín<sup>2</sup>, Laura Flórez<sup>3</sup>, David Wheeler<sup>5</sup>, John H Werren<sup>6</sup>, Amparo Latorre<sup>2,7</sup>, Jürgen Heinze<sup>1</sup>, Martin Kaltenpoth<sup>3,4</sup>, Andrés Moya<sup>2</sup> and Jan Oettler<sup>1</sup>

<sup>1</sup> Institut für Zoologie, Universität Regensburg, Regensburg, Germany

<sup>2</sup> Institut Canvanilles de Biodiversitat i Biologia Evolutiva (ICBiBE), Parc Científic de la Universitat de Valencia, Paterna (Valencia), Spain

<sup>3</sup> Max Planck Institute for Chemical Ecology, Jena, Germany

<sup>4</sup> Johannes Gutenberg University Mainz, Institute for Zoology, Department for Evolutionary Ecology, Mainz, Germany

<sup>5</sup> Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

<sup>6</sup> Department of Biology, University Rochester, Rochester, NY, USA

<sup>7</sup> Área de Genómica y Salud de la Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (FISABIO)-Salud Pública, Valencia, Spain

<sup>8</sup> These authors contributed equally to this work.

## 2.1 Abstract

The evolution of eukaryotic organisms is often strongly influenced by microbial symbionts that confer novel traits to their hosts. Here we describe the intracellular Enterobacteriaceae symbiont of the invasive ant *Cardiocondyla obscurior*, ‘*Candidatus Westeberhardia cardiocondylae*’. Upon metamorphosis *Westerberhardia* is found in gut-associated bacteriomes that deteriorate following eclosion. Only queens maintain *Westerberhardia* in the ovarian nurse cells from where the symbionts are transmitted to late-stage oocytes during nurse cell depletion. Functional analyses of the streamlined genome of *Westerberhardia* (533 kb, 23.41 % GC-content) indicate that neither vitamins nor essential amino acids are provided for the host. However, the genome encodes for an almost complete shikimate pathway leading to 4-hydroxyphenylpyruvate, which could be converted into tyrosine by the host. Together with increasing titers of *Westerberhardia* during pupal stage, this suggests a contribution of *Westerberhardia* to cuticle formation. Despite a widespread occurrence of *Westerberhardia* across host populations, one ant lineage was found to be naturally symbiont-free, pointing to the loss of an otherwise prevalent endosymbiont. This study yields insights into a novel intracellular mutualist that could play a role in the invasive success of *C. obscurior*.

## 2.2 Introduction

Interactions between organisms drive biological complexity (Maynard Smith & Szathmáry, 1997), shaping life as we know it. Symbioses with prokaryotes are considered to promote Eukaryote diversification (Brucker & Bordenstein, 2012), particularly in insects (Moya *et al.*, 2008; Gil *et al.*, 2010). Some bacterial symbionts provide novel ecological traits to their insect hosts, for example, defense against pathogens or parasitoids (Oliver *et al.*, 2003; Kaltenpoth *et al.*, 2005), enhanced stress tolerance (Russell & Moran, 2006) or nutrients (Douglas, 2009). Nutrient-providing symbionts are commonly found in hosts with restricted diets, for example, aphids feeding on phloem sap (Baumann, 2005), blood-feeding diptera (Wang *et al.*, 2013a) or grain weevils (Heddi *et al.*, 1999). Symbionts can provide essential amino acids, vitamins or help in nitrogen recycling (Nakabachi *et al.*, 2005; Feldhaar *et al.*, 2007; Michalkova *et al.*, 2014; Patino-Navarrete *et al.*, 2014). Such bacteria are commonly harbored in bacteriocytes, specialized host cells that sometimes form special organ-like structures, the bacteriomes (Baumann, 2005), or are confined to the insect gut (Engel & Moran, 2014). Provision-

ing with nutrients can lead to increased fitness (Michalkova *et al.*, 2014), which may enable invasive species to exploit novel habitats or food sources (Feldhaar, 2011).

*Cardiocondyla obscurior* (Wheeler, 1929) is an invasive ant that forms small multi-queen colonies in disturbed, arboreal habitats throughout the tropics. A peculiarity of the genus *Cardiocondyla* is the occurrence of wingless males that mate with closely related queens in their maternal nest (Oettler *et al.*, 2010). New colonies are established via colony splitting (Heinze *et al.*, 2006). This unique life history with frequent genetic bottlenecks and high levels of inbreeding makes it an interesting model for the study of rapid adaptation to novel environments (Schrader *et al.*, 2014).

Here we describe a so far unknown intracellular symbiont of *C. obscurior*, for which we propose the name '*Candidatus Westeberhardia cardiocondylae*' strain *obscurior* (from here on referred to as *Westerberhardia*). We analyzed its distribution within and across host populations, and compared infection of individual ants depending on morph and age. Furthermore, we localized *Westerberhardia* in the host and scrutinized its genome focusing on its metabolic functions. *Westerberhardia* has lost many metabolic capabilities, but retained most of the shikimate pathway and the ability to synthesize the tyrosine precursor 4-hydroxyphenylpyruvate. We suggest that its localization in gut-associated bacteriomes of pupae, and the increased titers during pupal development point to a role of *Westerberhardia* in cuticle formation.

## 2.3 Material and Methods

### Ant colonies

We reared *C. obscurior* colonies from Brazil (BR), Japan (JP) and Spain (SP) in the laboratory. The BR colonies originated from two collection sites ~70 km apart, a cacao plantation in Ilhéus (2009 and 2013) and a citrus plantation near Una (2012) (Brazilian Ministry of Science and Technology, permits 20324-1/40101-1). JP colonies were collected from two coral trees 100 m apart (lineages "OypB", "OypC") in the Oonoyama park in Naha, Okinawa (2011) and from additional trees of the same park ("OypU", 2013). SP colonies were collected at a campsite in Los Realejos, Tenerife in 2012 and 2013. All colonies were housed in plaster nests under 12 h 28°C light / 12 h 23°C dark cycles, with constant humidity and *ad libitum* provided honey and pieces of cockroaches. All animal treatment guidelines applicable to ants under international and German law have been followed.

## ***Westeberhardia* detection and phylogenomic analyses**

During analyses of the *C. obscurior* genome (Schrader *et al.*, 2014), we identified prokaryotic scaffolds and candidates for horizontal gene transfers (HGTs) (Wheeler *et al.*, 2013). These were then further characterized by blasting (blastx) all annotated genes against a database of prokaryotic proteins. Besides *Wolbachia*, we identified six scaffolds of an unknown Enterobacteriaceae. Following *de novo* genome assembly and annotation (see below), we used translated CDS sequences for phylogenetic placement following Husník *et al.* (2011). Briefly, we performed Dayhoff6 recoding followed by a phylogenomic reconstruction with PhyloBayes v3.3f (Lartillot *et al.*, 2009), based on 64 single-copy protein clusters (Supplementary Information).

We detected one prokaryotic gene incorporated into the host genome. After manual correction of the HGT gene model, we used blastx analyses against NCBI's non-redundant database to identify homologs. RNAseq data were used to verify expression in seven larval and seven adult queens by mapping reads against the *C. obscurior* genome (Schrader *et al.*, 2014). We generated count tables with htseq (Anders *et al.*, 2015) against *C. obscurior* gene annotations (including the manually corrected gene) and calculated untransformed, size factor-normalized read counts.

## **Genome assembly, annotation and functional analyses**

Paired-end Illumina reads from Schrader *et al.* (2014) were used for *de novo* assembly of the *Westeberhardia* genome. We removed *Wolbachia* sequences based on their blastx result, and then assembled remaining prokaryotic reads using SOAPdenovo2 (Luo *et al.*, 2012). The resulting contigs were scaffolded using a custom-modified version of SSPACE v2.0 (Boetzer *et al.*, 2011). Raw reads were then mapped back to the contigs using MIRA 4.0.1 (Chevreux *et al.*, 1999) and manually joined. Scaffold corroboration and visual inspection of contigs were performed in the Staden Package (Staden *et al.*, 2000). Inconsistencies were broken and manually reassembled. Base-calling correction was done using POLISHER (Lapidus *et al.*, 2008). No corrections were made to the consensus, which consisted of a single 532 684-bp contig (average coverage 204.5x).

The replication origin was predicted using originX (Worning *et al.*, 2006). A first round of open reading frame (ORF) prediction was performed using Prodigal v.2.5 (Hyatt *et al.*, 2010) and the predicted ORFs were annotated using the BASys server (Van Domselaar *et al.*, 2005). tRNAs were predicted using the "COVE-only" algorithm of tRNAscan-SE v.1.3.1 (Lowe & Eddy, 1997), and checked with TFAM v.1.4 (Tåquist *et al.*, 2007). tmRNAs and their tag-peptides were predicted using

ARAGORN v.1.2.36 (Laslett & Canback, 2004). The genome was searched against Rfam v11 (Burge *et al.*, 2013) using Infernal v.1.1 (Nawrocki & Eddy, 2013), and the resulting ncRNAs were manually integrated into the annotation following the INSDC's conventions ([http://www.insdc.org/files/feature\\_table.html](http://www.insdc.org/files/feature_table.html)). Ribosome-binding sites were predicted using RBSfinder (Suzek *et al.*, 2001) and signal peptides were predicted using SignalP v.4.1 (Petersen *et al.*, 2011). The resulting annotation was manually curated in Artemis (Rutherford *et al.*, 2000).

Metabolic functions were automatically predicted and analyzed using Pathway Tools v.17.5 (Karp *et al.*, 2010) against BioCyc and MetaCyc databases (Caspi *et al.*, 2012), following manual curation. Functional information was retrieved from the EcoCyc (Keseler *et al.*, 2013), KEEG (Ogata *et al.*, 1999) and BRENDA (Scheer *et al.*, 2011) databases.

## Coverage analyses

For comparison of *Westeberhardia* infection between the sequenced reference colonies from BR (Ilhéus, 2009) and JP (OypB) (Schrader *et al.*, 2014), we mapped genomic reads from pools of 30 BR and 26 JP males (140 million reads each) against the *C. obscurior* and *Westeberhardia* genomes, and compared coverage between BR and JP with samtools' depth algorithm (Li, 2011) and custom bash/perl/R scripts as described in Schrader *et al.* (2014).

## Analyses of intraspecific infection dynamics by PCR and real-time quantitative PCR (qPCR)

To assess *Westeberhardia* presence across host populations, we screened 42 *C. obscurior* samples collected worldwide and the sister species *Cardiocondyla wroughtonii* (Forel, 1890) by performing a diagnostic PCR assay on a 204-bp fragment of the *nrdB* (ribonucleoside diphosphate reductase 1 subunit beta) gene of *Westeberhardia* (*WEOB\_403*) (*nrdB*-for: 5'-GGAAGGAGTCCTAATGTTGCG-3', *nrdB*-rev: 5'-ACCAGAAATATCTTTTGCACGTT-3'), using the ant housekeeping gene *elongation factor 1-alpha 1* (*Cobs\_01649*) (*EF1*-for: 5'-TCACTGGTACCTCGCAAGCCGA-3', *EF1*-rev: 5'-AGCGTGCTCACGAGTTTGTCCG-3', 104-bp fragment) as a control. We used DNA from a previous study (Oettler *et al.*, 2010), samples from laboratory colonies and stored tissue from which we extracted DNA using a chloroform-based method (Sambrook & Russell, 2001) (Table 2.1). To verify infection with the same *Westeberhardia* species, we

sequenced a 917-bp fragment of the 16S rRNA gene of *Westeberhardia* (*WEOB\_122*) (*WE16S*-for: 5'-CATTTGAATATGTAGAATGGACC-3', *WE16S*-rev: 5'-AACTTT-TACAAGATCGCTTCTC-3') from one individual each of the BR, JP and SP populations and of *C. wroughtonii* (see Supplementary Information for PCR details).

We assessed inter- and intrapopulational *Westeberhardia* prevalence in laboratory colonies using PCR and quantitative (qPCR) assays for workers and queens, respectively (Supplementary Information). For this purpose, we sampled 6-10 dealate queens and 9-10 workers from 6-8 colonies from three lineages from BR and JP, respectively, and from the SP population, resulting in a total of 538 analyzed workers and 517 queens. Worker and queen DNA was extracted using the hotshot method (Alasaad *et al.*, 2008) and the NucleoSpin®Tissue XS Kit (Macherey-Nagel, Düren, Germany), respectively. To control for age effects on infection (see below) we selected freshly eclosed workers when possible.

We quantified *Westeberhardia* of single individuals by determining normalized *nrdB* copy numbers with qPCR (Supplementary Information) across developmental stages (larvae and prepupae of unknown sex and caste, young and old female pupae) for JP (OypB) and BR (Una, 2012) and across morphs (queens, workers, winged males, wingless males), ages (queens = 2, 14, 28 and 48 days; workers = 2, 14 and 28 days), and fertilization state of queens (virgin, mated) for only the BR population (Una, 2012).

## Fluorescence *in situ* hybridization

To localize *Westeberhardia*, we performed fluorescence *in situ* hybridization (FISH) as described previously (Kaltenpoth *et al.*, 2012, 2014, and Supplementary Information) on abdominal sections of queen, worker and wingless male pupae from BR (Ilhéus, 2009) and adult queens from BR (Ilhéus, 2009) and JP (OypB) with the general eubacterial probe *EUB338* (5'-GCTGCCTCCCGTAGGAGT-3') (Amann *et al.*, 1990) and one of the *Westeberhardia*-specific probes *Wcard1* (5'-ATCAGTTTTCGAACGCCATTC-3') and *Wcard2* (5'-CGGAAGCCACAATTCAAGAT-3'), targeting the 16S rRNA gene. Probes were labeled with Cy3 or Cy5, and samples were counterstained with DAPI (4',6-diamidino-2-phenylindole).

## Test for reproductive manipulation and paternal inheritance

Several bacterial symbionts are known to be reproductive manipulators, with cytoplasmic incompatibility (CI) and parthenogenesis induction (PI) being the most common phenotypes (Cordaux *et al.*, 2011). While CI results in the incompatibility of crosses

between infected males and uninfected queens, PI causes parthenogenetic production of diploid female offspring in infected females. We crossed uninfected JP queens (OypB) with infected BR males (Ilhéus, 2009), by placing sexual pupae together with 20 workers into new nests ( $n=10$ ), which were observed twice a week for the presence of male and female brood.

To test for paternal inheritance of *Westeberhardia*, crosses of infected males and uninfected freshly eclosed virgin queens were initiated in a mating arena overnight. The following day, we dissected and macerated the spermathecae of queens ( $n=8$ ) in dH<sub>2</sub>O and isolated DNA using the NucleoSpin<sup>®</sup>Tissue XS Kit (Machery-Nagel). We performed a diagnostic PCR assay with the *nrdB* gene and the housekeeping gene *EF1* as positive control. We further analyzed infection status of two worker pupae each emerging from four of the above crosses using the *nrdB* PCR assay.

## 2.4 Results

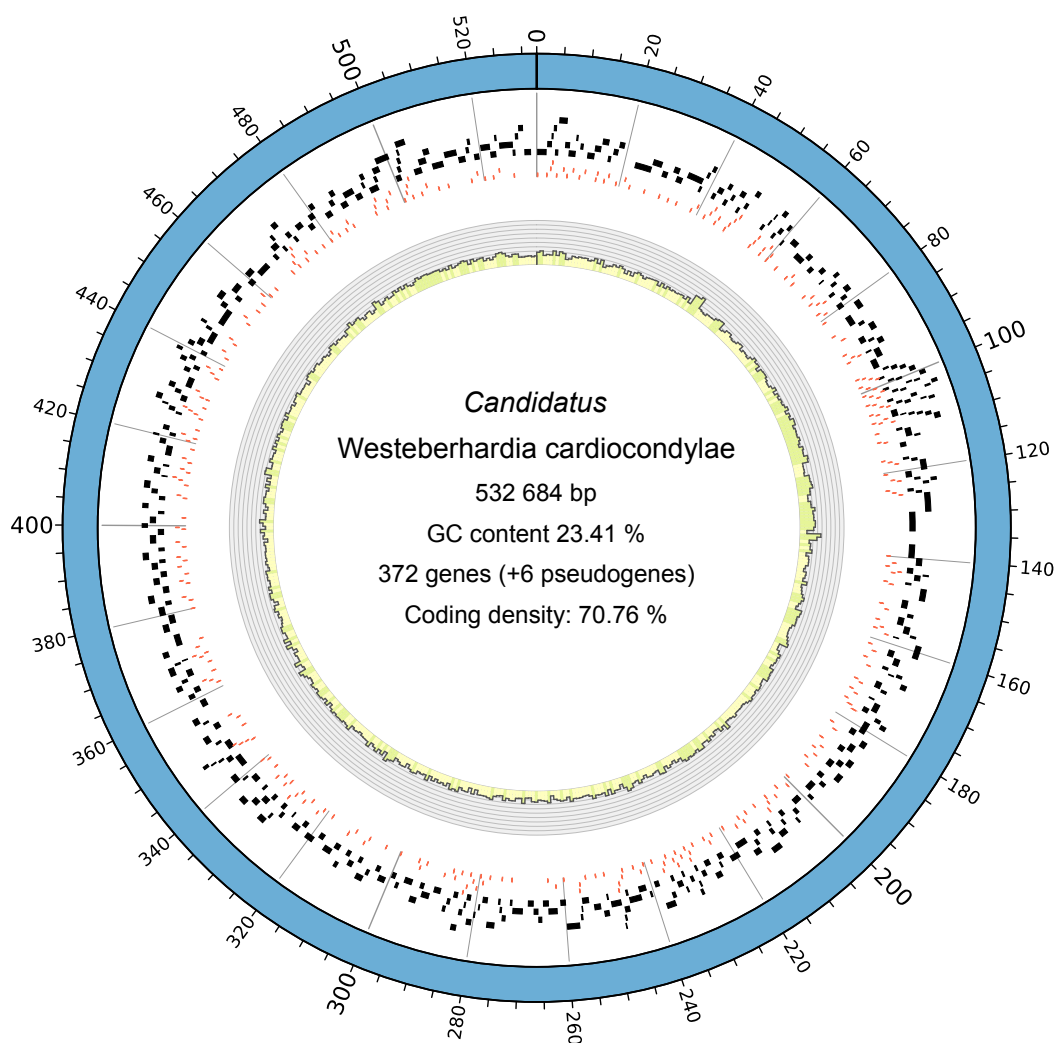
### Microbial associates of *C. obscurior*

Blastx analyses of the *C. obscurior* hologenome (Schrader *et al.*, 2014) retrieved 1.5 Mb of *Wolbachia* sequence and 543 172-bp in six scaffolds of an unknown  $\gamma$ -proteobacterium. A preliminary assembly of the *Wolbachia* sequences is hosted on antgenomes.org. The 16S sequence of the  $\gamma$ -proteobacterium showed 98.4% identity with an Enterobacteriaceae of a *C. obscurior* sample from Florida, USA (voucher RA0330, Genbank: GQ275143), detected during a survey of ant-associated bacteria (Russell *et al.*, 2009b). Blastx analyses further revealed a 360-bp intronless gene of putative prokaryotic origin encoding a xanthine-guanine phosphoribosyltransferase (EC: 2.4.2.22), which is incorporated into the host genome and has its closest ortholog in *Enterobacter cloacae* (WP\_023478997). Xanthine-guanine phosphoribosyltransferase has a central role in the synthesis of purine nucleotides through the salvage pathways, converting xanthine and guanine to XMP and GMP, respectively. The gene is present in genomic reads of *C. obscurior* from BR (Ilhéus, 2009) and JP (OypB). We used published RNAseq data from adult queens and queen-destined larvae (Schrader *et al.*, 2014), to confirm *in vivo* transcription of the HGT and found a fivefold increased expression in larvae compared to adults (median<sub>larvae</sub>=1140.1, median<sub>queens</sub>=223.2, Mann-Whitney *U*-test:  $W=79$ ,  $p<0.001$ ).



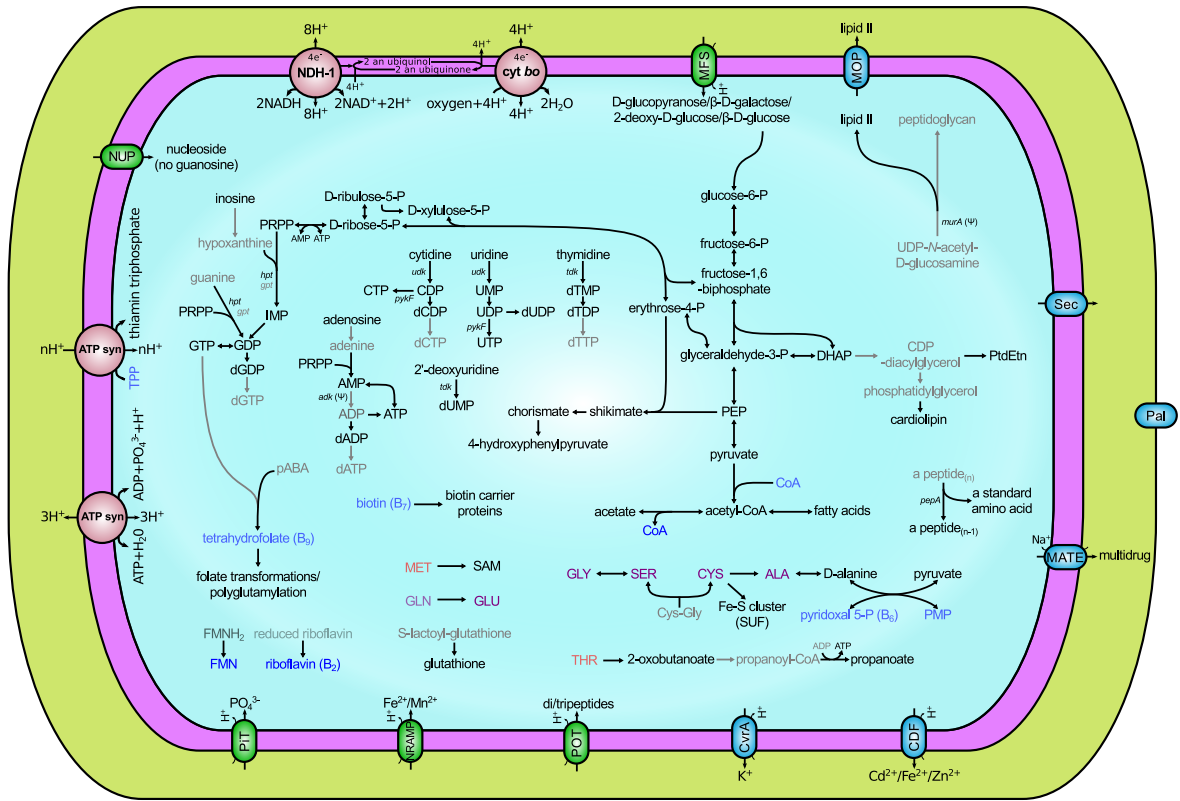
## ***Westeberhardia* genome assembly, annotation and functional and phylogenomic analyses**

*De novo* assembly of the *Westeberhardia* genome produced a single scaffold representing a circular chromosome of 532 684-bp (23.41 % GC-content) with 372 protein-coding genes and six pseudogenes Figure 2.1. Sequences are accessible through the European Nucleotide Archive (<http://www.ebi.ac.uk/ena/data/view/LN774881>) under study number PRJEB8217, chromosome accession number LN774881 and are hosted on [antgenomes.org](http://antgenomes.org).



**Figure 2.1: Genomic structure of *Westeberhardia*.** Representational circos plot (Krzywinski *et al.*, 2009) illustrating genomic properties of *Westeberhardia*. Tile plots show the distribution of protein-coding genes (black bars) and ribosomal binding sites (red bars). The histogram in the inner circle shows GC-content in percent for 1 kb windows.

The phylogenomic analysis placed *Westeberhardia* within a group of Enterobacteriaceae symbionts comprising both facultative and obligate insect endosymbionts, including *Sodalis*, *Baumannia*, *Blochmannia* and *Wigglesworthia* (Supplementary Figure S2.1). The genome codes for a simplified but functional informational machinery, with complete setups for DNA replication, transcription, translation and protein folding, but few genes involved in DNA repair (Figure 2.2). *Westeberhardia* has a limited metabolic repertoire but is capable of glycolysis, pentose phosphate pathway, fatty acid biosynthesis, nucleotide synthesis and ATP production through oxidative phosphorylation. The pathway for glycerophospholipid biosynthesis is impaired. Metabolite transport appears to be based on electrochemical-potential driven transporters and inorganic phosphate transporters, while ATP-binding cassette transporter as well as phosphotransferase system transporter genes are missing. *Westeberhardia* has lost the pathways for synthesis of all essential and most non-essential amino acids and cofactors, but has retained an incomplete shikimate pathway. Thus, it is able to synthesize chorismate, a central metabolite in the biosynthesis of many aromatic compounds (e.g. phenylalanine, tryptophan, *p*-hydroxybenzoate or enterobactin). However, it can only use chorismate for the biosynthesis of 4-hydroxyphenylpyruvate, a precursor of tyrosine (Hopkins & Kramer, 1992; Andersen, 2012). Although *Westeberhardia* cannot complete the last step in this pathway, the host genome codes for tyrosine aminotransferase (EC 2.6.1.5) converting 4-hydroxyphenylpyruvate to tyrosine (*Cobs\_01567*). Further conversion of tyrosine to DOPA (3,4-dihydroxyphenylalanine), an important component of insect cuticles (Andersen, 2012), might occur through tyrosine 3-monooxygenase (EC 1.14.16.2) encoded in the host genome (*Cobs\_14710*).



**Figure 2.2: *Westerberhardia* metabolic reconstruction.** Intact pathways are shown in black lines, unclear pathways (missing a specific gene or having it pseudogenized) are shown in gray lines. Exporters are represented using green ovals, whereas exporters/importers are represented using blue ovals with the name of the family/superfamily they belong to, otherwise the protein or complex name is used. ATP synthase is shown twice to represent an additional metabolic capability. Essential and non-essential amino acids are shown in red and purple lettering, respectively. Cofactors and vitamins are represented in blue. Blurred compounds represent those for which biosynthesis or import cannot be accounted for based on the genomic data, according to MetaCyc. Relevant genes involved in the biosynthesis of nucleotides and peptidoglycan are indicated. A single frameshift is found in *adk* and *murA*, therefore they might be young pseudogenes, or an RNA polymerase or ribosomal slippage would be required to produce a functional protein.

## Intraspecific infection dynamics

Coverage analysis of genomic reads showed that, in contrast to males from a BR lineage (Ilhéus, 2009), males from a JP lineage (OypB) are devoid of *Westerberhardia* (coverage BR: 30x, JP: 0.21, Figure 2.3A). qPCR of the *Westerberhardia*-specific *nrdB* gene in female pupae as well as larvae and prepupae of unknown sex and caste verified that *Westerberhardia* is completely absent in OypB (Figure 2.3B). Accordingly, *Westerberhardia* was not detected by FISH in sections of adult OypB queens.

Analyses of *C. obscurior* samples collected worldwide showed that *Westerberhardia* is present in 34 of 42 tested samples (81.0%), including all samples from BR, but ab-

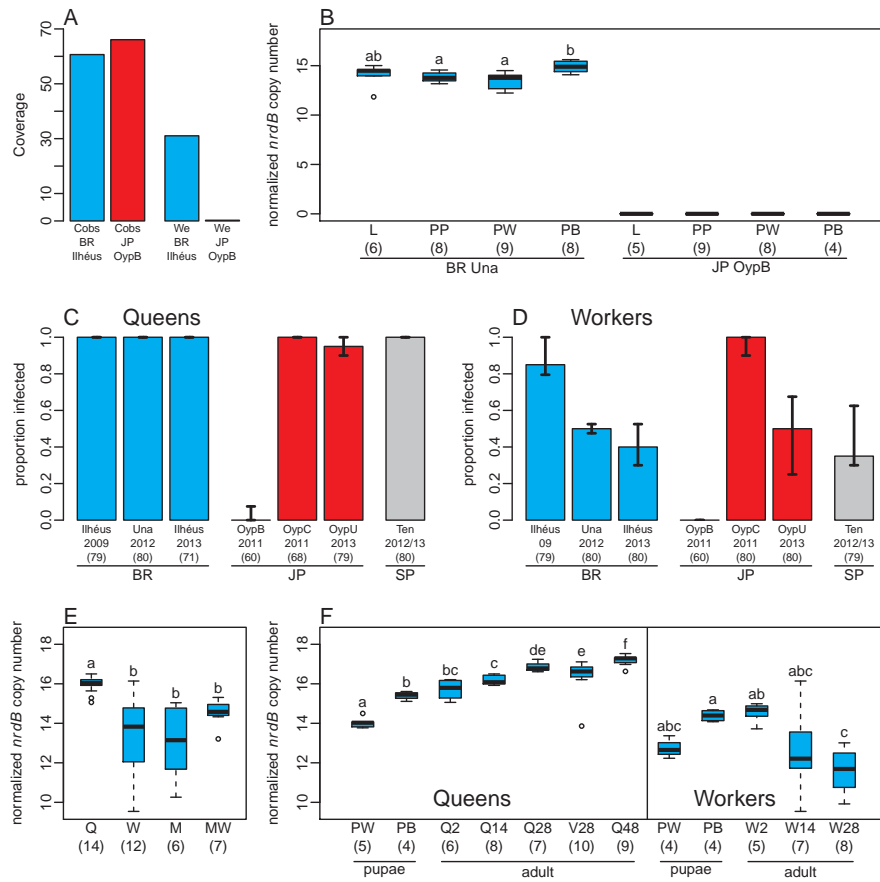
sent in some JP populations and in material from Egypt and Sri Lanka (Table 2.1). The closely related species *C. wroughtonii* also contains *Westeberhardia*. A 917-bp 16S rDNA fragment of *Westeberhardia* is identical between the three *C. obscurior* populations from BR, JP and SP, and between *C. obscurior* and *C. wroughtonii*.

*Westeberhardia* infection of workers varies considerably within and among infected lineages, ranging from 42.5 % to 96.3 % (Figure 2.3D), whereas queen infection is almost fixed in all populations (88.6 % to 100 %, Figure 2.3C, Supplementary Table S2.5). However, in the OypB lineage, only one of 60 workers and two of 60 queens were infected at low levels (indicated by weak bands on the agarose gel or late  $C_q$  values in the qPCR, respectively). These values are not significantly different from zero (one-sample  $t$ -tests: workers:  $t(59)=1$ ,  $p=0.32$ ; queens:  $t(59)=1.43$ ,  $p=0.16$ ), and could possibly be caused by contamination. Intriguingly, individuals from colonies collected in a tree merely 100 m away (OypC) show infection rates of 96.3 % (workers) and 100 % (queens).

In the BR (Una, 2012) population *Westeberhardia* relative densities increase significantly during pupal development (worker and queen pupae combined) from white (early) to brown (late) pupae ( $N_{\text{pupa white}}=9$ ,  $N_{\text{pupa brown}}=8$ ;  $t$ -test:  $t(14.7)=-4.3$ ,  $p<0.001$ ), but is not different between larvae, prepupae and early pupae (Figure 2.3B, Supplementary Table S2.1). *Westeberhardia* titers are higher in queen compared to worker pupae ( $t$ -test:  $t(13.3)=2.6$ ,  $p=0.023$ ). A comparison of 2- to 14-day-old adults of each morph (queens, workers, winged and wingless males) shows that *Westeberhardia* titers differ significantly across castes (Kruskal-Wallis:  $X^2=24.2$ ,  $df=3$ ,  $p<0.001$ , Figure 2.3E), with queens having significantly more *Westeberhardia* than the other morphs which are not different from each other (pairwise Mann-Whitney  $U$ -tests with Benjamini-Hochberg correction for multiple testing, Benjamini & Hochberg, 1995, Supplementary Table S2.2). We calculated generalized linear models (GLMs) with a Gaussian distribution and identity link function to model age dependency of *Westeberhardia* in adult females (Figure 2.3F). In workers, infection decreases with age (GLM:  $df=18$ ,  $F=12.7$ ,  $p=0.002$ ; Supplementary Table S2.3) and is significantly more variable than in queens (Fligner-Killeen-Test:  $X^2=17.0$ ,  $df=1$ ,  $p<0.001$ ). In queens, *Westeberhardia* significantly increase with age from day 2 after eclosion to day 48 (GLM:  $df=38$ ,  $F=29.4$ ,  $p<0.001$ ; Supplementary Table S2.4). Virgin and mated 28-day-old queens show no significant difference in infection (Wilcoxon's rank-sum test:  $W=49$ ,  $p>0.05$ ).

**Table 2.1:** Prevalence of *Westeberhardia* across populations of *Cardiocondyla obscurior* and the closely related species *C. wroughtonii* based on a diagnostic PCR screen using the *nrdB* gene. Abbreviations: EtOH, ethyl alcohol; Q, queen; N, no; W, worker; Y, yes

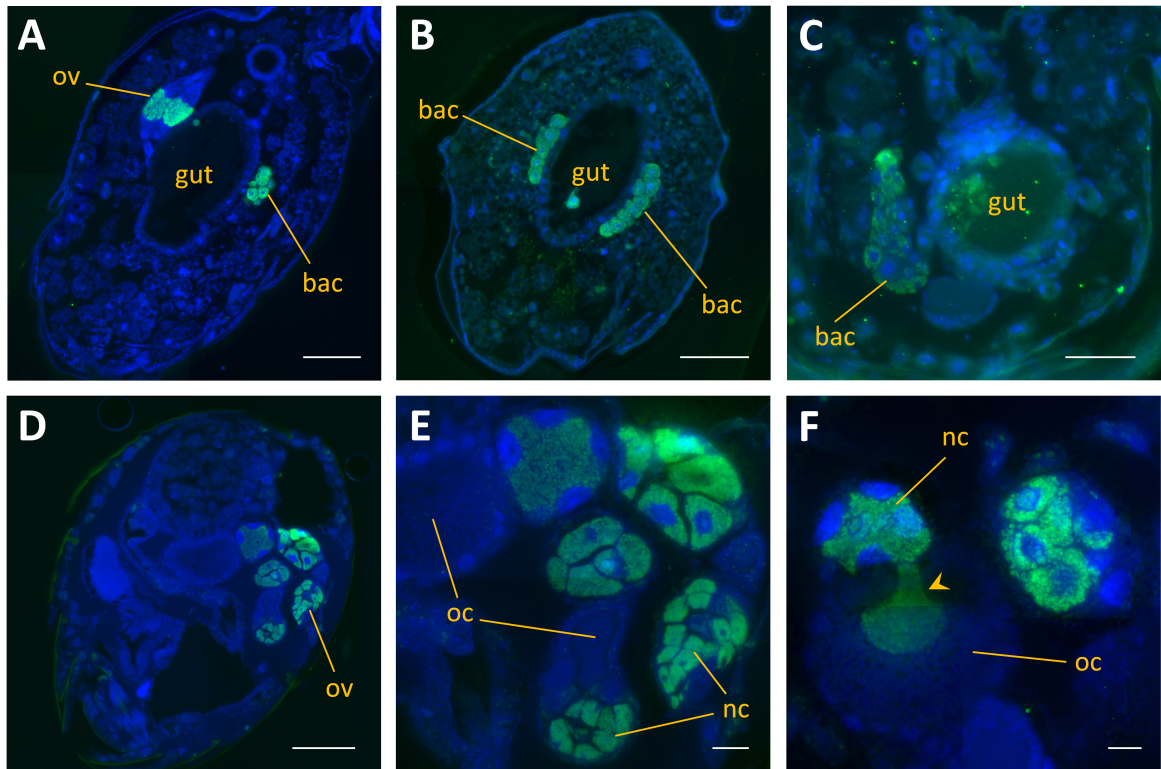
Sampling site (year)	Sample description	Morph	<i>Westeberhardia</i> (sample size)
<i>C. obscurior</i>			
Brazil: Ilhéus (2004)	Laboratory colonies	W	Y (3)
Brazil: Ilhéus (2009)	Laboratory colonies	Q	Y (4)
Brazil: Una (2012)	Laboratory colonies	Q	Y (4)
Brazil: Ilhéus (2013)	Laboratory colonies	Q	Y (4)
Japan: Ishigaki (2002)	Oettler <i>et al.</i> (2010)	W	N (2)
Japan: Naha (2011) 'OypB'	Laboratory colonies	Q	N (4)
Japan: Naha (2011) 'OypC'	Laboratory colonies	Q	Y (4)
Japan: Naha (2013) 'OypU'	Laboratory colonies	Q	Y (4)
Tenerife (2012)	Laboratory colonies	Q	Y (4)
Egypt: Talkha (2003)	Oettler <i>et al.</i> (2010)	W	N (1)
Fiji (2007)	EtOH material	W	Y (1)
Malaysia: Ulu Gombak (2002)	Oettler <i>et al.</i> (2010)	W	Y (1)
Sri Lanka (2006)	Oettler <i>et al.</i> (2010)	W	N (1)
USA: Lake Alfred, Florida (2004)	Oettler <i>et al.</i> (2010)	W	Y (3)
<i>C. cf. obscurior</i> Singapore (2014)	EtOH material	W	Y (2)
<i>C. wroughtonii</i>			
Japan: Nakijin (2013)	Laboratory colonies	W	Y (2)



**Figure 2.3: Intraspecific and temporal dynamics of *Westerberhardia* infection** (A) In genomic coverage data for pooled haploid males mapped against the *Westerberhardia* reference, *Westerberhardia* reads (We) were exclusively present in the Brazil, Ilhéus (2009) sample (BR, blue) and no reads could be detected in the OypB, Japan (JP, red) sample, while coverage of *C. obscurior* reads (Cobs) mapped against the *C. obscurior* reference is similar. (B) Real time-quantitative PCRs on DNA level for the *nrdB* gene confirm the absence of *Westerberhardia* in larvae (L), prepupae (PP) and female (queen and worker) pupae (PW = pupa white, PB = pupa brown) of the JP OypB population, whereas all those developmental stages are infected in the BR Una (2012) population (letters indicate significances for within-population comparisons for BR). (C+D) Prevalence of *Westerberhardia* in queens (C) and workers (D) across different populations of *C. obscurior* from BR (blue), JP (red) and Tenerife, Spain (SP, gray), as revealed by qPCR (C) and diagnostic PCR (D), of the *nrdB* gene. For each lineage 6-8 colonies and per colony 9-10 young workers and 6-10 queens were tested. Bars represent medians and whiskers denote quartiles. Note that while *Westerberhardia* infection status of workers varies between and within populations of *C. obscurior*, it is almost fixed in queens across all lineages except OypB. (E+F) Morph (E) and age (F) dependency of relative amounts of *Westerberhardia* in *C. obscurior* individuals from BR (Una, 2012) determined by real time-quantitative PCR. Normalized *nrdB* copy numbers are elevated in queens compared to all other morphs (Q=queens, W=workers, M=males winged, MW=males wingless) (E), increase with age in queens, but decrease with age in workers (numbers after Q/W show age in days, V=virgin queens, letters indicate significant differences for within-caste comparisons) (F). Sample sizes are given within parentheses.

## Localization of *Westeberhardia*

*Westeberhardia* is localized intracellularly in bacteriomes connected to the gut in queen, worker and wingless male pupae (Figure 2.4A-C). As in many other insect taxa with bacteriomes, bacteriocytes are densely packed with symbiont cells and exhibit enlarged host cell nuclei. In some sections, individual symbiont cells or symbiont-filled bacteriocytes were found in the gut lumen, suggesting beginning degradation of the bacteriome in later pupal stages. Concordantly, no bacteriomes were detected in adult queens (Figure 2.4D). However, both pupal and adult queens show high *Westeberhardia* abundances in the ovaries (Figure 2.4A,D-F). In particular, *Westeberhardia* is localized predominantly in the nurse cells. Several sections captured transgenerational infection events of the symbiont from the maternal nurse cells into late stages of the developing oocyte (Figure 2.4F).



**Figure 2.4: Localization of *Westeberhardia* in adults and pupae of *C. obscurior* (from Brazil, Ilhéus, 2009).** Symbionts were stained in longitudinal sections through the abdomen with the *Westeberhardia*-specific probe Wcard2-Cy5 (green), and host cell nuclei were counterstained with DAPI (blue). (A-C) Localization of *Westeberhardia* in gut-associated bacteriomes (bac) in pupae of a queen (A), a worker (B), and a male (C). Note the additional presence of symbionts in the queen ovaries (ov). (D) Section of the abdomen of an adult queen, with symbionts visible in the ovaries (ov). (E,F) Ovaries of an adult queen. Symbionts are mainly localized in the nurse cells (nc), but enter the developing oocyte (oc), probably during nurse cell depletion (arrowhead). Scale bars: 100  $\mu\text{m}$  (A,B,D), 50  $\mu\text{m}$  (C), and 20  $\mu\text{m}$  (E,F).

## Test for reproductive manipulation and paternal inheritance

*Westeberhardia*-free OypB queens mated to *Westeberhardia*-infected BR males produced viable diploid female and haploid male offspring. Furthermore, infected females generally also produce male offspring and unfertilized females exclusively produce males. In addition, *Westeberhardia* does not seem to be transmitted paternally, as revealed by the absence of the diagnostic *nrdB* gene in the spermatheca content of *Westeberhardia*-uninfected queens mated with infected males. Worker pupae emerging from the above crosses were also uninfected.

## Description of ‘*Candidatus Westeberhardia cardiocondylae*’

In accordance with the guidelines of the International Committee of Systematic Bacteriology, unculturable bacteria should be classified as *Candidatus* (Murray & Stackebrandt, 1995). We propose the name ‘*Candidatus Westeberhardia cardiocondylae*’ strain *obscurior* for this newly discovered  $\gamma$ 3-proteobacterium. The genus name *Westeberhardia* refers to Mary Jane West-Eberhard, expressing our admiration for her far-reaching advances in evolutionary developmental biology. The specific epithet, *cardiocondylae*, indicates that it is an endosymbiont of *Cardiocondyla* ants.

## 2.5 Discussion

The 16S sequence of an unknown Enterobacteriaceae isolated from *C. obscurior* was previously published (Russell *et al.*, 2009b), but the specificity and functionality of this association had not been addressed. Here we describe it as ‘*Candidatus Westeberhardia cardiocondylae*’ strain *obscurior* and provide a first characterization of its relationship with *C. obscurior*. Phylogenomic analysis indicates that *Westeberhardia* is closely related to *Blochmannia*, the obligate endosymbiont of *Camponotus* ants (Feldhaar *et al.*, 2007). Nevertheless, its phylogenetic placement has to be considered with caution, because of long-branch attraction. As already observed by Husník *et al.* (2011), the monophyly of the cluster formed by *Sodalis*, *Baumannia*, *Blochmannia* and *Wigglesworthia*, in which *Westeberhardia* appears, needs to be further tested.



## Transmission of *Westeberhardia*

Maternal transmission of *Westeberhardia* occurs through a different process than described for other endosymbionts (Koga *et al.*, 2012; Balmand *et al.*, 2013). In adult queens *Westeberhardia* is localized in ovarian syncytial nurse cells, which originate from the same germline stem cell as the oocyte and are responsible for provisioning of the oocyte with metabolites. During the process of nurse cell depletion, when large amounts of cytoplasmic material are channeled into the oocyte (Mahajan-Miklos & Cooley, 1994), cytoplasmic *Westeberhardia* are swept into the developing oocyte, ensuring complete vertical transmission (Figure 2.4F).

CI is a widespread phenotype induced by some bacterial endosymbionts (Gotoh *et al.*, 2006; Werren *et al.*, 2008), but has to our knowledge not been shown for any social insect. In social Hymenoptera, males develop from haploid, unfertilized eggs and females from diploid, fertilized eggs, consequently CI would only affect diploid offspring. *Westeberhardia* does not appear to induce strong CI, if any, as uninfected queens mated to infected males produced diploid F1 females. Another common phenotype caused by reproductive manipulators, the induction of thelytokous parthenogenesis (PI), can also be excluded. This would lead to exclusive female offspring in infected queens, as *C. obscurior* does not show diploid male production (Schrempf *et al.*, 2006). The regular occurrence of male offspring and the exclusive production of males by unfertilized females argue against PI. However, *Wolbachia*, also present in *C. obscurior*, could exert effects on host reproduction as well. In contrast to some other intracellular symbionts (Moran & Dunbar, 2006; Damiani *et al.*, 2008; Watanabe *et al.*, 2014), paternal transmission of *Westeberhardia* is unlikely as we did not detect *Westeberhardia* DNA in transferred sperm and/or seminal fluids stored in the spermatheca of uninfected queens mated to infected males.

## *Westeberhardia* as a possible source of a HGT event

The bacterial gene *gpt* encoding the xanthine-guanine phosphoribosyltransferase (XPRT, EC: 2.4.2.22), which was horizontally transferred into the host genome has its closest ortholog in *Enterobacter*, an Enterobacteriaceae. This indicates that *Westeberhardia* could be the origin of the HGT event. However, it could also be a relict of a former symbiont no longer present in *C. obscurior* (Husník *et al.*, 2013). Homologs of the *gpt* have been identified in most bacterial endosymbionts, including *Buchnera*, *Moranella*, *Blochmannia*, *Sodalis* and *Wigglesworthia*. The presence of the HGT in the OypB

lineage suggests either an ancestral association between *C. obscurior* and *Westeberhardia* and a secondary loss of the symbiont in OypB, or the origin of the HGT from an unknown bacterium in the ancestor of both lineages. *Westeberhardia* is not capable of *de novo* synthesis of purines, but it is capable of producing all purine nucleotides from recovered bases and nucleosides. A functional genome annotation revealed the presence of *hpt*, a gene with similar function to *gpt*, in the *Westeberhardia* genome. However, RNAseq data show that infected hosts transcribe *gpt*. Therefore, *Westeberhardia* might be reliant on an effective salvage utilizing the host-encoded *gpt* in some conditions (O'Reilly *et al.*, 1984). Interestingly, *gpt* expression is higher in larval compared to adult queens, indicating that it is not correlated with *Westeberhardia* titers. Inhibiting *gpt* expression in *Westeberhardia*-infected and -uninfected individuals will help elucidate a putative effect of this gene on host and bacteria fitness.

## Dependency of *Westeberhardia* on host-provided metabolites

With a genome size reduction to 533 kb and a GC-content of 23.4 %, the *Westeberhardia* genome exhibits features of degenerative genome evolution following the transition to obligate symbiosis (Moya *et al.*, 2008). In addition to reduced effective population sizes in host-associated bacteria compared to free-living relatives, small effective population size of *C. obscurior* (Schrader *et al.*, 2014) and social insects in general (Romiguier *et al.*, 2014) could lead to even faster genome degeneration. With a coding density of 70.76 %, the genome is surprisingly loosely packed, compared to other endosymbionts with similar-length genomes (88 % coding density on average, McCutcheon & Moran, 2012). Furthermore, the occurrence of six pseudogenes indicates that genome erosion in *Westeberhardia* is still incomplete. It was previously shown that even in advanced mutualistic relationships, endosymbiont genome reduction continues (Gil *et al.*, 2002). Nevertheless, despite the substantial genome reduction, *Westeberhardia* appears capable of DNA replication, transcription, translation and protein folding, suggesting that it is close to a minimal cell status (Gil *et al.*, 2004). On the other hand, the lack of *dnaA* and any other alternative mechanism for replication initiation might be an indication that bacterial replication could fall under host control, as suggested for *Wigglesworthia* and *Blochmannia* (Akman *et al.*, 2002; Gil *et al.*, 2003), although no potential mechanism has been identified yet. It has also been suggested that, in the absence of *recA*, the maintenance of *recBCD* might indicate that this complex has a role in DNA replication initiation instead of recombination (Wu *et al.*, 2006). The gene count with only 372 coding genes and the impairment of essential pathways like cofactor and essential amino acids biosynthesis indicate a metabolic dependency on ex-

trinsic resources. In particular, a highly simplified cell envelope and the absence of most transporter genes point towards dependency on the host machinery. In this, *Westeberhardia* resembles *Buchnera aphidicola* BCc, which also lacks the ability to synthesize cell surface components (Pérez-Brocal *et al.*, 2006). Intracellularity allows the host to control endosymbiont populations (Vigneron *et al.*, 2012), which together with the lack of *dnaA* suggests that *Westeberhardia* populations are controlled by *C. obscurior*.

## Potential for mutualism: Shikimate pathway

*Westeberhardia* has retained almost the complete shikimate pathway, which produces chorismate, the precursor for tryptophan, tyrosine and phenylalanine, but lacks the downstream enzymes necessary for synthesis of these aromatic amino acids. However, it can produce 4-hydroxyphenylpyruvate, which can then be converted to tyrosine by the host. Tyrosine is a precursor for DOPA and thereby essential for cuticle formation in insects (Hopkins & Kramer, 1992; Andersen, 2012). Insects cannot synthesize aromatic amino acids and acquisition from diet and/or provisioning by endosymbionts is a common phenomenon. For example, *B. aphidicola* has evolved overproduction of phenylalanine and tryptophan (Lai *et al.*, 1994; Jiménez *et al.*, 2000). Similarly, *B. floridanus* can synthesize tyrosine, and increased tyrosine biosynthesis during the host's pupal stage (Zientz *et al.*, 2006) coincides with elevated *Blochmannia* titers (Stoll *et al.*, 2010; Ratzka *et al.*, 2013). Accordingly, we found high densities of *Westeberhardia* in late *C. obscurior* pupae and young adults. Together with the detection of gut-associated bacteriomes in pupae, this suggests a role of *Westeberhardia* in cuticle synthesis during metamorphosis. Although the precise metabolites provided to the host are unclear at this stage, we hypothesize that the retention of the shikimate pathway is key to the mutualistic relationship between *Westeberhardia* and *Cardiocondyla*.

After eclosion, *Westeberhardia* declines slowly in workers but proliferates in queens with age. Although virgin queens exhibit significantly reduced egg laying rates compared to mated queens (Schrempf *et al.*, 2005), we did not find an increase of *Westeberhardia* infection with reproductive output. Instead, it appears as if the mere availability of reproductive tissue allows proliferation of *Westeberhardia*. As a consequence of reproductive division of labor in an eusocial host, the major proportion of *Westeberhardia* encounter a dead end. *Cardiocondyla* workers completely lack ovaries (Heinze *et al.*, 2006), thus likely impeding *Westeberhardia* proliferation in the absence of the appropriate microhabitat. In *Camponotus floridanus* ants, mid-gut connected bacteriomes populated by *Blochmannia* during the pupal stage become symbiont-free in adult queens and workers, whereas queens retain *Blochmannia* in their ovarioles (Sauer *et al.*, 2000;

Wolschin *et al.*, 2004). Similarly, symbionts localized in gut-associated bacteriomes of cereal weevils are actively eliminated by initiation of apoptosis after cuticle formation is finished, but ovary-associated symbionts are retained for vertical transmission (Vigneron *et al.*, 2014). Probably due to slow degeneration of bacteriomes, *Westeberhardia* was still present in young adult males and workers. Bacteria detected in the gut lumen (from degrading bacteriomes) may be the source of continued bacterial infections found in adult workers and males. We have not ruled out that the symbiont continues to play a role in adult workers, although its general decline suggests that the role is not vital. It remains to be investigated if active degradation of bacterial populations in workers benefits individuals and, on a higher level, colony performance (Wenseleers *et al.*, 2002).

## Population differences cast doubts about the symbiosis status

We found a naturally occurring uninfected host lineage that continues to thrive in the laboratory, questioning the essentiality of *Westeberhardia*, at least under conditions including *ad libitum* protein provisioning. We verified absence of *Westeberhardia* in freshly collected field colonies and established laboratory colonies with different methods and across different developmental stages. It remains elusive why *Westeberhardia* prevalence is so substantially different between colonies of two lineages (OypB, OypC) separated by such short distances (<100 m) in the field. This could indicate a facultative status of *Westeberhardia* as seen for many insect-microbe symbioses (e.g. Moran *et al.*, 2005). However, *Westeberhardia* lacks the main characteristics shared by facultative symbionts, even those of a facultative symbiont in the transition to becoming obligate (i.e., large genomes with low coding density and abundance of pseudogenes, presence of repetitive sequences and transposable elements, high GC-content, Manzano-Marín & Latorre, 2014). The occurrence of *Westeberhardia* in *C. wroughtonii* and in different populations of *C. obscurior* indicates an ancestral infection and secondary loss of *Westeberhardia* in OypB. Since the impact of facultative symbionts depends on the particular environmental conditions (Dale & Welburn, 2001; Hansen & Moran, 2014), a shift in diet or different gut microbiota could explain symbiont loss by natural selection. Due to proximity of host lineages, an alternative, more parsimonious explanation is mutational loss of *Westeberhardia* and subsequent fixation through drift (Reuter *et al.*, 2005). Future comparisons between infected and uninfected hosts with the same or different genetic backgrounds under varying environmental conditions will help to reveal potential effects of *Westeberhardia* on host fitness.

## Conclusion

Our study describes a novel intracellular bacterium that maintains a mutualistic relationship with its host but can be lost in some conditions. Its genomic organisation, metabolic capabilities, localization and prevalence during host development indicate a role of *Westeberhardia* in host cuticle formation, possibly facilitating an invasive lifestyle in nutrient poor arboreal environments. The putative monophyly with other insect endosymbionts possibly facilitating cuticle buildup during development (Zientz *et al.*, 2006; Vigneron *et al.*, 2014), suggests a single origin of metamorphosis-based symbiosis.

Owing to novel traits emerging through host-symbiont associations, it is indispensable to evaluate possible fitness effects of symbionts on hosts, which are used as model organisms for broad biological questions. Although symbionts are situated along the boundary between biotic environmental factors and genomic composition of the host, it becomes obvious that selection pressures acting on the holobiont must be considered when studying adaptation: “*Contrary to common belief, environmentally initiated novelties may have greater evolutionary potential than mutationally induced ones.*” (West-Eberhard, 2005).

## 2.6 Acknowledgements

We are most grateful to Masaki Suefuji, Tina Wanke, and Jacques HC Delabie for collecting the JP, SP and BR colonies. We like to thank Helena Lowack for help in the laboratory, the "Molecular Ecology" class of 2014 (University Regensburg) and Franziska Scherl for help with the population screens. We thank Benjamin Weiss for help with the sectioning of ant specimens for FISH. We thank Yannick Wurm, for hosting Wobs and WEobs on antgenomes.org. We also like to thank Horacio Frydman, Abdelaziz Heddi, Filip Husník, Eva Schultner and Mary Jane West-Eberhard for helpful discussions. This study was funded by a grant from the German Science Foundation (DFG He1623/31) to JO and JH. AL received grant BFU2012-39816-C02-01 from the Spanish Ministry of Economy and Competitiveness co-financed by FEDER funds, and AM was funded by grant PrometeoII/2014/065 from Valencian Community (Spain). AMM was supported by a Marie Curie grant (FP7-PEOPLE-2010\_ITNSYMBIOMICS 264774) and a fellowship from the Mexican Federal Government (CONACYT 327211/381508). MK and LF were funded by the Max Planck Society, and an additionally grant from the German Science Foundation to MK (DFG KA2846/2-1). DW and JW were sup-

ported by US NSF DEB0821936; JW was further supported by US NSF DEB1257053, and DW was further funded by a startup grant from Massey University.

# **3 Co-option of sex differentiation pathways facilitates the evolution of social insect polyphenism**

Antonia Klein<sup>1</sup>, Eva Schultner<sup>2</sup>, Helena Lowak<sup>1</sup>, Lukas Schrader<sup>1</sup>, Jürgen Heinze<sup>1</sup>,  
and Jan Oettler<sup>1</sup>

<sup>1</sup> Institut für Zoologie, Universität Regensburg, Regensburg, Germany

<sup>2</sup> Centre of Excellence in Biological Interactions, Department of Biosciences, University of Helsinki, Viikinkaari 1, 00014, Helsinki, Finland

### 3.1 Abstract

The origin of major evolutionary transitions is a principal yet largely unresolved question in biology. One such major transition is the evolution of eusociality, with morphologically, physiologically and behaviorally distinct female castes developing from the same genetic background. It is currently unknown whether this is achieved by genomic idiosyncrasies. Alternatively, we propose a concept of co-option of conserved gene sets, where changes in the regulation of highly conserved pathways link new input stimuli with novel target traits, thereby providing a powerful source for evolutionary novelty. We validate this concept by identifying candidate sex differentiation genes that have been co-opted in the developmental processes of female and male morph differentiation in *Cardiocondyla obscurior*, an ant with four discrete phenotypes. Among these candidates is the master sex differentiation gene *doublesex* (*dsx*), a highly conserved gene whose sex-specific isoforms act as transcription factors in downstream developmental processes. Differential expression of *dsx* isoforms across morphs in pupae and adults points towards a critical role of *dsx* in modulating polyphenic development. During larval development, *dsx* as well as seven other conserved candidate genes, exhibit both sex- and morph-specific expression patterns. By generating new interaction networks between alternative primary signals (genetic or environmental) and the same set of effector genes, phenotypic differentiation on two levels, into females or males and discrete morphs, is achieved. We argue that co-option of core gene sets with ancestral function may have facilitated the evolution of eusociality because these pathways were already deeply rooted in development and exerted a switch-like behavior.

### 3.2 Introduction

A major question in biology is how evolutionary novelty originates. Evolution requires variation, which can be generated by mutations and neo-functionalization of duplicated genes, or acquired by gene transfer and through recombination of existing genotypes or developmental recombination (West-Eberhard, 2003). However, these mechanisms provide only small-scale steps. In contrast, co-option of conserved gene sets, where changes in the regulation of highly conserved pathways link new input stimuli with novel target traits, constitutes a rapid source for evolutionary novelty because even minor alterations in regulatory mechanisms can produce substantially different phenotypes (“theory of facilitated variation”, Gerhart & Kirschner, 2007).



A prime example for conserved regulatory gene sets that produce substantial phenotypic divergence are sex differentiation genes (Kopp, 2012). Intriguingly, in social Hymenoptera, phenotypic divergence between queens and workers or among worker castes can be as extreme as divergence between sexes (Wheeler, 1986), with castes differing in morphology (e.g. body size, presence of wings), physiology (e.g. rate of senescence, fertility) and behavior (e.g. foraging, mating). Proximally, genetic components (Schwander & Keller, 2008; Smith *et al.*, 2008), hormones (Wheeler, 1990), nutrition (Kamakura, 2011), methylation patterns (Simola *et al.*, 2013) and different transcription factor binding site landscapes (Simola *et al.*, 2013; Schrader *et al.*, 2015) translate into morph-biased gene expression during larval development. Despite extensive efforts however, we still lack a general developmental evolutionary framework to explain the repeated independent evolution of polyphenism in the social Hymenoptera.

Here, we propose that co-option of conserved sex differentiation gene sets underlies the evolution of morphological castes in the ant *Cardiocondyla obscurior*. The key regulatory pathways for sex differentiation involving genes of the *doublesex-mab3* (DM) family are highly conserved across the metazoa (Yi & Zarkower, 1999; Kopp, 2012; Matson & Zarkower, 2012). In contrast, there is rapid turnover in both primary signals for sex determination and sexually dimorphic phenotypic traits. In insects, the evolution of sex determination mechanisms is assumed to follow an hourglass model (Bopp *et al.*, 2014), with variable genetic primary signals (i.e. sex-determining signals), a conserved part including the genes *transformer* (*tra*) and *doublesex* (*dsx*) and a variable set of downstream genes regulating sex-specific tissue differentiation. *dsx* is assumed to be the most conserved player in this cascade (Bopp *et al.*, 2014), with its sex-specific isoforms acting as transcription factors that induce tissue sex-specificity (Burtis & Baker, 1989) by regulating expression of target genes involved in basal functions like cell-signaling and transcriptional regulation (Chatterjee *et al.*, 2011; Shukla & Palli, 2012a; Clough *et al.*, 2014). Expression of sex differentiation genes in turn modulates sex-specific phenotypic traits including brain development (Tomchaney *et al.*, 2014), genital disc formation (Chatterjee *et al.*, 2011), reproductive physiology (Shukla & Palli, 2012a) and development of male sexual ornaments (Kijimoto *et al.*, 2012).

Like most ants, *C. obscurior* produces two distinct female castes (queens=QU, workers=WO). In addition, *C. obscurior* expresses two alternative male phenotypes (winged=WM, wingless=EM from ‘ergatoid’, literally ‘worker-like’), adding an additional level of phenotypic complexity compared to most social Hymenoptera (Oettler *et al.*, 2010; Schrader *et al.*, 2015). For simplicity, we refer to all four phenotypes as “morphs” in the following. Like in most social Hymenoptera, sex in this ant is determined by the

fertilization status of the egg, with unfertilized haploid eggs developing into males and fertilized diploid eggs developing into females. Like in other insects, this primary genetic signal induces sex-specific gene expression in the developing egg. In turn, female and male morph differentiation is induced by environmental conditions experienced during development and is fixed in 3rd (final) instar larvae (Schrempf *et al.*, 2006; Schrader *et al.*, 2015).

To test the hypothesis that conserved regulatory sex differentiation gene sets are involved in this coordinated developmental switch into discrete morphs in *C. obscurior*, we identify *dsx* and its female- and male-specific splicing variants, and corroborate the link between phenotype and isoform expression in sex mosaics that aberrantly develop male and female morphologies in one individual. Importantly, we find that *dsx* sex-specific isoforms are expressed not only sex-specifically, but also morph-specifically across different life stages. Finally, morph-specific expression patterns in a wide range of sex-biased genes in RNAseq data of individual larvae of known developmental trajectory (Schrader *et al.*, 2015) provide strong evidence for extensive co-option of existing sex differentiation pathways for female caste and male morph differentiation.

## 3.3 Results

### Verification of haplodiploidy

Like most haplodiploid Hymenoptera, male *C. obscurior* develop from unfertilized eggs while females develop from fertilized eggs. Although the mode of sex determination in *C. obscurior* is unknown, single locus sex determination (like in the honeybee, Gempe *et al.*, 2000) is unlikely (Schrempf *et al.*, 2006). Using inter-population crosses we verified haplodiploidy and excluded unusual modes of reproduction found in other ants (Leniaud *et al.*, 2012). Hybrid queens and workers were heterozygous for diagnostic microsatellite markers, whereas emerging winged and wingless males as well as one sex mosaic individual exclusively carried the maternal alleles (Table S3.5).

### Identification of *C. obscurior dsx*

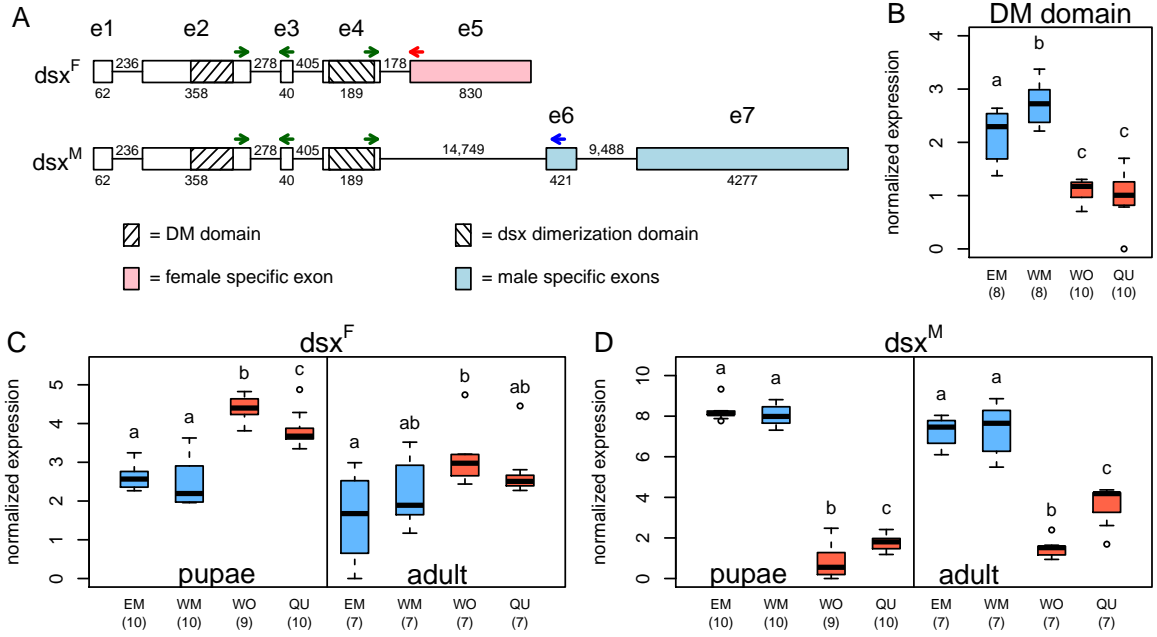
The *C. obscurior* genome (Schrader *et al.*, 2014) has four paralogs containing the DM domain of *dsx* (pfam00751; *Cobs\_01393*, *Cobs\_07724*, *Cobs\_09254* and *Cobs\_18158*) representing the ancestral state in holometabolous insects (Oliveira *et al.*, 2009). Sex-specific splice forms are only known from one gene per species, e.g., in *Apis* (Cho *et*

*al.*, 2007) and *Nasonia* (Oliveira *et al.*, 2009), and the function of the other paralogs is unclear. To find the functional *dsx* ortholog in *C. obscurior*, we identified DM domain-containing proteins of *Drosophila melanogaster*, *Nasonia vitripennis*, *Apis mellifera*, *Pogonomyrmex barbatus*, *Acromyrmex echinator* and *C. obscurior* by BLASTp and tBLASTn analyses (Table S3.1) and aligned them with MUSCLE (Edgar, 2004). We extracted the DM domain region from the manually corrected alignment (Figure S3.1) and built a phylogenetic tree in MEGA (Hall, 2013), applying a WAG+G+I phylogenetic model and bootstrap resampling with 1,000 replicates (Figure S3.3). This revealed that *Cobs\_01393* shows the highest sequence homology to functional *dsx* in other insects. We further verified functionality of *Cobs\_01393* by identifying *C. obscurior prospero* by tBLASTn of *N. vitripennis prospero* (Genbank NP\_001164363) against predicted *C. obscurior* transcripts. Microsynteny of *dsx* and *prospero* is conserved across the Hymenoptera (Oliveira *et al.*, 2009). Ortholog *C. obscurior prospero* sequence is located within ~79 kb of *Cobs\_01393*, while the other paralogs are located on distinct scaffolds in the genome assembly. Finally, we analyzed differential expression of the four DM-containing transcripts in RNAseq data from female and male larvae from (Schrader *et al.*, 2015) with DESeq2. Only *Cobs\_01393* was differentially expressed between the sexes (Table S3.3), confirming that this transcript represents the functionally active *dsx* in *C. obscurior*.

## Sex- and morph-specific expression of *dsx* isoforms

We applied 3' rapid amplification of cDNA ends (RACE) to identify the full-length sequence and sex-specific isoforms of *dsx* (Figure S3.2). The first four exons are identical in both isoforms. The DM domain (pfam00751) is located in exon 2 and the *dsx* dimerization domain (pfam08828) in exon 4. The female transcript *dsx<sup>F</sup>* contains one female-specific exon, whereas the male transcript *dsx<sup>M</sup>* is generated by alternative splicing, skipping the female exon but including two male-specific exons (Figure S3.1A, Table S3.4). This splicing pattern with a shortened female transcript has been inferred for the fire ant *Solenopsis invicta* from RNAseq data (Nipitwattanaphon *et al.*, 2014), and matches *dsx* sex-specific isoforms in *D. melanogaster* and *A. mellifera*, but not in *N. vitripennis* (Oliveira *et al.*, 2009). While the sex-signaling function of *dsx* is highly conserved across species, recent evidence shows that *dsx* evolves rapidly even on small evolutionary timescales (Loehlin *et al.*, 2010; Kunte, 2014; Eirín-López & Sánchez, 2015), explaining variation in *dsx* sequence and splicing patterns between species. A higher level of divergence in *dsx* compared to other DM domain-containing proteins in our phylogenetic analysis confirms this result. For subsequent real-time quanti-

tative PCR (qPCR) we designed intron-spanning primers targeting the DM domain-containing exon (i.e., measuring overall expression of both isoforms) as well as specific primers for both isoforms (Figure 3.1A). We found significantly higher expression of the DM domain in adult males compared to females ( $n_{EM}=8$ ,  $n_{WM}=8$ ,  $n_{WO}=10$ ,  $n_{QU}=10$ ; Welch two sample t-test:  $t(25.7)=-8.7$ ,  $p<0.001$ , Figure 3.1B). While expression of the DM-domain was similar in queens and workers (pairwise t-test with Benjamini-Hochberg correction:  $p=0.672$ ), winged males showed higher expression than wingless males (pairwise t-test with Benjamini-Hochberg correction:  $p=0.009$ ). We then compared expression of  $dsx^F$  and  $dsx^M$  across all four morphs in pupae ( $n_{EM}=10$ ,  $n_{WM}=10$ ,  $n_{WO}=9$ ,  $n_{QU}=10$ ) and adults ( $n_{EM}=7$ ,  $n_{WM}=7$ ,  $n_{WO}=7$ ,  $n_{QU}=7$ ; Figure 3.1C,D). We found morph-specific signatures of expression in both life stages for  $dsx^F$  (ANOVA: pupae:  $F(3,35)=42.33$ ,  $p<0.001$ ; adults:  $F(3,24)=3.75$ ,  $p=0.0243$ ) as well as for  $dsx^M$  (Kruskal-Wallis rank sum test: pupae:  $X^2=30.2$ ,  $df=3$ ,  $p<0.001$ ; adults:  $X^2=22.6$ ,  $df=3$ ,  $p<0.001$ ). Worker pupae showed significantly higher  $dsx^F$  expression than queen pupae (pairwise t-test with Benjamini-Hochberg correction:  $p=0.013$ ) and worker pupae and adults showed significantly lower  $dsx^M$  expression than queen pupae and adults, respectively (pairwise Wilcoxon Tests with Benjamini-Hochberg correction: pupae:  $p=0.012$ ; adults:  $p=0.0014$ ). In males, neither  $dsx^F$  nor  $dsx^M$  expression differed significantly between the two male morphs. However, overall expression of both isoforms showed higher  $dsx$  expression in winged compared to wingless males (Figure 3.1B). Combined, differential expression of  $dsx$  isoforms across morphs in pupae and adults points towards a critical role of  $dsx$  in modulating polyphenic development.

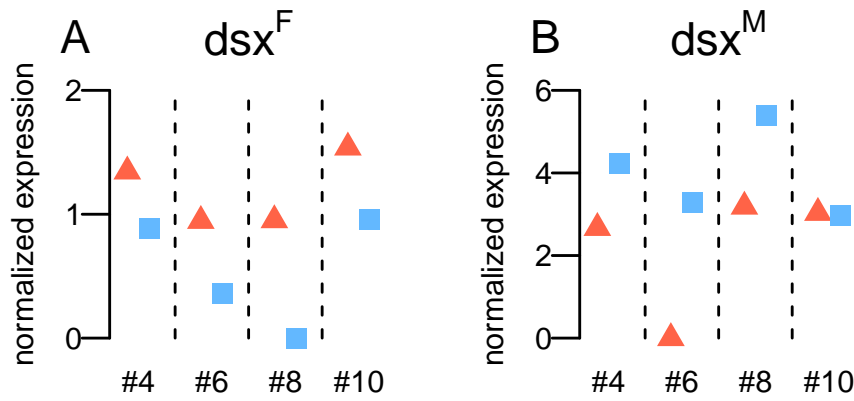


**Figure 3.1: Sex- and morph-specific expression of *doublesex* isoforms in the ant *Cardiocondyla obscurior*.** (A) Schematic illustration of the sex-specific *dsx* isoforms *dsx<sup>F</sup>* and *dsx<sup>M</sup>*. Numbers show sizes of exons and introns in bp. Arrows indicate approximate positions of primers (green: primers targeting both isoforms, red/blue: specific for *dsx<sup>F</sup>*/*dsx<sup>M</sup>*, respectively; primers: green-exon2: *dsx4\_for4*, green-exon3: *dsx4\_rev1*, green-exon4: *4for*, red: *F5rev*, blue: *M5rev*). (B-D) Normalized gene expression of *dsx* measured by qPCR in wingless (ergatoid) males (EM), winged males (WM), workers (WO) and queens (QU). (B) Expression of the DM domain-containing exon in adults is higher in males than in females, and higher in winged than in ergatoid males. (C) *dsx<sup>F</sup>* is significantly higher expressed in female than in male pupae, whereas *dsx<sup>M</sup>* is significantly higher expressed in males than in females (D). Worker pupae express significantly more *dsx<sup>F</sup>* than queen pupae and significantly less *dsx<sup>M</sup>* than queen pupae, and adult workers express significantly less *dsx<sup>M</sup>* than adult queens (C+D). Letters indicate significant differences. Sample sizes are given within parentheses, boxplots show the median, interquartile ranges (IQR) and 1.5 IQR.

## Verification of sex-specificity of *dsx* isoforms in sex mosaics

To confirm that expression of *dsx* isoforms corresponds with phenotypic tissue differentiation we used qPCR to analyze *dsx<sup>M</sup>* and *dsx<sup>F</sup>* expression in aberrantly produced sex mosaic individuals that express both male and female characters. We dissected heads and thoraces from four sex mosaic pupae (for morphological descriptions see Table S3.6), and separated female and male tissue in RNeasy<sup>®</sup>-Lysis Buffer (Qiagen). *dsx<sup>F</sup>* and *dsx<sup>M</sup>* splice forms were detected in both tissue types but expression levels were consistent with morphology in all individuals, except for one sample that had similar levels of *dsx<sup>M</sup>* in both tissue types (Figure 3.2). In most documented cases, *C. obscurior* sex mosaics were laterally separated into female and male halves, indicating that inter-

sexuality is caused by single and early developmental aberrations such as anomalous fertilization events, loss of sex locus expression or inheritance of maternal effects (Yang & Abouheif, 2011; Kamping *et al.*, 2007; Michez *et al.*, 2009). Concordantly to studies in other *Cardiocondyla* species (Heinze & Trenkle, 1997; Yoshizawa *et al.*, 2009), we observed either the combination of queen and winged male traits (“gynandromorphs”), or the combination of worker and wingless (ergatoid) male traits (“ergatandromorphs”), whereas other combinations of male and female traits were absent (Table S3.6). This suggests that common pathways control *Cardiocondyla* morph differentiation in both sexes. Alternatively, observational bias due to mortality in early developmental stages may explain the absence of other combinations.

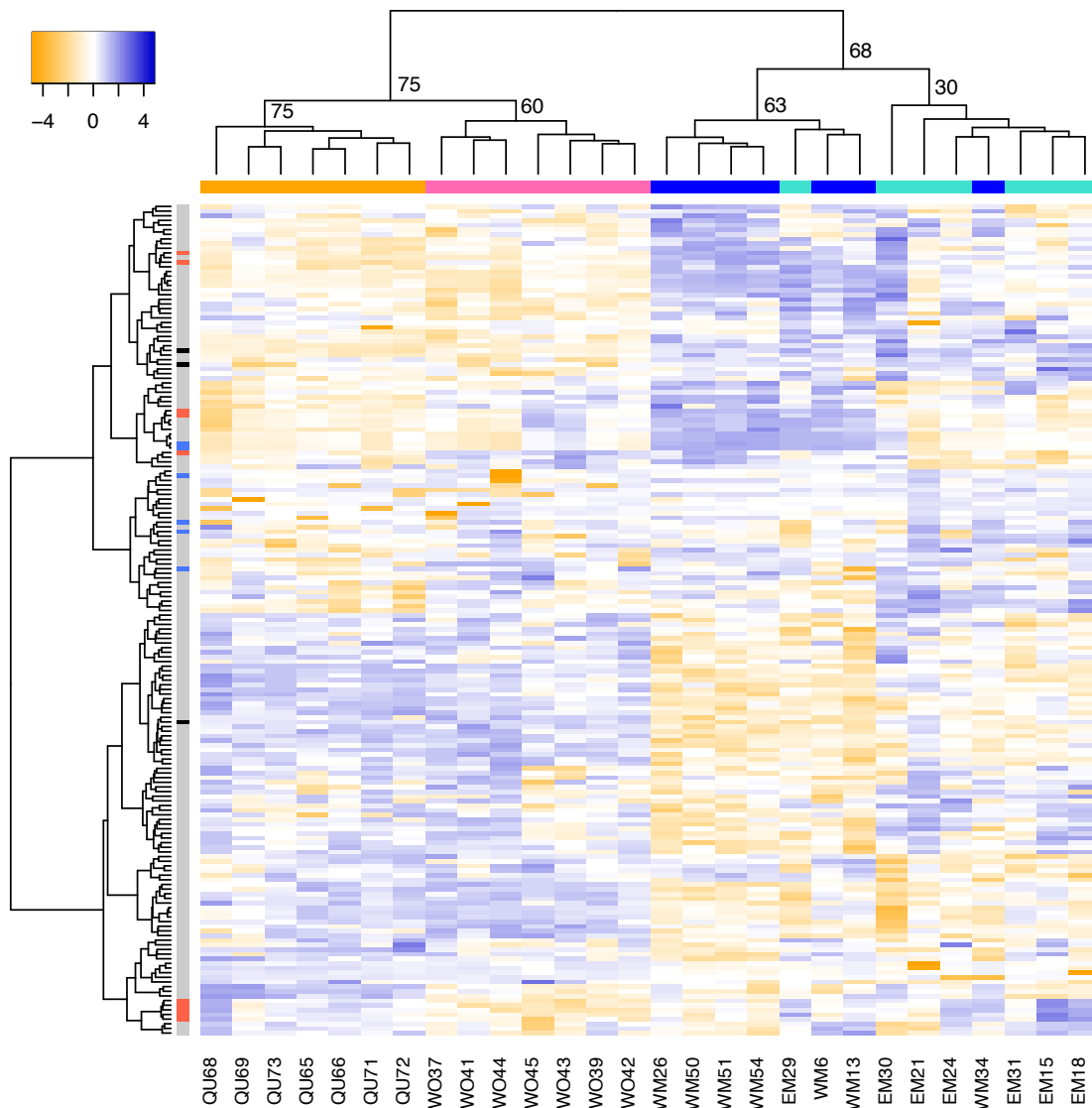


**Figure 3.2: Tissue-specific expression of *dsx* isoforms in sex mosaics of the ant *Cardiocondyla obscurior*.** (A) *dsx*<sup>F</sup> expression is higher in female (red triangles) than male halves (blue squares) of each individual, whereas *dsx*<sup>M</sup> expression is higher in male than female halves, with one exception (#10) (B).

## Co-option of *dsx* and other candidates in larval morph differentiation

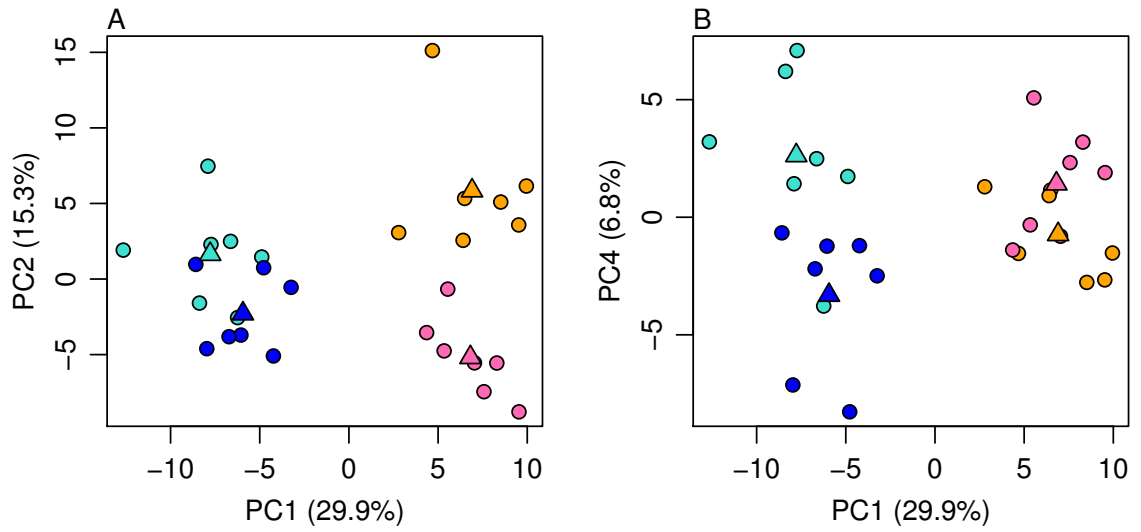
We analyzed published RNAseq data from individual early 3rd instar larvae (QU, EM, WM, WO; n=7 each; Schrader *et al.*, 2015) on an exon-level with DEXSeq (Anders *et al.*, 2012). We found morph-biased expression in each of the seven *dsx* exons and confirmed sex-specific expression of the DM domain, *dsx*<sup>F</sup> and *dsx*<sup>M</sup> in the early 3rd larval stage (Figure S3.4, Table S3.7). Overall, *dsx* expression was higher in males than in females, and higher in wingless morphs compared to winged morphs (EM > WM, WO > QU). Similar to *dsx*, other genes with sex-biased expression of alternative splice forms may have been co-opted for morph differentiation. At a conservative false discovery

rate of 0.005, DEXSeq analysis revealed 179 exons of 91 genes with sex-biased expression (Table S3.8). To test for co-option of this set of exons into morph differentiation, we performed a hierarchical clustering analysis based on log-transformed exon counts. Queens and workers as well as winged and wingless males were clearly separated by the set of sex-biased exons with the exception of two male samples (bootstrap node support: QU/WO=75, WM/EM=68) (Figure 3.3, see Figures S3.5 S3.6 for bootstrap support for all nodes and exon names). As terminal switch points for morph differentiation in male and female larvae may differ (Schrader *et al.*, 2015), misclassification of two male samples (WM34 & EM29) in hierarchical clustering may reflect higher plasticity in males compared to females at this particular developmental stage. Indeed, in *C. obscurior* 3rd instar larvae more genes are differentially expressed between queens and workers than between winged and wingless males (Schrader *et al.*, 2015). To identify sex-biased exons that most strongly affect separation between sexes and morphs, we performed a principal component analysis (PCA) of the 179 normalized exon counts (prcomp in R 3.0.3 GUI). PC1 separated sexes (29.9 % of the variation in the data explained) while PC2 (15.3 %) and PC4 (6.8 %) separated female and male morphs, respectively (Figure 3.4). We confirmed that PCs 1, 2 and 4 significantly separate between sexes and morphs with linear discriminant analysis (Wilks test; factor sex:  $F(1,28)=71.33$ ,  $p<0.001$ ; factor morph:  $F(3,28)=22.80$ ,  $p<0.001$ ). From the 179 exons we identified those exons with the strongest influence on sex (PC1), female morph (PC2) and male morph (PC4) by extracting the exon loadings that fell in either the 10 % or 90 % quantiles for each PC (Table S3.8). We further filtered this list to extract only those genes that contained multiple exons with strong influence on both sex (PC1) and morph (PC2 and/or PC4). This gave a list of eight highly conserved sex-biased genes showing clear morph bias in exon-specific expression (Table 3.1). Using a Gene Ontology (GO) term enrichment analysis with topGO (Alexa & Rahnenfuhrer, 2010) to infer biological function we found that these candidate genes consistently perform basic metabolic functions (Table S3.9).



**Figure 3.3: Heat map representing hierarchical clustering of normalized counts for sex-biased exons across the four morphs of the ant *Cardiocondyla obscurior*.** Females are clustered into the two castes: workers (WO, pink) and queens (QU, orange). Males also cluster into the two morphs: winged males (WM, dark blue) and ergatoid males (EM, light blue), with the exception of two samples (WM34, EM29). In the heat map, yellow indicates low expression and blue indicates high expression. Columns represent exons that exhibit sex-biased expression. As revealed by principal component analysis (PCA), exons whose loadings fell into the 10 % or 90 % quantile on PC1 and either PC2 (separating female morphs) or PC4 (separating male morphs) are indicated in red and blue, respectively. *dsx* exons are highlighted in black (top: exon 7, middle: exon 6, bottom: exon 5). Numbers in the sample-tree show bootstrap values.





**Figure 3.4: Principal component analysis (PCA) of the 179 sex-biased exons in the ant *Cardicondyla obscurior*.** PC1 separates the sexes (A+B), whereas PC2 separates workers (WO, pink) and queens (QU, orange) (A) and PC4 separates winged males (WM, dark blue) and ergatoid males (EM, light blue) (B). Triangles show group centroids.

**Table 3.1:** Candidate genes for co-option from sex differentiation into female caste differentiation (F) or male morph differentiation (M) pathways.

<i>Gene ID</i>	<i>Dmel ortholog</i>	<i>Blastp value/Query/Ident</i>	<i>Alternative function</i>	<i>GO IDs</i>	<i>GO terms (biological process)</i>
<i>Cobs_01393</i>	<i>dsx</i>	Manually corrected gene model	Exon 5,6: M Exon 7: Sex+M	GO:0043565, GO:0006355, GO:0005634, GO:0007548	regulation of transcription, DNA-templated, sex differentiation
<i>Cobs_03321</i>	<i>Mtpalpha</i>	0, 98%, 65%	Exon 1,2: F Exon 4,5: Sex	GO:0003857, GO:0016507, GO:0006635, GO:0050662, GO:0000166, GO:0004300, GO:0006550, GO:0006552, GO:0006554, GO:0006568, GO:0006574, GO:0006633, GO:0018874, GO:0019482, GO:004625 GO:0006468, GO:0005524, GO:0004674, GO:0051301, GO:0009069	limonene catabolic process, beta-alanine metabolic process, benzoate metabolic process, tryptophan metabolic process, isoleucine catabolic process protein phosphorylation, cell division, serine family amino acid metabolic process
<i>Cobs_08682</i>	<i>Cdk4</i>	3e-102, 55%, 51%	Exon 1: F Exon 3: Sex+M		
<i>Cobs_04840</i>	<i>Cap-D2</i>	0, 72%, 30%	Exon 12,13: Sex Exon 19,20: M	GO:0007067, GO:0005634, GO:0030261, GO:0051301, GO:0055114, GO:0008137, GO:0005488	cell division, mitotic nuclear division, oxidation-reduction process
<i>Cobs_05728</i>	<i>RpL5</i>	2e-165, 30%, 76%	Exon 12, 13: M Exon 15,16: Sex	GO:0005840, GO:0008097, GO:0003735, GO:0006412, GO:0042254, GO:0005524, GO:0030983, GO:0006298	translation, ribosome biogenesis, mismatch repair
<i>Cobs_12024</i>	<i>CG4822</i>	2e-152, 67%, 41%	Exon 3: Sex Exon 6,9,10:F	GO:0016021, GO:0006200, GO:0005524, GO:0016887, GO:0003723, GO:0033897	ATP catabolic process
<i>Cobs_14042</i>	<i>Kr-h2</i>	5e-74, 90%, 48%	Exon 1: Sex Exon 3,4: F	GO:0016021	No annotated GO term for Biological Process
<i>Cobs_15702</i>	<i>asp</i>	1e-150, 59%, 36%	Exon 2,6: Sex	GO:0005815, GO:0008017, GO:0005875, GO:0051295, GO:0004672, GO:0007282, GO:0000922, GO:0030723, GO:0048132, GO:0007052, GO:0051297, GO:0051642, GO:0030037, GO:0051383, GO:0030706, GO:0032027, GO:0045298	kinetochore organization, centrosome localization, cystoblast division, actin filament reorganization involved in cell cycle

### 3.4 Discussion

In the social Hymenoptera, eusociality with reproductive division of labor and accompanied morphological specialization has evolved several times independently. Repeated independent evolution of caste systems, together with an increasing amount of data showing the influence of conserved genes on phenotypic variation (Kijimoto *et al.*, 2012; Alvarado *et al.*, 2015; Schrader *et al.*, 2015), point towards repeated recruitment of basal developmental pathways as key factors in generating phenotypic divergence among social Hymenopteran castes. Here, we identify candidate sex differentiation genes that have been co-opted in the developmental processes of female and male morph differentiation in an ant. Among these is the master sex differentiation gene *dsx*, which along with most of the genes we identified is involved in highly conserved pathways across animal taxa.

Sex differentiation genes, and in particular genes of the *doublesex-mab3* family, and their effects on sex-specific phenotypic divergence are well documented in insect model species. For example, differential expression of *dsx* isoforms controls the formation of genital discs (Chatterjee *et al.*, 2011) and secondary sexual traits such as sex combs (Tanaka *et al.*, 2011) and abdominal pigmentation (Clough *et al.*, 2014) in fruit flies and modulates the development of sexual ornaments in beetles (Kijimoto *et al.*, 2012). Intriguingly, the results of these studies also indicate that differential *dsx* expression can be associated with the origin and diversification of novel phenotypic traits: in *Drosophila*, three distantly related species have independently evolved novel male-specific structures and display corresponding gains of *dsx* expression (Tanaka *et al.*, 2011). Similarly, manipulation of *dsx* expression in *Onthopagus* beetles induces the formation of horns, a typical male-specific trait, in otherwise hornless females (Kijimoto *et al.*, 2012). This indicates that small regulatory changes in core gene pathways can have extensive impact on a wide range of morphological, physiological and behavioral characters and lead to the evolutionary change required to generate adapted phenotypes.

Pleiotropy of these genes makes them suitable candidates for co-option, i.e., the evolution of phenotypic novelty through re-wiring of core gene sets. The evolutionary advantage of re-wiring core gene sets is rooted in their wide spectrum of targets, as weak linkage is an important prerequisite for changes in regulatory pathways (Gerhart & Kirschner, 2007). Accordingly, in developing *C. obscurior* we found only four genes (*dsx*, *CycB*, *aub*, *CG12316*) with large effect on separation between sexes that also show sex-specific expression patterns in other insect models (Shukla & Palli, 2012a;

Clough *et al.*, 2014). However, the lack of overlap between sex-biased genes across taxa may in part be due to differences in sampling, as sex-biased gene expression is known to be highly tissue- and age-specific (Clough *et al.*, 2014).

Compared to constitutively expressed genes that generally underlie some degree of redundancy to buffer the effects of deleterious mutations (Ohno, 1970), sex-biased genes already carry an inherent switch-like behavior and alternative states of function, a predicted requirement for co-option of conserved genes in novel developmental contexts (Gerhart & Kirschner, 2007). Our study suggests that in *C. obscurior*, sex-biased genes function in two distinct developmental contexts. Early in development, sex specific expression of *dsx* isoforms lead to differential regulation of *dsx* target genes in a tissue-specific manner resulting in sex-specific morphological differentiation (Bopp *et al.*, 2014). Later during larval development, this regulatory gene network is again initiated, this time by an environmental stimulus such as larval nutrition. Environmentally induced differential transcription of *dsx* and other sex-biased genes then initiates developmental divergence into distinct female and male morphs. Such co-option of sex-biased genes in morph differentiation pathways has been demonstrated in beetles, where nutrition-mediated changes in *dsx* expression (Kijimoto *et al.*, 2012) regulate tissue-specific juvenile hormone (JH) sensitivity and result in development of alternative male morphs (Gotoh *et al.*, 2014). Similarly, morph determination in *C. obscurior* is environmentally induced and can be influenced with JH-analogues (Schrempf *et al.*, 2006; Schrader *et al.*, 2015), making *dsx* a likely player to control development of intra-sexual dimorphisms via regulation of JH-sensitivity in divergently developing tissues and life stages.

Our study suggests that regulatory changes in larval development, induced by environmentally mediated transcription of *dsx*, *kr-h2* and six other sex-biased genes (Table 3.1) and most likely involving JH signaling, provide the basis for morph differentiation in *C. obscurior*. Specifically, successful induction of winged morphs with JH-analog, together with higher expression of *dsx* in wingless compared to winged morphs during larval development, point toward a role for *dsx* in mediating JH action in developing tissues. Differential *dsx* expression seems to play a particularly prominent role in male morph development, with wingless males showing significantly higher expression in six of seven *dsx* exons. Conversely, differential expression of *kr-h2* separated strongly between female morphs. *kr-h2* has structural similarity to the JH-inducible transcription factor *kr-h1* (Benevolenskaya *et al.*, 2000), which mediates initiation of metamorphosis in flies and beetles (Minakuchi *et al.*, 2009, 2008). *Kr-h2*-induced differences in developmental timing may explain why metamorphosis is delayed in queens compared to workers (Schrempf *et al.*, 2006; Schrader *et al.*, 2015). Overall, the strong effect of dif-

ferent isoforms of our eight candidate genes on separation of sexes and morphs suggests that these genes play a dual role in *C. obscurior*, mediating sex- and morph-specific development.

To conclude, the involvement of a diverse range of mechanisms that have co-evolved in concert with eusociality (from hormones, Wheeler (1990), to methylation patterns, Simola *et al.* (2013)) have made it difficult to pinpoint the evolution of caste determination systems. While we cannot dissect cause from effect, our study provides substantial evidence that the evolution of eusociality and caste polyphenism is likely mediated and maintained by a set of core genes implicated in the conserved developmental process of sex differentiation. Co-option of sex-biased genes allows exploitation of evolutionarily refined regulatory networks in a novel context, thus facilitating the coordinated developmental switch necessary for morph determination.

## 3.5 Material and Methods

### Verification of haplodiploid genetic sex determination

Crosses between five queens of a *C. obscurior* population from Japan (JP) and five wingless males of a *C. obscurior* population from Brazil (BR) were set up by placing the sexual pupae together with some brood and 20 workers in plaster-filled Petri dishes. Nests were checked twice a week, provided with water, honey and pieces of dead insects and kept at constant conditions (12 h 28 °C light, 12 h 23 °C dark). We sampled emerging F1 hybrid QU, WO, EM and WM pupae and extracted DNA from the 10 parental and 71 F1 individuals (23 EM, 3 WM, 22 QU, 22 WO, 1 GY = gynandromorph, for sample sizes per family see Table S3.6) using a modified protocol after Sambrook & Russell (2001). Each individual was analyzed at three microsatellite loci (Cobs\_1.1, Cobs\_8.3, Cobs\_8.4; for primer sequences see Table S3.3), which are variable between the two populations. PCRs were performed using the BIO-X-ACT™ Short Mix (Bioline) and microsatellite analyses were carried out on an ABI PRISM® (Applied Biosystems).

## Verification of functional *dsx* and its sex-specific isoforms in *C. obscurior*

We reanalyzed previously published RNAseq data of larvae (Schrader *et al.*, 2015). After raw-reads were cleared from adapter sequences with cutadapt and quality filtration was performed with Trimmomatic (Bolger *et al.*, 2014), reads were mapped against the reference genome with tophat2 (v2.0.8) and bowtie2 (v2.1.0) in sensitive mode. We generated count tables with HTseq based on the Cobs 1.4 official gene set and used DESeq2 (Love *et al.*, 2014) to assess sex-specific expression of the four *dsx* paralogs following size factor normalization. We applied RACE (Rapid Amplification of cDNA Ends) for identification of *dsx* isoforms. Total RNA was extracted from three females (QU adult, QU pupa, WO pupa) and three wingless males (one pupa, two adults) using the peqGOLD MicroSpin Total RNA Kit (peqlab). Transcription to cDNA was performed with the AffinityScript Multiple Temperature cDNA Synthesis Kit (Agilent Technologies), using the 3' RACE Adapter GCGAGCACAGAATTAAT-ACGACTCACTATAGGTTTTTTTTTTTNN. 3' RACE was performed in a nested PCR using two gene-specific 3' primers (*dsx4\_for4*, *Co\_dsx\_p3\_for*, for primer sequences see Table S3.2) and the 5' primer provided in the First Choice<sup>®</sup> RLM-RACE Kit (Ambion). PCRs were performed using the BIO-X-ACT<sup>™</sup> Short Mix (Bioline) with the following protocol: 94 °C (3 min), followed by 35 cycles 94 °C (30 sec), 60 °C (30 sec), 72 °C (2 min) and a final elongation of 72 °C (7 min). The products were purified with the NucleoSpin<sup>®</sup> Gel and PCR Clean-up (Macherey-Nagel) and Sanger sequenced at LGC Berlin.

## Expression of *dsx* and its isoforms in morphs and sex mosaics with real-time quantitative PCR (qPCR)

Total RNA was extracted from adults (8 EM, 8 WM, 10 QU, 10 WO) using the RNeasy Plus Mini Kit (Qiagen) and transcribed to cDNA using the AffinityScript Multiple Temperature cDNA Synthesis Kit (Agilent Technologies). Expression of the DM domain was quantified by qPCR using the primer pair *dsx4\_for4/dsx4\_rev1* and normalized with two housekeeping genes (*RPS2\_new*, *RPL32*; see Table S3.2 for primer sequences).

We further used qPCR to measure isoform-specific *dsx* expression (*dsx<sup>M</sup>* and *dsx<sup>F</sup>*), in pupae and adults of all four morphs, respectively, and in tissue from four sex mosaic pupae. We dissected the head and thoraces of the sex mosaics (for morphological descriptions see Table S3.6) laterally in male and female halves and stored male and

female tissue parts separately in RNAlater<sup>®</sup>-ICE (Ambion), resulting in one female and one male sample per individual. We extracted total RNA from 9-10 pupae and seven adults of each of the four phenotypes and of the sex mosaic tissue using the peqGOLD MicroSpin Total RNA Kit (peqlab) including a DNA digestion step with the peqGOLD DNase I Digest Kit (peqlab). After cDNA synthesis with iScript<sup>™</sup> cDNA Synthesis Kit (Bio-Rad) we quantified gene expression of *dsx*<sup>F</sup> and *dsx*<sup>M</sup> using isoform specific, intron-spanning primers ( *dsx*<sup>F</sup>: 4for/F5rev, *dsx*<sup>M</sup>: 4for/M5rev; see Figure 3.1A for position of primers) and two housekeeping genes (RPS2\_new, Y45F10D\_JO1). All qPCR reactions were performed in triplicates. Data analysis was carried out according to Livak & Schmittgen (2001), using the geometric mean of two housekeeping genes for normalization.

## Co-option of *dsx* and other candidates in morph differentiation

We analyzed published RNAseq data (Schrader *et al.*, 2015) from 3rd star instar larval QU, WO, WM and EM (n=7 each) and assessed differential exon-specific expression with DEXSeq (Anders *et al.*, 2012). Raw reads were trimmed and passed through quality filtration as described in Schrader *et al.* (2015) and mapped with STAR (Dobin *et al.*, 2013) against the reference genome Cobs1.4. (Schrader *et al.*, 2014). We corrected the *dsx* and *tra* gene model using the RACE results for *dsx* and splitting the *tra* gene model into two paralogs (*tra1* and *tra2*), as observed in other ants (Schmieder *et al.*, 2012; Privman *et al.*, 2013; Koch *et al.*, 2014). For all other genes we used gene models of the Cobs 1.4 official gene set. We followed the default workflow of DEXSeq (Anders *et al.*, 2012) and tested for differential exon usage between males and females based on a false discovery rate of 0.005. In the resulting 179 sex-specific exons we analyzed morph-specific signatures using a hierarchical clustering approach (ward.D2 method, Murtagh & Legendre, 2014) of pairwise Manhattan distances between log-transformed normalized exon counts of all four morphs, using bootstrap resampling of 10,000 replicates in gplots (Warnes *et al.*, 2015) and pvclust (Suzuki & Shimodaira, 2015) in R 3.1.2 (R Core Team, 2013).

Moreover we conducted a PCA with normalized exons counts. We visually identified principal components that best separated between sexes (PC1), female morphs (PC2) and male morphs (PC4) and confirmed that these components suffice to separate among sexes and morphs with linear discriminant analysis and subsequent Wilks tests in R 3.1.2. Based on loadings of exons on each component we identified exons that fell in either 10 % or 90 % quantile (Table S3.8) as those with the strongest influence on PC1, PC2 and PC4. From this list, we extracted only those genes that contained mul-

multiple exons with strong influence on both sex (PC1) and morph (PC2 and/or PC4). This gave a list of eight highly conserved sex-biased genes showing clear morph bias in exon-specific expression (Table 3.1).



# **4 Local recombination rate is positively correlated with genetic diversity and negatively linked to transposable element content in an inbred ant**

Antonia Klein<sup>1</sup>, Olav Rueppell <sup>2</sup>, Jürgen Heinze<sup>1</sup>, and Jan Oettler<sup>1</sup>

<sup>1</sup> Institut für Zoologie, Universität Regensburg, Regensburg, Germany

<sup>2</sup> Department of Biology, University of North Carolina, Greensboro, North Carolina, United States of America

## 4.1 Abstract

Eusocial insects have the highest rates of homologous recombination reported for multicellular organisms so far, which may have promoted the evolution of the worker phenotype or may underlie traits generally associated with genetic heterogeneity, e.g. division of labor and immune repertoire. Here we present a linkage map of the ant *Cardiocondyla obscurior*, based on 85 recombinant F2 males from a cross between two *C. obscurior* populations. The map contains 3,130 SNPs and spans 1,793.9 cM. Overall recombination rate of 10.0 cM/Mb lies within the range of other eusocial species. Local recombination rates were determined for 53.2 % of the genome with a median of 8.08 cM/Mb (25 % quartile: 0.00; 75 % quartile: 22.58). However, several regions where interpopulational recombination probably has been impeded by larger structural variants between the two populations suggest that intrapopulational recombination rates could be higher. Low recombining regions are enriched for transposable elements, an alternative putative source of genetic variation. On the other hand, recombination rates correlate positively with SNP, InDel and SSR content, showing that recombination drives divergence between the two populations. Using published RNAseq data on larval phenotypes, we found elevated recombination rates in wingless male-biased genes (peculiar fighter males typical for the genus *Cardiocondyla*) compared to winged male-biased genes (the ancestral phenotype) but not in worker-biased genes. This finding suggests that the evolution of wingless males under strong directional selection could be facilitated by high recombination rates and therefore rapid sequence evolution and highlights the importance of species-specific selection pressures on local recombination rates.

## 4.2 Introduction

Homologous recombination is conserved across all domains of life. On the one hand meiotic recombination is important for ensuring proper segregation of homologous chromosomes and to act as a repair mechanism caused by double strand breaks (Davis & Smith, 2001). On the other hand homologous recombination generates offspring genetic diversity required for selection and simultaneously leads to a disruption of parental genotypes that have been tested successful by natural selection. Despite being one of the oldest recognized genetic phenomena (Morgan, 1911) the role of recombination in evolution is still not satisfactory understood. Recombination rates vary across species, between sexes (Dumont & Payseur, 2011), and also within the genomes of

single species (Baudat *et al.*, 2010), possibly reflecting differences in effective population sizes, adaptation to different selection pressures on sexes as well as properties of genome architecture.

Comparisons of genetic linkage maps to physical genome size across honey bees, bumblebees, ants, and wasps (Beye *et al.*, 2006; Meznar *et al.*, 2010; Sirviö *et al.*, 2006, 2010, 2011) suggest that eusocial insects have evolved increased recombination rates compared to parasitic Hymenoptera (Holloway *et al.*, 2000; Niehuis *et al.*, 2010) and other insects (Yamamoto *et al.*, 2006). However, the underlying ultimate causes and proximate mechanisms are not yet resolved (Wilfert *et al.*, 2007). It has been suggested that high recombination rates could have been important during the evolution of eusociality and in particular may have facilitated the evolution of the worker phenotype (Kent *et al.*, 2012). Alternatively high recombination could result in an increase of genetic diversity, predicted to be important for the immune repertoire of social insect colonies that are anticipated to be highly susceptible to pathogens and parasites because of high densities of group-living individuals (Fischer & Schmid-Hempel, 2005). Common to both hypotheses is the prediction of high recombination rates in worker-biased genes, but data is scarce and remains controversial (Liu *et al.*, 2015b; Wallberg *et al.*, 2015). Moreover, a simulation study showed that recombination did not contribute to offspring genetic diversity, calling for alternative explanations of high recombination rates in eusocial insects (Rueppell *et al.*, 2012).

The ant species *Cardiocondyla obscurior* is an emerging genomic model for eusocial evolution (Schrader *et al.*, 2014; Klein *et al.*, 2015; Schrader *et al.*, 2015; Schrempf *et al.*, 2015). In addition to reproductive queens and non-reproductive workers, a peculiar male diphenism with two distinct male morphs makes *C. obscurior* especially suited for analysing correlations between recombination rates and morph biased-gene expression, adding one level of complexity compared to most social insects (Oettler *et al.*, 2010). The wingless, so-called “ergatoid” males are the prevalent male morph, whereas winged males are only rarely produced in *C. obscurior*, presumably under stressful environmental conditions (Cremer & Heinze, 2003). Wingless males remain in the maternal nest and compete for mating opportunities with closely-related virgin queens (Heinze & Hölldobler, 1993).

This unique life history leads to exceptionally high levels of inbreeding and genetic homogeneity within and among colonies. One mechanism to overcome depletion of genetic variation are transposable elements (TEs) which in *C. obscurior* accumulate in clusters (so called “TE islands”) and underlie the evolutionary divergence of two populations from Japan and Brazil (Schrader *et al.*, 2014). Here, we test the hypothesis

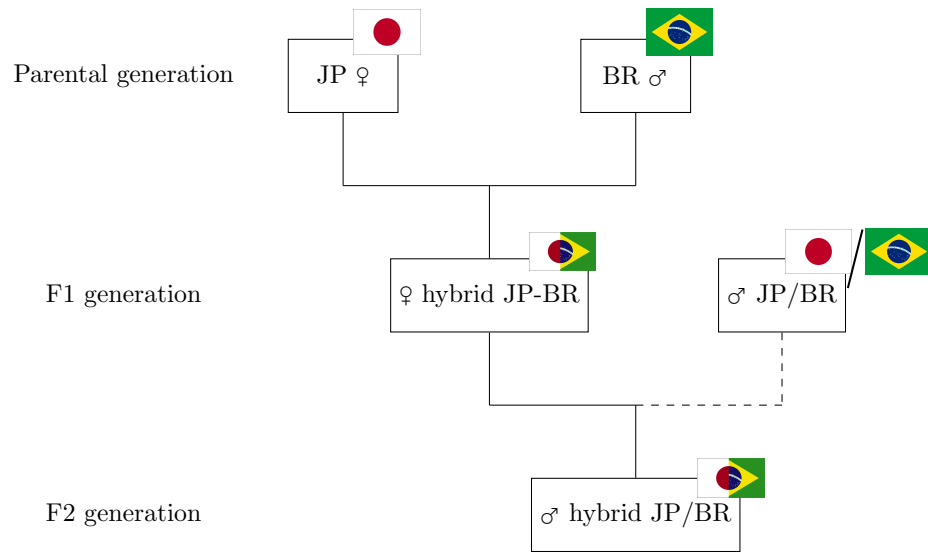
that in addition homologous recombination is largely responsible for generating genetic variation under extreme inbreeding. Second, we address whether morph-biased genes during the critical larval stage are correlated with increased recombination rates. In line with predictions, we find low rates of recombination in TE islands and a correlation of recombination rates with repeat content and genomic divergence between the two populations. Recombination in distinct genomic areas is inhibited and associated with the accumulation of structural variants such as deletions and duplications, whereas inversions were not correlated with low recombination rates. While we found no evidence for high recombination rates of worker-biased genes, increased recombination rates in wingless-male-biased genes compared to winged-male-biased genes may reflect the consequences of strong sexual selection.

## 4.3 Material and methods

### Crosses to generate mapping population

The haplodiploid Hymenoptera are ideal for linkage mapping because reproductive females produce haploid sons with a genome that directly represents meiotic products of the females. The construction of a genetic linkage map for the 177.9 Mb draft genome of *Cardiocondyla obscurior* (Schrader *et al.*, 2014) by an intrapopulational cross is complicated by high levels of inbreeding which result in genetic homogeneity within introduced founder populations. Thus, an interpopulational cross between two highly inbred populations of *C. obscurior* from Japan (JP) and Brazil (BR), which differ in 553,052 SNVs on a genomic level (Schrader *et al.*, 2014), was performed and F2 males were used as mapping population for linkage mapping (Figure 4.1).

The accuracy of linkage maps increases with sample size. However, the number of male offspring produced by one *C. obscurior* queen is insufficient to produce a high-quality map (e.g. Cremer & Heinze, 2002; Schrempf *et al.*, 2005). Thus, a multi-family design was used to sample sufficient numbers of F2 males. Four queens from JP (offspring of one highly inbred colony) were mated to a single male from BR. Although F2 males are produced via arrhenotokous parthenogenesis in haplodiploids, and fathers consequently do not contribute genetically, mating will increase the reproductive output of queens (Schrempf *et al.*, 2005), and the F1 queens were thus mated to males from either JP or BR. F2 males originating from the five F1 queens with the greatest reproductive output were sampled at the pupal stage as mapping population (Supplementary Figure S4.1).



**Figure 4.1: Experimental design.** Queens from Japan (JP) were all mated to the same male from Brazil (BR). Hybrid F1 queen pupae were transferred to new nests, backcrossed to ergatoid males either from JP or BR and hybrid F2 males were sampled at pupal stage. Backcrossing is represented by dashed lines, as in haplodiploids fathers do not contribute to the genetic material of their sons. We used F2 males originating from five different F1 queens, which were daughters of four different parental queens (for graphical explanation of the multi-family design see Supplementary Figure S4.1).

## DNA extraction and RADtag sequencing

Total genomic DNA of 90 F2 male pupae, the parental male, one parental queen and three F1 queens was extracted using a phenol-chloroform method (Gadau, 2009) followed by a RNase A digestion step. Genomic DNA was converted to nextRAD sequencing (next Restriction site associated DNA sequencing) libraries (SNPsaurus, Eugene, OR, USA). The nextRAD method (Russello *et al.*, 2015) for genotyping-by-sequencing uses a selective PCR primer to amplify genomic loci consistently between samples. Similarly to how RAD-Seq sequences the DNA flanking a restriction enzyme cut site, nextRAD sequences the DNA downstream of a short selective priming site. Genomic DNA (2.5 ng for haploid males, or 5.0 ng for diploid females) from each sample was first fragmented using a partial Nextera reaction (Illumina, Inc), which also ligates short adapter sequences to the ends of the fragments. Fragmented DNA was then amplified using Phusion Hot Start Flex DNA Polymerase (NEB), with one of the Nextera primers modified to extend six nucleotides into the genomic DNA with the selective sequence TGCAGG. Thus, only fragments starting with a sequence that can be hybridized by the selective sequence of the primer were efficiently amplified. The following PCR parameters were used: 72 °C for 3 minutes, 98 °C for 3 minutes, 22 cycles

of 98 °C for 45 seconds followed by 74 °C for 1 minute, then hold at 4 °C. The 95 dual-indexed samples were pooled and the resulting libraries were purified using Agencourt AMPure XP beads at 0.7x. The purified library was then size selected to 350-800 base pairs. Sequencing was performed using two lanes of an Illumina HiSeq2500 (Genomics Core Facility, University of Oregon) for the set of 95 individuals.

A list of loci was developed by gathering reads from a sampling of the population and removing those with read counts above 1,500, suggesting misassembled repetitive genomic sequence, or low read counts suggesting they are off-target or error reads in nature. The remaining read instances were aligned to each other to identify alleles and a single read instance was chosen to represent the locus in a pseudo-reference. All reads from each sample were then aligned to the pseudo-reference with bwa-mem (Li & Durbin, 2009) and converted to a vcf genotype table using Samtools.mpileup (filtering nucleotides with a quality of 13 or better), and bcftools call (Li, 2009). This resulted in 6,189 SNPs that were distinct between the BR and the JP *C. obscurior* populations. Two F2 males and one F1 queen were discarded because of low-quality libraries.

## Linkage Mapping

For linkage mapping, two additional quality filters were applied: First, only markers that showed distinct genotypes in the sequenced parental queen and the parental male were selected, reducing the dataset to 4,049 markers. Second, to account for the multi-family design, markers that did not segregate in a specific family were excluded by recoding them as missing data in that particular family. However, this was only the case for 68 markers in total (replaced markers per family: F21: 5; F96: 18; F152: 5; F204: 7; F252: 33).

Mapping was done ‘phase unknown’ as described in Sirviö *et al.* (2006). In short, the whole dataset was duplicated and phase-inverted to assign linkage groups. Mapping was done with the qtl package (Broman *et al.*, 2003) in R (R Core Team, 2013). Three males were excluded because of too many missing data points, resulting in a final mapping population of 85 males from five families, with 17, 18, 21, 16 and 13 males respectively. After discarding markers with significant segregation distortion based on a FDR of 0.05 as the last quality-filtering step, the final dataset consisted of 3,130 markers. For mapping, the Kosambi map function (Kosambi, 1944) was applied, which takes potential double recombination events into account and is generally applicable to social Hymenoptera (Solignac *et al.*, 2007).

In a first map construction step, linkage groups were formed with a minimum LOD score of 5.0 and a maximum recombination fraction of 0.1, meaning that two loci are

regarded as linked when at most 10 % of the mapping population show recombinant genotypes. Marker order on each linkage group was optimized with the `orderMarker` function, combined with manual corrections by checking recombination fractions of exterior markers of the linkage group and exterior markers of its subclusters for all regions on the linkage map with gaps >10 cM. When the gap size between two marker-clusters on a linkage group exceeded 50 cM, this linkage group was split in two groups. Finally, the exterior markers of each linkage group were checked for connections to other linkage groups and linkage groups were merged, if recombination frequencies of <20 % were found.

To estimate the percentage of the physical assembly covered by the linkage map, we summed up the distances between the marker with the lowest and the marker with the highest position for each scaffold on each linkage group. Scaffolds that were represented by <5 markers were discarded and all distances for all scaffolds and all linkage groups were summed up and divided by the size of the draft genome (182 Mb). To estimate overall recombination rate of *C. obscurior*, the size of the genetic map (1,793.9 cM) can be divided by the physical size of the genome covered by the genetic map (180 Mb). However, in ~50 % of the genome local recombination rates could not be estimated as the linkage map was not supported the physical assembly in these areas (see below). Consequently, a better approximation of *C. obscurior* overall recombination rate is the median of all local recombination rates.

To analyse the distribution of markers across exons, introns and intergenic regions, markers within these genomic elements were counted with `BEDtools intersect` (Quinlan & Hall, 2010). The observed numbers were compared to the expected numbers, based on the overall abundance of exons, introns and intergenic regions in the reference genome Cobs 1.4, using Chi-squared tests.

## Detection of structural variants between BR and JP

Paired end genomic reads with an insert size of 200 bp from a pool of 26 JP ergatoid males generated on an Illumina HiSeq platform (Schrader *et al.*, 2014) were used. After quality filtration and adapter trimming with `Trimmomatic 0.33` (Bolger *et al.*, 2014), the JP reads were mapped against the BR reference with `BWA` (Li & Durbin, 2009). Subsequently, `Pindel` (Ye *et al.*, 2009) was used to detect structural variants (inversions, deletions, duplications) between the two populations. While `InDels` separating the BR and JP *Cardiocondyla obscurior* population detected by Schrader *et al.* (2014) showed a maximal length of 66 bp, `Pindel` is able to detect longer structural variants due to

usage of paired end information including insert size. Results were filtered for structural variants supported by at least 10 reads of at least 100 bp length.

## Estimation of local recombination rates and correlation with genomic features

For estimation of local recombination rates genomic regions where genetic and physical map corresponded sufficiently were selected, meaning that at least five markers of the same scaffold were adjacent on the genetic map. For each linkage group, markers not belonging to the scaffolds mostly represented on the linkage map were discarded, as well as markers with no or multiple matches in the physical genome assembly. Next, the remaining markers were ordered according to the physical assembly and the resulting map was compared to the original version of the genetic map before marker discarding and ordering. As this did not result in severe map extension for any of the used linkage groups, this new map was regarded as a correct representation of recombination events (Supplementary Table S4.2). For each pair of adjacent markers recombination rate was calculated by dividing the genetic distance between them (in cM) by the physical distance between these markers (in Mb). Subsequently, recombination rate was calculated in 100 kb sliding windows (slide=10 kb). For each window, the weighted mean of recombination rate was then calculated. All adjacent marker pairs falling within this window were chosen and the physical distance between them was used for weighting. “Hotspots” were defined as windows with recombination rates  $>50$  cM/Mb (94.9 % quantile), “coldspots” were defined as windows with zero recombination (0 cM/Mb, 35.8 % quantile). Genes located in those windows were identified with BEDtools intersect (Quinlan & Hall, 2010) and Gene Ontology (GO) term enrichment analyses were performed with topGO (Alexa & Rahnenfuhrer, 2010) to identify biological processes enriched in hot- and coldspots, respectively.

Correlations between local recombination rates (100 kb window) with the following genomic features were performed: GC-content, gene- exon- and intron content, overall repeat content, content of class I retrotransposons, class II DNA-transposons and simple repeats (SSRs), content of LINE, LTR and TIR elements, overall SNP content, content of homozygous and heterozygous SNPs and InDel content. Correlations were performed using the number of the corresponding genomic feature per 100 kb window as well as using the number of basepairs of the genome feature per 100 kb window. We used adjacent, non-overlapping windows (1/10 of the total dataset) to avoid pseudoreplication resulting from the sliding window approach. The genome features in these windows were determined by custom R and bash scripts, the R package “seqinr”



(Charif & Lobry, 2007) for calculating GC content and BEDtools intersect (Quinlan & Hall, 2010). As the distribution of recombination rates was left-skewed and to account for overdispersion, generalized linear models (GLM) of the quasipoisson family were used (McGaugh & Noor, 2012).

To test for the effect of longer structural variants (detected with Pindel) on local recombination rates, recombination rates of windows with or without structural variants were compared using Wilcoxon rank sum tests.

## Regions not supported by the physical assembly

The linkage map was divided into regions supported by the physical assembly, i.e. the marker order on the genetic map corresponds to the marker order in the physical genome assembly (supported regions = SR), and on the other hand into regions where the genetic map is not supported by the physical assembly (not supported regions = NR). In order to compare SR and NR, we analysed gene density, exon content, SNP content, InDel content, overall repeat content, content of class I transposons, class II transposons and SSRs, as well as localization in TE islands (cf. Schrader *et al.*, 2014). We analyzed in detail the most abundant repeat families, which are LINE and LTR for class I and TIR for class II elements. The 3,130 marker of the linkage map were divided in markers falling in SR or NR regions (SR: 2,089 markers, NR: 956 markers), and 1 kb windows around each of the markers (500 bp up- and downstream, respectively) were defined. The numbers of genomic features and/or the number of bases of the genomic features in these 1 kb windows were counted and summed up for SR and NR regions using BEDtools intersect (Quinlan & Hall, 2010) and custom R scripts and divided by the total number of markers in SR (2,089) or NR (956) for normalization.

Next, we analyzed whether structural variants detected by Pindel were enriched in NR compared to SR. For each of the three types of structural variants (inversions, deletions, duplications), markers located within the structural variant were selected with BEDtools intersect (Quinlan & Hall, 2010). The expected counts were calculated as the percentage of markers located in NR (28 %) and SR (72 %) of the total number of markers located in the structural variant. The observed versus expected counts in SR and NR regions were compared by a Chi-square test.

Finally, a Gene Ontology (GO) term enrichment analysis was performed with topGO (Alexa & Rahnenfuhrer, 2010) to identify biological processes overrepresented in NR regions.

## Correlation between recombination rate and morph-biased gene expression

For studying the relation between local recombination rate and morph-biased gene expression, published RNAseq data from Schrader *et al.* (2015) generated from larvae with known developmental trajectory of all four morphs (N=7 each) were used. From this data set we calculated the “norm”, a measurement for morph-bias, as described in Schrader *et al.* (unpublished). In short, for each gene a particular morph bias is represented as a vector in a three-dimensional space based on expression log-fold changes of this morph compared to all other three morphs. Those genes with the highest morph-bias based on the vector length “norm” (falling in the 95 % quantiles) were used for assessment of a correlation with recombination rates. This resulted in 78 genes for queens, 70 genes for workers, 60 genes for winged males and 69 genes for ergatoid males, while the 70 most unbiased genes were used as reference. For each gene, only one value for recombination rate was determined. If the gene was only located in one 100 kb window, the recombination rate of this window was assigned to the gene. If the gene was spanning more than one 100 kb window, the mean recombination rate of the corresponding windows was calculated. Recombination rates across gene sets were compared with a Kruskal Wallis test followed by pairwise comparisons using Wilcoxon rank sum test with Benjamini-Hochberg correction (Benjamini & Hochberg, 1995).

## 4.4 Results

### Linkage Map of *Cardiocondyla obscurior*

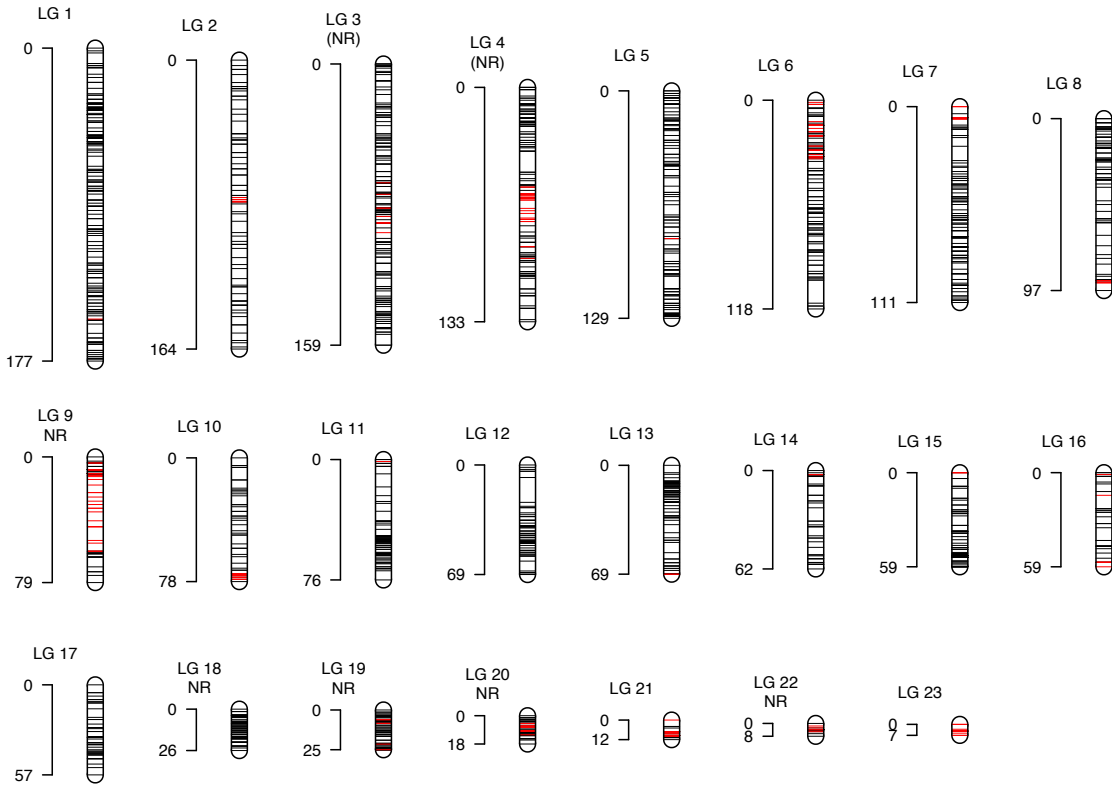
The physical sequence contained in the linkage map added up to 180 Mb. With a genome size of 182 Mb, the map covers 98.9 % of the draft assembly Cobs 1.1. The map contains all of the first 95 largest scaffolds of the genome assembly (Supplementary Figure S4.2), which contain 172,033,427 bp corresponding to 94.5 % of the draft genome assembly Cobs 1.1. 62.3 % of the markers are located in intergenic regions, 19.9 % in exons and 17.8 % in introns. Compared to the overall abundance of exons, introns and intergenic regions (in bp) in the *C. obscurior* genome (exons: 11.8 %, introns: 20.4 %, intergenic regions: 67.8 %), there is an overrepresentation of markers in exons compared to introns and intergenic regions (Pearson’s Chi-squared test:  $X^2=75.782$ ,  $df=2$ ,  $p<0.001$ , Supplementary Figure S4.3, Supplementary Table S4.1).

The initial map consisted of 24 linkage groups. Manual correction led to the split of

one linkage group and to combination of two linkage groups in two instances where markers were linked with recombination frequencies  $\sim 15\%$  and where the physical assembly supported linkage. The final linkage map consists of 23 linkage groups and spans 1793.9 cM (Table 4.1, Figure 4.1, Supplementary Figure S4.4). Based on the draft genome size of 182 Mb *C. obscurior* has an overall recombination rate of 10.0 cM/Mb. The 16 linkage groups that are completely supported by the physical assembly (=SR) span 1345.8 cM on the map and 85.1 Mb on the physical assembly, which results in a recombination rate of 15.8 cM/Mb.

**Table 4.1: Summary of the linkage map of *Cardiocondyla obscurior*.** Linkage groups (LG) are ordered by length in cM. Estimated physical size could not be calculated for regions where genetic map and physical assembly are inconsistent (LG 9, 18, 19, 20, 22 and inner parts of LG 3 and LG 4). Mean RR per LG was calculated by dividing length in cM by the estimated physical size in Mb.

LG	length in cM	number of markers	average spacing	maximum spacing	estimated physical size in Mb	mean RR in cM/Mb
1	177.2	279	0.6	6.7	13.7	12.9
2	163.6	93	1.8	7.1	5.8	28.2
3	159.2	280	0.6	6.0	>4.0	?
4	132.7	221	0.6	11.9	>7.7	?
5	128.8	172	0.8	7.1	6.2	20.8
6	118.2	208	0.6	13.1	8.1	14.6
7	110.9	167	0.7	8.1	6.5	17.1
8	97.3	104	0.9	7.6	4.6	21.2
9	79.2	143	0.6	8.6	?	?
10	77.9	80	1.8	7.1	3.7	21.1
11	75.8	112	0.7	8.2	5.1	14.9
12	69.0	103	0.7	15.7	4.5	15.3
13	68.6	103	0.7	6.4	6.0	11.4
14	62.0	54	1.2	8.4	2.3	26.9
15	59.3	130	0.5	6.3	5.9	10.1
16	59.2	47	1.3	8.3	1.3	45.5
17	58.8	59	1.0	5.7	2.8	21.0
18	26.1	172	0.2	2.4	?	?
19	25.0	160	0.2	2.4	?	?
20	17.8	187	0.1	2.5	?	?
21	12.3	95	0.1	4.0	3.5	3.5
22	8.1	107	0.1	1.6	?	?
23	6.9	54	0.1	3.1	5.1	1.4
total	1,793.9	3,13	0.7	15.7	96,8	17.9



**Figure 4.2: Linkage Map of *Cardiocondyla obscurior*.** Y-axes show the size of each linkage group (LG) in cM. LGs are sorted by genetic size. Red markers indicate markers in TE islands as defined by Schrader *et al.* (2014). Linkage groups which are completely inconsistent with the physical assembly are indicated by “NR” (not supported regions), linkage groups which are partly inconsistent with the physical assembly are indicated by “(NR)”.

## Estimation of local recombination rates and correlation with genomic features

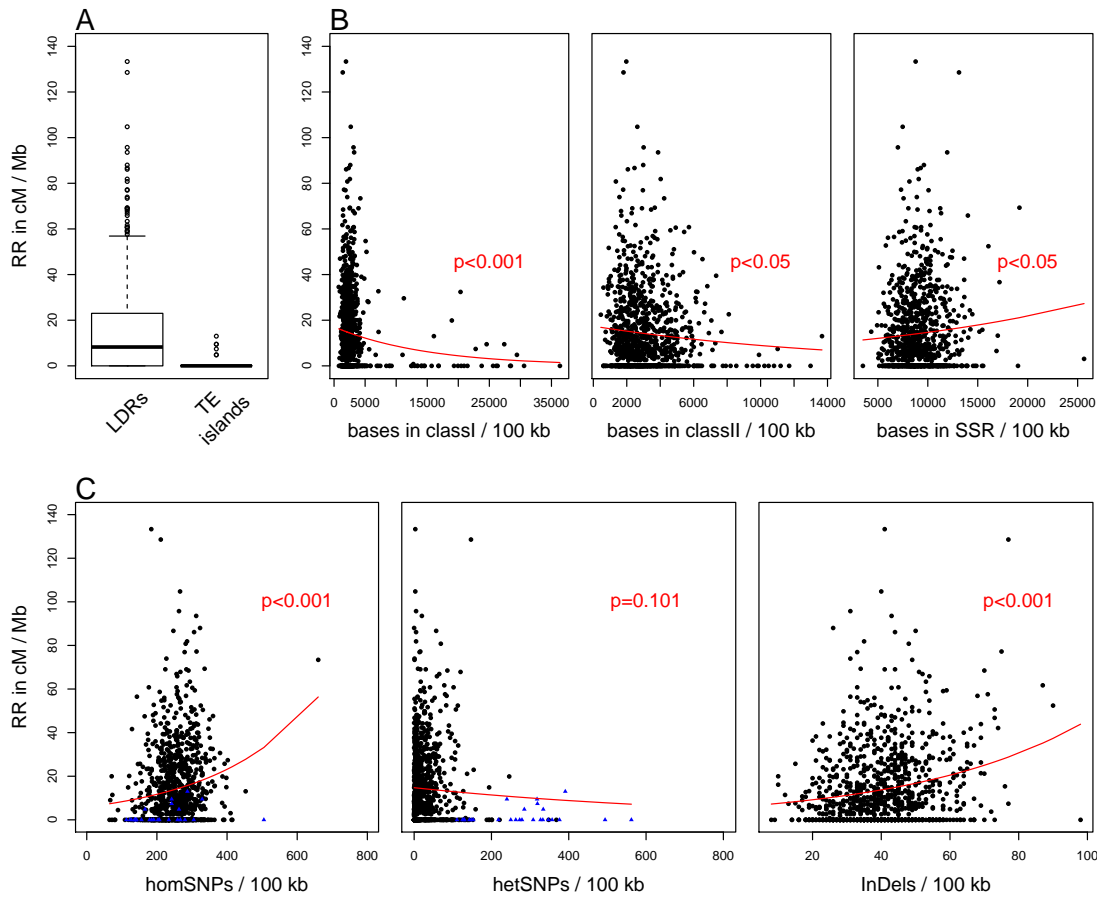
The physical assembly supported 16 of the 23 linkage groups (i.e. marker order between genetic and physical map is in accordance). Two linkage groups (LG 3 und LG 4) were only partly supported. The regions in LG 3 and LG 4 not supported are located in the middle of the linkage group. All supported regions (96.8 Mb) were used for calculation of local recombination rates, corresponding to 53.2% of the genome. Here, median recombination rate was 8.08 cM/Mb (25 % quartile: 0.00; 75 % quartile: 22.58). The recombination landscape of *C. obscurior* is not uniform (Supplementary Figure S4.5), with 35.8% of the windows showing zero recombination. Furthermore, 6.3% of the windows have a recombination rate between 0 and 5, 12.7% between 5 and 10 cM/Mb

and 45.1 % over 10 cM/Mb.

Based on a conservatively chosen FDR of 0.01 we found that “hotspots” (RR >50 cM/Mb) were enriched for genes involved in spermatocyte division, very long-chain fatty acid metabolic process and several GO terms associated with muscle development (Supplementary Table S4.3). In contrast, “coldspot” (RR=0 cM/Mb) were enriched in regulation of transcription and several processes involved in development of organs, neuronal tissue, glands and exocrine systems (Supplementary Table S4.4).

Local recombination rates neither correlate with GC content (overall and in exons, introns and intergenic regions, respectively), nor with gene, exon or intron content (Supplementary Table S4.5). In repeat-rich sequence formerly identified as concatenated TE islands (Schrader *et al.*, 2014) recombination rate was significantly reduced compared to regions with low TE content (low density regions; LDRs) (Wilcoxon rank sum test:  $W=18885$ ,  $p<0.001$ , Figure 4.3A). 963 windows are located in LDRs compared to 26 windows in TE islands. Consequently, 1,000 replicates of the test were performed to account for this bias. To this end 26 random samples of the LDR dataset were drawn and compared with the TE island dataset, resulting in a mean p-value of 0.005 with a SD of 0.02. Number of bases per 100 kb window in class I and class II elements are negatively correlated with recombination rate, respectively, whereas no correlation was found for the number of these elements per 100 kb window (Figure 4.3B, Supplementary Table S4.5). A closer inspection of class I and class II transposons showed that the number of bases of class I LTR and non-LTR LINE elements is negatively correlated with recombination rate, whereas no correlation with bases in class II TIR elements was found (Supplementary Table S4.5). However, most correlations between recombination rate and transposable element content dropped below significance when considering only LDRs (Supplementary Table S4.5). SSR content is positively correlated with recombination rate overall and also in LDRs (Figure 4.3B, Supplementary Table S4.5). Also recombination rate is positively correlated with homozygous SNPs and InDels, but not with heterozygous SNPs (Figure 4.3C, Supplementary Table S4.5).

Next, we analysed the effect of longer structural variants on local recombination rates in supported regions (SR). Windows that contain inversions (inv) and deletions (del) show significantly lower local recombination rates than windows without, respectively (Wilcoxon rank-sum tests: inversions:  $\text{median}_{\text{inv}}=5.3$ ,  $\text{median}_{\text{no inv}}=9.4$ ,  $W=14060368$ ,  $p<0.001$ ; deletions:  $\text{median}_{\text{del}}=7.4$ ,  $\text{median}_{\text{no del}}=8.5$ ,  $W=13495128$ ,  $p<0.001$ ). In contrast, recombination rates are significantly higher in windows with duplications (dup) than in windows without (wilcoxon rank sum test:  $\text{median}_{\text{dup}}=8.6$ ,  $\text{median}_{\text{no dup}}=7.5$ ,  $W=12140536$ ,  $p<0.01$ ).



**Figure 4.3: Local recombination rate (RR) is influenced by repeat content and degree of divergence in *Cardiocondyla obscurior*.** (A) RR is significantly higher in LDRs (low density regions) than in TE-islands. As sample sizes are very unbalanced (NLDR=963; NTE=26), we verified the result via random subsampling from the LDR dataset. (B)+(C) We used generalized linear models (GLMs) of the quasipoisson family to model correlations between RR and genomic features. Red lines depict GLMs. (B) Number of bases in class I and class II TEs and in SSRs per 100 kb window. RR is decreased in regions with many class I repeats ( $p < 0.001$ ) as well as class II repeats ( $p < 0.05$ ), but increased in regions with many SSRs ( $p < 0.05$ ). (C) Number of homozygous SNPs (homSNPs), heterozygous SNPs (hetSNPs) and Insertions-Deletions (Indels) counted per 100 kb window. Blue triangles represent windows located in TE-islands. This shows that RR is positively correlated with homSNPs ( $p < 0.001$ ), which are not enriched in TE-islands, but no correlation could be found for hetSNPs ( $p = 0.118$ ), which are enriched in TE-islands. RR is positively correlated with InDels ( $p < 0.001$ ).

## Regions not supported by the physical assembly

Marker order on five linkage groups (LG 9, 18, 19, 20 and 22) is inconsistent with the physical assembly. The size of these five linkage groups is relatively short (8.2 - 79.2 cM) in relation to the high number of markers they contain (107 - 187 markers). In addition, the middle regions of LG 3 and LG 4 were also not supported by the physical

assembly. The estimated physical size of the supported regions (SR) is half of the total genome size, and recombination rates could be calculated for 96.8 Mb, corresponding to 53.2 % of the total physical genome size.

We could not find evidence for decreased gene or exon content in NRs (Table 4.2). A GO term enrichment analysis at a FDR of 0.01 revealed that genes involved in reproduction, cytoskeleton organization and organ growth are enriched in non-recombining regions (Supplementary Table S4.6).

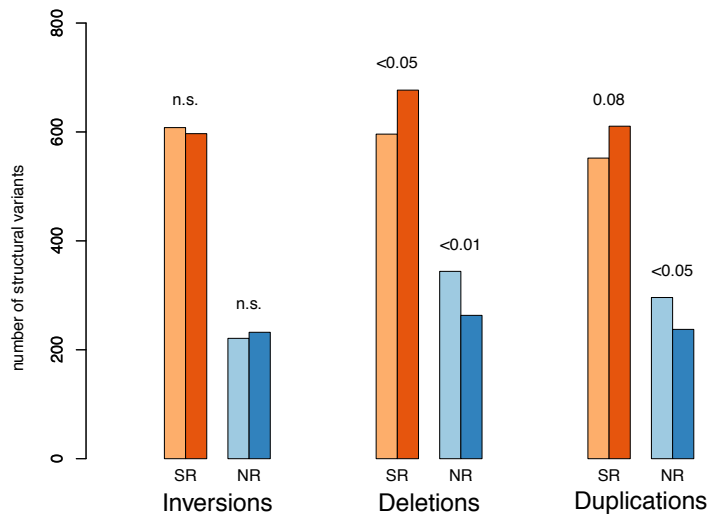
Using Pearson's Chi-squared test with Yates' continuity correction we compared observed vs. expected number of genomic elements in SR vs. NR regions. Markers in NRs appear to be located in TE-islands more often than expected, although the trend is only marginal significant ( $p=0.051$ , Table 4.2). NRs exhibit an increase in overall repeat content, caused by an increase of class I and class II transposons but not SSRs (Table 4.2). Especially the number of bases of LINE and LTR elements is higher than expected in NRs (Table 4.2). SNP and InDel content were not different between SR and NR, but heterozygous SNPs are reduced in NR, whereas homozygous SNPs are enriched in NR compared to SR.

For analyzing the influence of large structural variants (SVs) between the populations from BR and JP on recombination, we compared the numbers of SVs between SR and NR regions. 442 inversions, 2,113 deletions and 1,644 duplications were detected by Pindel. Inversions (inv) and duplications (dup) were longer than deletions (del): 25 % quantile/median/75 % quantile in basepairs: inv: 5,792/34,400/892,500; del: 838/8,201/233,600; dup: 1,113/30,300/676,600.

Markers in NR were significantly more often located in deletions and duplications than expected. In contrast, markers in SR were significantly less often located in deletions and duplications than expected (Pearson's Chi-square test with Yates' continuity correction: deletions:  $X^2=15.491$ ,  $df=1$ ,  $p<0.001$ ; duplications:  $X^2=9.061$ ,  $df=1$ ,  $p<0.01$ ), whereas no bias from the expected values was found for inversions (Figure 4.4, Supplementary Table S4.7).

**Table 4.2: Comparison between regions supported by the physical assembly (SR) and regions not supported by the physical assembly (NR) regarding different genomic features.** Number of elements and number of bases in elements were counted for 1 kb windows 500 bp up- and downstream of the markers and divided by the total number of markers in SR (2089) or NR groups (956) for normalization. Pearson's Chi-squared test with Yates' continuity correction were performed to compare observed and expected values in SR and NR.

SR	NR	X <sup>2</sup>	df	p-Value	
number of genes / kb	0.461	0.479	0.2	1	0.673
number of exons / kb	0.546	0.578	0.5	1	0.460
bases in exons / kb	156.069	155.320	1.3	1	0.258
% markers in TE islands	5.314	8.264	3.8	1	0.051
bases in repeats / kb	189.018	202.807	316.5	1	<0.001 ***
bases in classI / kb	42.488	53.826	876.5	1	<0.001 ***
bases in LINE elements / kb	17.318	23.250	578.9	1	<0.001 ***
bases in LTR elements / kb	25.260	30.383	333.2	1	<0.001 ***
bases in classII / kb	21.874	24.180	75.4	1	<0.001 ***
bases in TIR elements / kb	14.082	14.441	6.5	1	<0.05 *
bases in SSR / kb	91.990	91.381	1.4	1	0.237
number of SNPs / kb	4.786	4.929	1.3	1	0.247
number of hetSNPs / kb	0.621	0.350	49.6	1	<0.001 ***
number of homSNPs / kb	3.808	4.119	7.9	1	<0.01 **
number of indels / kb	0.429	0.368	2.9	1	0.089

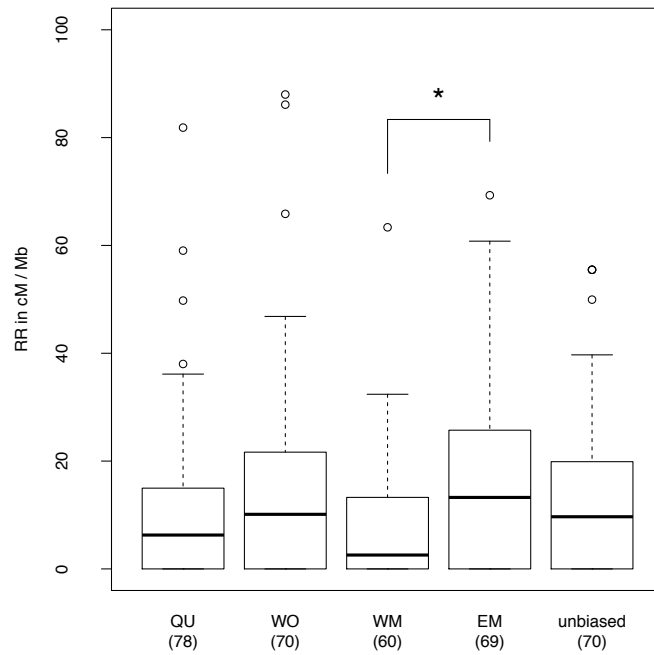


**Figure 4.4: Observed and expected number of structural variants in regions supported by the physical assembly (SR) and regions not supported by the physical assembly (NR).** Lightorange and lightblue bars show observed numbers, darkorange and darkblue bars show expected numbers. There is a significantly enrichment of deletions and duplications in NR compared to expected values, whereas no difference could be found for inversions. Values show significant levels based on Chi-square tests.



## Correlation between recombination rate and morph-biased gene expression

Overall, recombination rates were significantly different between morph-biased genes (Kruskal-Wallis rank sum test:  $X^2=12.44$ ,  $df=4$ ,  $p<0.05$ ; Figure 4.5). There was no significant difference between worker-biased genes and queen-biased genes ( $p=0.124$ ), although worker-biased genes tend to have higher recombination rates. Ergatoid-male-biased genes showed significantly increased recombination rates compared to winged-male-biased genes ( $p<0.05$ ), but were neither different from queen-biased nor from worker-biased genes. Overall differences in recombination rates in morph-biased genes do not seem to be strong in *C. obscurior*, as none of the morph-biased genes sets is significantly different from the unbiased gene set.



**Figure 4.5: Ergatoid-male-biased genes show increased recombination rates compared to winged-male-biased genes in *Cardiocondyla obscurior*.** Based on gene expression data of developing 3rd instar larvae published in Schrader *et al.* (2015) worker-biased genes (WO) show higher rates of recombination than queen-biased genes (QU), but the difference is not significant ( $p=0.124$ ). Ergatoid-male-biased genes (EM) show significantly higher recombination rates than winged-male-biased genes (WM) ( $p<0.05$ ). Overall differences in recombination rates in caste-biased genes do not seem to be strong in *C. obscurior*, as none of the caste-biased genes sets is significantly different from the unbiased gene set. Boxplots show median, interquartile ranges (IQR) and 1.5 IQR, the asterisk indicates the significant difference, all other group comparisons were not significant. Samples-sizes are given within parentheses.

## 4.5 Discussion

In this study we present a linkage map of the ant *Cardiocondyla obscurior*, consisting of 23 linkage groups and spanning 1793.9 cM. The map contains the 95 largest scaffolds, covering 96.7 % of the Cobs 1.1 assembly. The linkage map suggests that the *C. obscurior* draft genome assembly only partly succeeded in assembling repeat rich regions. Several markers assigned to different TE accumulations based on the physical assembly (Schrader *et al.*, 2014) are in fact located in close proximity to each other. Likewise, markers assigned to single TE accumulations are located between markers from low density regions (i.e. regions with low TE content) on the linkage map. In total, the marker order of the genetic map supports the draft genome assembly for 53.2 % of the physical genome size.

Overall recombination rate of *C. obscurior* (based on the 182 Mb draft genome size: 10.0 cM/Mb; based on the 85.1 Mb sequence in supported regions: 15.8 cM/Mb) falls within rates reported for other eusocial species, adding further support for elevated rates of recombination in eusocial species. We were able to determine local recombination rates for 53.2 % of the genome, revealing substantial variation in local recombination rates and 18.1 % of the genome (combined ~33 Mb) without signs of recombination. The median local recombination rate for supported regions is slightly lower (8.08 cM/Mb) than overall recombination rate of the complete map (10.0 cM/Mb). Furthermore, the enrichment of deletions and duplications in regions not supported by the assembly suggests that large structural variants impeded recombination between phenotypically divergent populations, similar to what has been found for large inversions in the genomes of *Solenopsis invicta* and *Formica selysi* ants (Wang *et al.*, 2013a; Purcell *et al.*, 2014). Thus, we hypothesize that recombination rate of *C. obscurior* could be considerably higher in intrapopulational matings, therefore the given overall recombination rate is a conservative estimate.

### Correlation of local recombination rates with genomic features

Correlations of local recombination rate with GC content and gene content were absent in *C. obscurior*. GC content and recombination rates correlate positively in other species (Fullerton *et al.*, 2001; Marsolier-Kergoat & Yeramian, 2009; Chan *et al.*, 2012; Clément & Arndt, 2013; Arbeithuber *et al.*, 2015; Wallberg *et al.*, 2015), presumably caused by the process of GC-biased gene conversion (gBGC) (Leseque *et al.*, 2013; Arbeithuber *et al.*, 2015). However, under inbreeding heterozygosity is greatly reduced and leads to a reduction of effective recombination, which consequently renders gBGC

ineffective. Thus, under high inbreeding no positive correlation between recombination rate and GC content is expected (Marais *et al.*, 2004), which has also been empirically shown in the inbred plant *Arabidopsis* (Giraut *et al.*, 2011). A negative correlation between recombination rate and gene content or gene density has been reported in *Apis* (Liu *et al.*, 2015b) and is explained by a deleterious effect of recombination on genes. In *C. obscurior* high levels of inbreeding could likewise impede a deleterious effect of recombination, as recombination will lead to novel genotypes only in rare cases.

Several GO categories are enriched in recombination “coldspots” and “hotspots”. The enrichment of conserved developmental processes in “coldspots” reflects that these are shielded from potential deleterious sequence evolution through recombination events. In contrast, regions of high recombination are enriched for genes involved with spermatocyte division, very long-chain fatty acid metabolic process and muscle development, a suite of genes expected to evolve relatively fast and to underlie only weak purifying selection.

In line with empirical studies (Boissinot *et al.*, 2001; Bartolomé *et al.*, 2002; Rizzon *et al.*, 2002; Song & Boissinot, 2007) and theoretical predictions (Dolgin & Charlesworth, 2008) local recombination rate in *C. obscurior* is negatively correlated with TE content. This can be explained either by the accumulation of TEs in regions of low or absent recombination. Alternatively, ectopic recombination between heterologous TEs is predicted to have deleterious effects (such as substantial chromosomal rearrangements) leading to selection for low rates of recombination in TE rich regions (Petrov *et al.*, 2003; Abrusán & Krambeck, 2006; Song & Boissinot, 2007). Under inbreeding and high degrees of homozygosity reduced levels of ectopic recombination are expected (Charlesworth & Charlesworth, 1995), because ectopic recombination should be more effective in removing heterozygous TE insertions under the necessary assumption that these pose fitness costs. Consequently selection against TE-insertions should be less effective in inbred species, leading to an absence of a correlation between recombination rate and TE content such as reported for the inbred plant *Arabidopsis thaliana* (Wright *et al.*, 2003). Regions with high TE content are a source for genetic variation in *C. obscurior*, comprising elevated SNV rates compared to regions with low TE content (Schrader *et al.*, 2014). The accumulation of heterozygous SNP calls in TE islands (i.e. duplicated loci collapsed in the assembly, Schrader *et al.*, 2014) suggests that selection acts against recombination of these regions to avoid the costs of ectopic recombination and demonstrates how recombination shapes TE content and *vice versa*. Recombination thus plays a major role in modulating genomic architecture in *C. obscurior* and, more generally, in driving the evolutionary potential of TEs (Santana *et al.*, 2014; Stapley *et al.*, 2015). Interestingly, we find that class I LINE

and LTR TE families are more tightly correlated with recombination rate than DNA class II TIR elements. This is likely related to the differences in transposition where class I elements are copied via an RNA intermediate and remain at their original site whereas class II DNA transposons are jumping genes and are excised from their site of origin. Consequently, if recombination is biologically linked to TE abundance and control thereof a correlation would only be expected for stationary elements such as retrotransposons. Future studies will focus in detail on genome integrity and retrotransposons as a means to generate genetic variation in the absence of effective sexual recombination under inbreeding.

The correlation of reduced levels of recombination with the presence of large scale inversions and deletions supports the hypothesis that structural rearrangements are caused by ectopic recombination events especially in TE rich regions. Subsequently, structural rearrangements between populations suppress recombination in experimental cross. Whereas the presence of deletions or inversions will likely prevent an alignment of homologous sequence, the positive correlation of duplications with recombination rate, however, likely results from elevated ectopic recombination events.

We found a positive correlation between recombination and SSR content, similar to reports for *Caenorhabditis elegans* (Duret *et al.*, 2000). Together with the positive correlations of recombination rate with homozygous SNPs and InDels, respectively, this suggests that recombination is in fact a major player in generating genetic diversity. Recombination additionally increases the fixation probability of such new alleles through Hill-Robertson-Interference (Roze & Barton, 2006), allowing for generation of selectable phenotypes under inbreeding.

## Regions not supported by the physical assembly

Marker order in the genetic and the physical map are not consistent for nearly half of the physical genome assembly. The enrichment of GO terms involved in conserved processes like reproduction, cytoskeleton and organ growth supports the idea that recombination in regions not supported by the physical assembly are indeed regions of low recombination. On the other hand the enrichment of these regions with deletions and duplications suggests that structural variants suppress homologous recombination similar to a case study in maize (Rodgers-Melnick *et al.*, 2015). Consequently, in hybrid F1 queens crossover events could have been inhibited by pairing of non-homologous sequence. The increased TE content in regions not supported by the physical assembly indicates that ectopic recombination events in the past have led to structural rearrangements between JP and BR, which now impeded recombination. The higher than

expected occurrence of homozygous SNPs in these regions suggests that here SNPs have come to fixation after divergence of the lineages, again arguing for suppressed recombination in these regions.

Chromosome number can vary within ant species (Crozier, 1969; Imai & Taylor, 1989), and different chromosome numbers have been reported for *C. obscurior* populations from Florida, USA and Okinawa, Japan (Seifert, 2003). Rates of chromosome number changes have been found to be elevated in eusocial insects compared to solitary species (Ross *et al.*, 2015), probably due to increased recombination rates under eusociality. While the karyotypes of the populations used in this study are unknown despite significant efforts, we cannot rule out that differences in chromosome numbers contributed to the lack of recombination of large parts of the genome.

## Correlation of recombination rate with morph-biased genes

High recombination in morph-biased genes is expected if genetic diversity would be causally linked to traits that putatively require diversity. Such assumptions have been made for social immunity (Ugelvig & Cremer, 2012), division of labor (Wiernasz *et al.*, 2008) and worker polymorphism (Rheindt *et al.*, 2005). Multiple mating (polyandry) and presence of multiple queens per nest (polygyny) contribute to generate offspring genetic diversity and likewise recombination rate has been suggested to be a driving factor in diversity (Liu *et al.*, 2015b). However, a simulation study predicted that recombination rate is not sufficient to contribute significantly to genetic variation (Rueppell *et al.*, 2012).

This study is the first correlating gene expression data with recombination rates in an ant. Our approach to extract vectors as a measure for morph-bias of a given gene during a specific developmental window takes into account that genes are only rarely expressed in only one morph, species or condition (Morandin *et al.*, 2015; Schrader *et al.*, 2015; Smith *et al.*, 2015). The absence of significantly increased recombination rates of worker-biased genes is in line with recent findings in *Apis* (Wallberg *et al.*, 2015) and suggests that features other than worker traits underlie elevated recombination rates in social insects. Together with elevated rates of chromosomal changes in social insects, genomic regions with high recombination may in fact be linked to adaptation and speciation in social insects which in general have drastically reduced effective population sizes  $N_e$  (Romiguier *et al.*, 2014). Recombination rate, which varies within genomic regions, in turn is negatively correlated to the level of adaptation via Hill-Robertson effects (Charlesworth, 2009). Therefore one may assume that elevated rates of recombination act as a compensatory mechanisms to account for low  $N_e$  in

both individual genomes and on a larger scale also in populations.

Higher recombination rates of ergatoid-male-biased genes compared to winged-male-biased genes add support for the view that the genome structure is shaped by Hill-Robertson effects. Differences in male-specific recombination rates may be linked to higher speed of evolutionary processes in ergatoid males under strong sexual selection (Schrempf *et al.*, 2005).

## Conclusions

Our study adds further support for exceptionally high recombination rates in eusocial insects. Our finding of elevated recombination rates in ergatoid-males-biased genes rather than worker-biased genes suggests that  $N_e$ , evolutionary novelty and level of adaptation but not specific biological functions (i.e. immunity, worker polymorphism) shape genome evolution. Moreover, high recombination rates in *C. obscurior* could lead to the accumulation of small-step mutations (SNPs, SSRs) into fixed genetic divergence between populations via Hill-Robertson-effects (Roze & Barton, 2006) and thus produce selectable phenotypes. On the other hand, low recombination rates in TE rich regions allow for fixation of variants produced by TE activity (Schrader *et al.*, 2014). Thus elevated rates of recombination as well as TE activity in low-recombining regions are suggested to act as two complementary mechanisms to drive adaptation and speciation in *C. obscurior*.

## Acknowledgments

This study was funded by a grant from the German Science Foundation (DFG He1623/31) to JH and JO. OR was additionally supported by the Regensburger Universitätsstiftung and NIGMS (R15GM102753).

## 5 General Discussion

### 5.1 Mechanisms to generate evolutionary novelties in *C. obscurior*

Research on the Myrmicine ant species *Cardiocondyla obscurior* is motivated by the occurrence of ergatoid males, a peculiarity of the genus *Cardiocondyla* (Kugler, 1983). Stricter definitions (e.g. Müller & Wagner, 1991) would not regard ergatoid *Cardiocondyla* males as an evolutionary novelty, because their characteristics such as enlarged mandibles or the absence of wings would not represent novelty without homology to other known structures *per se*. However, on the one hand broader definitions indeed include homologous structures given they provide new adaptive functions (Pigliucci, 2012) and on the other hand lifelong spermatogenesis (Heinze *et al.*, 1998) clearly is a novel trait which evolved once in the genus *Cardiocondyla*. Moreover, it is indisputable that ergatoid males beside winged males occupy a second fitness optimum in *Cardiocondyla* male-adaptive space, according to Hallgrímsson *et al.* (2012). Hence ergatoid males can safely be regarded as a major evolutionary step in the biology of *Cardiocondyla* ants.

The evolution of these wingless males thus leads to the obvious question how this evolutionary novelty has arisen. Can gradual change by mutation and selection induce novelty? Which alternative mechanisms can generate such large-step or saltatory phenotypic changes? Mechanisms to generate evolutionary novelties in general either involve the recruitment of novel genes or traits, or the co-option of existing genetic material (Pigliucci, 2012). This thesis provides evidence that both mechanisms play a role in the adaptive evolution of the ant model organism *Cardiocondyla obscurior*.

The finding of co-option of *doublesex* and other genes involved in sex differentiation into female caste and male morph determination in *C. obscurior* (Chapter 3) could be a concept of general importance for the evolution of eusociality, which emerged multiple times independently. Another mechanism to generate novelty is homologous

recombination, which is suggested to have potential for generating adaptive phenotypes. Recombination leads to chromosomal rearrangements and by interaction with transposable elements shapes the *C. obscurior* genome (Chapter 4). On the other hand, novel traits are likely introduced into *C. obscurior* biology *de novo* via microbial symbionts, which are generally known to provide novel ecological traits to their hosts (Feldhaar, 2011) eventually inducing speciation (Brucker & Bordenstein, 2012). While the potential of the mutualistic symbiont *Westeberhardia* most likely lies in nutritional upgrading during pupal stage (Chapter 2), the potential of *Wolbachia*, a second significant symbiont in *C. obscurior*, is less well-known. The current knowledge of *Wolbachia* in *C. obscurior* and its potential for novelty will be discussed in the next section.

## 5.2 Endosymbionts

### 5.2.1 *Wolbachia* - a further player influencing *C. obscurior* biology?

The alpha-proteobacterium *Wolbachia* is a prevalent symbiont in diverse arthropod hosts. *Wolbachia* phenotypes range from reproductive manipulators to mutualistic benefactors (Cordaux *et al.*, 2011; Zug & Hammerstein, 2015). While reproductive manipulators may induce speciation events (Bordenstein *et al.*, 2001), mutualistic *Wolbachia* may confer novel traits to the host (Hosokawa *et al.*, 2010). In either way, *Wolbachia* has the potential to influence host biology massively. In ants, *Wolbachia* have been found to be prevalent (Wenseleers *et al.*, 1998), but knowledge about *Wolbachia* phenotypes in ant hosts is scarce. A mutualistic association has been suggested in *Acromyrmex octospinosus* where *Wolbachia* can be found extracellular in the workers' gut lumen and in faecal droplets. Here, *Wolbachia* has been suggested to play a role in fungus garden maintenance (Andersen *et al.*, 2012). In the ant *Formica truncorum* *Wolbachia* instead has been suggested to have deleterious effects on worker fitness, as colonies with fewer infected workers produce significantly more sexuals (Wenseleers *et al.*, 2002). These examples suggest that the effects of *Wolbachia* in ant hosts are diverse, but so far no evidence for reproductive manipulation was found.

Outcrossing experiments in *C. obscurior* revealed drastically decreased lifespan and fecundity of Brazilian queens mated to allopatric males from a second Brazilian population compared to queens mated with males from their own population (Schrempf *et al.*, 2015). This effect was even stronger when crosses were performed between Brazilian queens and Japanese males (Schrempf *et al.*, 2015). Although incompatible



matings between allopatric populations are a typical sign of reproductive manipulators, *Wolbachia*-mediated outbreeding depression was excluded, because sequencing of the *Wolbachia surface protein* (*wsp*) gene revealed that all populations used for this experiment were infected by the same *Wolbachia* strain (Schrempf *et al.*, 2015). However, the analysis of gene expression data suggested that the disruption of co-evolved immunity traits underlies outbreeding depression, possibly caused by either incompatible, alien seminal fluid proteins, subtle differences in *Wolbachia* strains not detected by the *wsp* sequencing or other unknown bacteria (Schrempf *et al.*, 2015). To test for infection with different *Wolbachia* strains among *C. obscurior* populations, I performed a multi-locus-sequence typing (MLST) approach according to Baldo *et al.* (2006). This showed that the *Wolbachia* alleles of the *C. obscurior* populations from Japan ("OypB", 2011) and Brazil ("Ilhéus", 2009) differed at each of the five characteristic loci (Table 5.1). Both strains belong to *Wolbachia*-supergroup A. The OypB-*Wolbachia* strain (wObs1) is identical to strain 19 of the MLST system and has been described in other ants, e.g. *Pheidole* from Thailand (Russell *et al.*, 2009a). In contrast, no match to the Ilhéus-*Wolbachia* strain (wObs2) is deposited in the database, suggesting rapid adaptation of *Wolbachia* after introduction to the new Brazilian habitat. Alternatively, a new *Wolbachia* strain could have been acquired via horizontal transmission, e.g. via parasitoids (Heath *et al.*, 1999). Either way, the absence of a homologous *Wolbachia* strain in the MLST database calls for future studies on the origin and evolution of *Wolbachia* in Brazilian *C. obscurior*. However, the alleles used for MLST were extracted *in silico* from next-generation sequencing data. Thus it was not possible to identify potential double infection events.

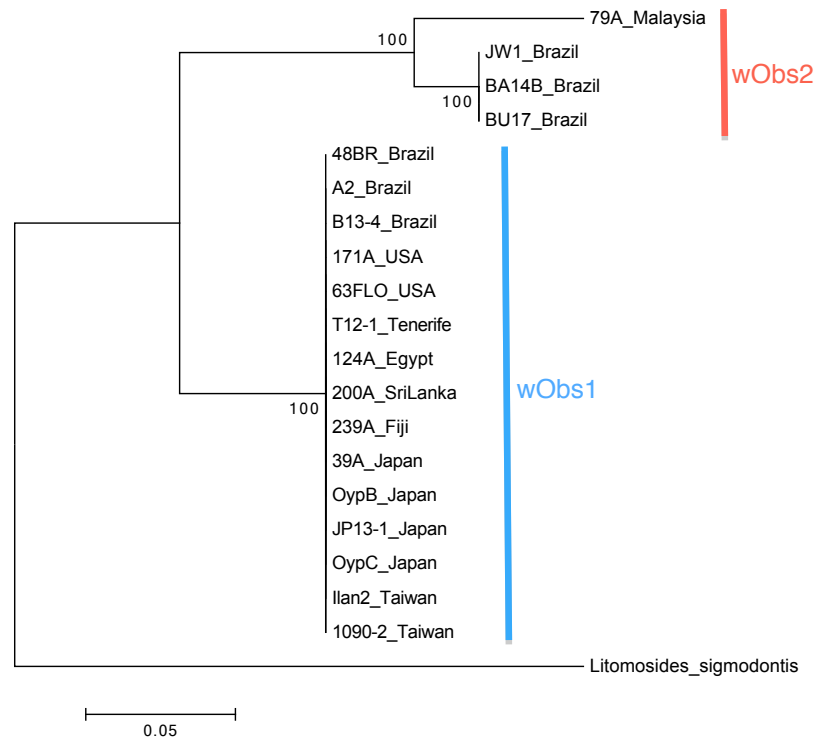
**Table 5.1: Results of the multi locus sequencing typing (MLST) approach to identify *Wolbachia* strains (Baldo *et al.*, 2006).** MLST was performed *in silico* and numbers indicate allele numbers according to the *Wolbachia*-MLST system (<http://pubmlst.org/wolbachia/>).

locus	wObs1	wObs2
	OypB, Japan	Ilhéus, Brazil
<i>gatB</i>	7	54
<i>coxA</i>	6	84
<i>hcpA</i>	7	23
<i>ftsZ</i>	3	70
<i>fbpA</i>	8	303

In order to analyze distribution of wObs1 and wObs2 among *C. obscurior*, I performed Sanger-sequencing of a 512 bp fragment of the *wsp* gene for *C. obscurior* samples from current laboratory colonies as well as preserved DNA and EtOH material (Table 5.2). This showed that wObs1 is present in *C. obscurior* from all over the world and suggests wObs1 as the ancestral strain, whereas wObs2 was only detected in samples from Brazil and in a single sample from Malaysia (Figure 5.1). In some *C. obscurior* samples from Brazil both *Wolbachia* strains were present (data not shown), supporting the hypothesis of wObs1 as the ancestral strain and rapid adaptation of *Wolbachia* after introduction into Brazil, with or without an horizontal transmission event.

**Table 5.2:** Sample overview and results of Sanger sequencing of a 512 bp fragment of the *wsp* gene.

Sample	Location (year)	Sample type	Strain
BA14B	BRAZIL: Ilhéus (2009)	Laboratory colonies	wObs2
B13-4	BRAZIL: Ilhéus (2013)	Laboratory colonies	wObs1
BU17	BRAZIL: Una, Bahia (2012)	Laboratory colonies	wObs2
JP13-1	JAPAN: Naja (2013)	Laboratory colonies	wObs1
OypB1	JAPAN: Naja (2011)	Laboratory colonies	wObs1
OypC1	JAPAN: Naja (2011)	Laboratory colonies	wObs1
Ten12-1	TENERIFE: Puerto de la Cruz los Realeios (2012)	Laboratory colonies	wObs1
79A	MALAYSIA: Ulu Gombak (2002)	DNA sample	wObs2
A2	BRAZIL: Itabuna, Bahia (2004)	ergatoid male pupae, frozen	wObs1
239A	FIJI: (2008)	DNA sample	wObs1
39A	JAPAN: Ishigaki (2005)	DNA sample	wObs1
124A	EGYPT: Talka, Elmansoga (2003)	DNA sample	wObs1
171A	USA: Lake Alfred (2004)	DNA sample	wObs1
200A	intercepted at LONDON airport in boxes with guavas from SRI LANKA	DNA sample	wObs1
48BR	BRAZIL: Itabuna, Bahia (2004)	EtOH material, pooled individuals	wObs1
63FLO	USA: Florida	EtOH material, pooled individuals	wObs1
JW1	BRAZIL: Itabuna, Bahia (2004)	Laboratory colonies	wObs2
Ilan2	TAIWAN: Northeastern-Taiwan (2014)	Laboratory colonies	wObs1
1090-2	TAIWAN: Taipeh, Academica Sinica (2014)	Laboratory colonies	wObs1

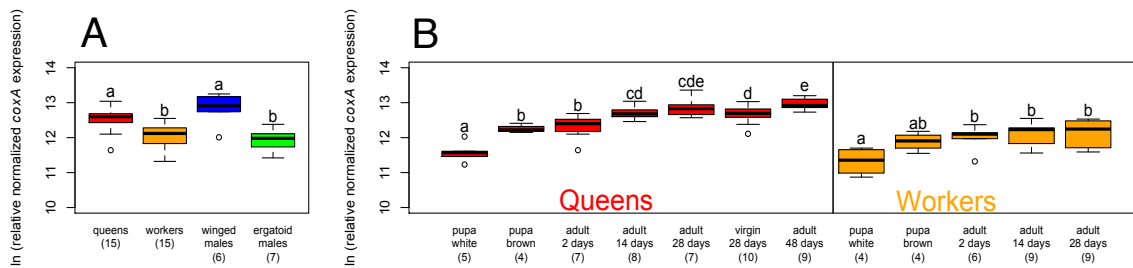


**Figure 5.1: Phylogenetic tree of a 512 bp fragment of the *Wolbachia-wsp* gene.** *Wolbachia* of the nematode *Litomosides sigmodontis* was used as outgroup. While the wObs1 strain was present in *C. obscurior* from all over the world, the wObs2 strain could only be identified in Brazilian *C. obscurior* and a closely related variant was found in a sample of Malaysia.

The *C. obscurior* populations from OypB and Ilhéus have probably been introduced by human activities and have likely experienced bottleneck effects. They differ in a range of phenotypic characters, i.e. queen number per colony, colony odor, behavior and body size (Schrader *et al.*, 2014). The finding of two distinct *Wolbachia* strains suggests that *Wolbachia* (besides *Westeberhardia*, which is absent in OypB) could act as a possible driver of divergence between the two *C. obscurior* populations. Hypothetically, *Wolbachia*-mediated incompatibilities could have preceded divergence between *C. obscurior* populations, similar to what is found in *Nasonia* (Bordenstein *et al.*, 2001). Indeed, crosses of queens from Ilhéus with males from OypB failed to produce viable brood, while the reciprocal crosses did well (A. Klein, D. Draxinger, J. Oettler, unpub.). This strongly suggests cytoplasmic incompatibility induced by the OypB-*Wolbachia* strain. Crossings with Ilhéus queens with antibiotic-treated *Wolbachia*-free OypB-males will allow testing for cytoplasmic incompatibility and a putative "rescue" of the phenotype in the future.

To assess the lifestyle of *Wolbachia* (mutualistic vs. parasitic) the growth of *Wolbachia* populations across different morphs and ages of the host could be a first indication. Consequently, I performed real-time qPCR analyses on bacterial genes specific for *Westeberhardia* and the wObs2 *Wolbachia* strain. The analyses were performed on ants from Brazil (Una, 2012) to be able to measure *Westeberhardia* and *Wolbachia* titers in one experiment. In case of *Westeberhardia*, this showed that *Westeberhardia* populations in adult workers decrease due to the degradation of bacteriomes (Chapter 2). Contrary, *Wolbachia* populations (=wObs2) in adult workers remain stable (Figure 5.2B). This suggests that *Wolbachia* strain wObs2 does not have deleterious effects on worker performance, like suggested for *Wolbachia* in *Acromyrmex* (Andersen *et al.*, 2012), but may even have mutualistic effects. If *Wolbachia* in *C. obscurior* would be harmful for worker performance, selection for low infection densities in workers would be expected.

Interestingly, the winged phenotypes (queens and winged males) show higher *Wolbachia* densities compared to the wingless phenotypes (workers and ergatoid males, Figure 5.2A). The reason for this remains elusive, but possibly higher amounts of muscle tissue with increased energy supply provide advantageous conditions for *Wolbachia* to proliferate in the winged compared to the wingless morphs.



**Figure 5.2: Dependency of *Wolbachia* densities on host-morph (A) and -age (B) as revealed by real-time quantitative PCR of the *Wolbachia* specific *coxA* gene.** (A) Queens and winged males show significantly higher *Wolbachia* titers compared to workers and ergatoid males. (B) *Wolbachia* titers show an increase during pupal development in queens and in workers, but titers in workers remain stable after eclosion, whereas titers increase with queen age. Sample sizes are given within parenthesis, letters indicate significant differences.

In summary, the preliminary results on the two distinct *Wolbachia* strains indicate a parasitic, CI inducing phenotype of wObs1 and a mutualistic phenotype of wObs2 and emphasize the diversity of the genus *Wolbachia*. Sequencing of the Ilhéus (BR) *C. obscurior* population led to the assembly of the genome of the wObs2 strain, whereas no complete genomic information of wObs1 is available. Comparative genomic analysis of both strains will further reveal their biology and their influence on host biology.

It has been shown that *Wolbachia* phenotypes can evolve rapidly from being a parasite to mutualism (Weeks *et al.*, 2007). Especially host populations introduced to new habitats and which experience genetic bottlenecks may provide great potential for rapid evolution of *Wolbachia* phenotypes towards mutualism, as the host has to adapt to new food sources and/or to new parasite and pathogen pressures. Moreover, horizontal transmission events probably are more common than previously thought (Viljakainen *et al.*, 2008; Schuler *et al.*, 2013) and will also contribute to rapid turnover of *Wolbachia* strains in introduced species. The ant *C. obscurior* is introduced to new habitats via human trading activities (Heinze *et al.*, 2006). Consequently, the *C. obscurior*-*Wolbachia* complex with its two different *Wolbachia* strains is an ideal candidate to study evolution of *Wolbachia* effects on hosts recently and repeatedly introduced to new habitats.

### 5.2.2 The tripartite interaction between *C. obscurior*, *Westeberhardia* and *Wolbachia*

The detection of two bacterial symbionts in *C. obscurior*, *Westeberhardia* and *Wolbachia*, calls for further investigations of their interaction. The presence of multiple endosymbionts in insects is well known, e.g. in grain weevils, tse tse flies or aphids (Heddi *et al.*, 1999; Gómez-Valero *et al.*, 2004; Wang *et al.*, 2013a). The complexity of such systems can reach extreme levels. For example *Planococcus citri* mealybugs host the beta-proteobacteria *Tremblaya princeps*, whose extensive genome reduction to 139 kb is probably facilitated by the acquisition of a *Tremblaya* symbiont, which itself is located in the cytoplasm of the mealybug symbiont (von Dohlen *et al.*, 2001; Husník *et al.*, 2013). Moreover, Husník *et al.* (2013) suggest that the acquisition of horizontal gene transfers (HGTs) from other organisms enabled the evolution of this three-way symbiosis. These examples illustrate the complexity of such interactions between multiple partners, including a host, a diverse range of symbionts and possibly also horizontally transferred genes originating from present or ancestral symbionts which combined contribute to the holobiont metabolism.

Co-existence of multiple symbionts in one host can lead to displacement of one symbiont. For example a secondary S symbiont partially replaces *Buchnera* in bacteriocytes in the aphid *Cinara cedri* (Gómez-Valero *et al.*, 2004). Interestingly, it was suggested that swapping of symbionts could have been caused by severe genome reduction of the primary symbiont (Pérez-Brocal *et al.*, 2006; Koga & Moran, 2014). *Westeberhardia* has a substantial reduced genome: with ~533 kb it is only about one third of the size

of free-living bacteria. However, the presence of pseudogenes indicates potential for further genome reduction, which eventually could lead to out competition of *Westeberhardia* by *Wolbachia*. Alternatively evolution may lead to a mutualistic symbiosis between two bacteria harbored in one host. For example *Wigglesworthia* provides thiamine to *Sodalis* in the tsetse fly (Snyder & Rio, 2013; Wang *et al.*, 2013a), resulting in a decline of *Sodalis* densities in the absence of *Wigglesworthia* (Wang *et al.*, 2013a). The sharing of common pathways of two symbionts in the host seems more likely than the sharing of host and symbiont pathways, as bacterial metabolites may be more compatible than bacterial and eukaryotic metabolites.

So far, no signs for a metabolic interaction between *Westeberhardia* and *Wolbachia* was found. However, a possible interaction between the two is indicated by their localization. *Westeberhardia* is located in gut-associated bacteriomes in pupae, which degrade following eclosion, and a second *Westeberhardia* population is located in the queens' ovaries for vertical transmission. Contrary to this locally restricted localization, *Wolbachia* is distributed across the whole body in adult queens, but likewise concentrates in ovary tissue (unpublished results of Fluorescence *in situ* hybridization (FISH) performed by M. Kaltenpoth, MPI Jena). The coexistence in ovary tissue is probably linked to vertical transmission of both symbionts. This co-localization could lead to cooperative metabolic interactions but also to conflict in that *Westeberhardia* and *Wolbachia* could compete for host-provided resources and/or transmission to the oocyte. It is unknown, where *Wolbachia* is located during the pupal stage but it should be further investigated if *Wolbachia* is enriched in bacteriomes, which is indicated by an increase of *Wolbachia* titers from early to late pupae (Figure 5.2B), similarly to *Westeberhardia*. This could either point towards metabolic interactions of *Wolbachia* and *Westeberhardia* in bacteriomes or alternatively to an ongoing process of outcompetition of *Westeberhardia* by *Wolbachia*. This latter explanation would fit the absence of *Westeberhardia* in the OypB (JP) host population, where *Wolbachia* titers are much higher than in the Brazilian population (A. Klein, L. Schrader, J. Oettler, unpub.), but remains highly speculative and needs to be tested in the future.

The detection of a prokaryotic xanthine-guanine phosphoribosyltransferase encoding gene incorporated into the host genome adds a further player in the metabolic network of the holobiont. It is unclear if the HGT originates from *Westeberhardia* or from a former symbiont no longer present in *C. obscurior*. Regardless, even HGTs which are relicts of former bacterial associates have been shown to take part in the holobiont metabolism (Nikoh *et al.*, 2010; Husník *et al.*, 2013). The HGT plays a central role in the nucleotide salvage in bacteria, converting xanthine and guanine to XMP and GMP, respectively. The gene is transcribed *in vivo* in the ant, which is not the case for most

horizontally transferred genes (Hotopp *et al.*, 2007) and suggests that it could play a crucial role in ancestral or extant symbiosis. Moreover, we found higher expression in larvae compared to adult queens, suggesting that the HGT plays a role for either host development or symbiont proliferation. At the moment, it is unclear if larvae already possess bacteriomes and how and when during development bacteriomes establish.

In summary, the finding of two endosymbiotic bacteria, *Westeberhardia* and *Wolbachia*, will have major impact on future work on *Cardiocondyla* ants and our interpretation of previous work on these ants. Both symbionts seem to be prevalent across the entire genus *Cardiocondyla* and comparisons of host and symbiont phylogenies may give insights into transmission routes and evolutionary history of the symbioses in the future. Naturally, it is indispensable to evaluate possible fitness effects of symbionts which are used as model organisms. Bacterial symbionts have been shown to influence host biology massively, e.g. they influence lifespan and fecundity (Prado *et al.*, 2009; Ben-Yosef *et al.*, 2014; Michalkova *et al.*, 2014). As *C. obscurior* is a model system for aging (Schrempf *et al.*, 2011), sexual selection (Schrempf *et al.*, 2005), social immunity (Ugelvig *et al.*, 2010) and adaptation (Schrader *et al.*, 2014) traits such as lifespan and fecundity are often used as fitness measures between different conditions. Likewise the study of phenotypic plasticity (Schrempf & Heinze, 2006; Schrader *et al.*, 2015) may be largely affected by symbiosis. Given that maternal transmission in social insect hosts is mainly restricted to queens, maternally inherited endosymbionts should evolve not only distortion of sex ratio but also caste ratio, with the aim to produce a queen biased caste ratio.

In the future great caution should be taken to choose *C. obscurior* populations with the same endosymbiotic background when comparing between different experimental conditions, a feat that has been largely neglected in the past. Moreover, the effects of *Westeberhardia*, the two *Wolbachia* strains wObs1 and wObs2 and the relevance of the HGT on the host should be analyzed. Schrader *et al.* (2014) showed that the two main laboratory population of *C. obscurior* from OypB (JP) and Ilhéus (BR) are distinct regarding several phenotypic traits (e.g. aggression, cuticular hydrocarbons, body size, Schrader *et al.*, 2014). It should be confirmed if these phenotypic differences are indeed differences between the BR and the JP populations resulting from divergence between host genomes (Schrader *et al.*, 2014) or if distinct endosymbiont assemblages have been the actual drivers of phenotypic divergence.

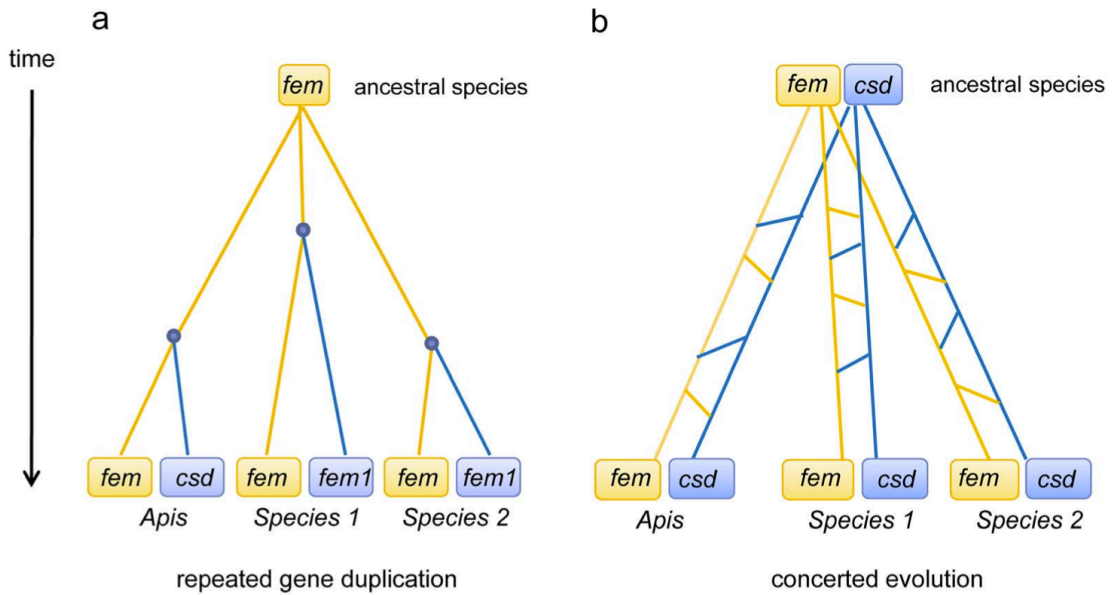
## 5.3 Sex determination in *Cardiocondyla obscurior*

### 5.3.1 Absence of single-locus sex determination in *C. obscurior*

The presence of a homolog of *Apis-csd* as primary sex determiner (Beye *et al.*, 2003) was assumed to be valid for all haplodiploid Hymenopterans, including ants, unless otherwise demonstrated (Heimpel & de Boer, 2008; Privman *et al.*, 2013). However, functional evidence is restricted to the honeybee (Hasselmann *et al.*, 2008). In support of a *csd* homolog in ants, it was suggested that the two *tra* paralogs result from an ancestral gene duplication in the common ancestor ants and bees and evolved under concerted evolution since then (Schmieder *et al.*, 2012; Privman *et al.*, 2013, Figure 5.3b). Under concerted evolution two paralogs in one species show higher sequence similarity due to gene conversion events caused by meiotic recombination than found in the orthologs across different species, despite the ancestral duplication event. However, a recent study by Koch *et al.* (2014) provides evidence in support of lineage-specific duplications in the bee, bumblebee and ant lineage, respectively (Figure 5.3a). Koch *et al.* (2014) argue that Privman *et al.* (2013) did not distinguish between alleles of the same gene (recombination events) and paralogous genes (concerted evolution). Based on this view of lineage-specific independent gene duplications (Koch *et al.*, 2014), ants do not have a true homolog of *Apis-csd*, and thus the assumption of single-locus sex determination (sl-sd) at the *csd/tra* locus in ants is highly implausible (Koch *et al.*, 2014). Nevertheless, alternative loci may act as primary sex determining signals corresponding to *csd* in *Apis* rendering sl-sd in ants still a possibility.

For *C. obscurior*, Schrempf *et al.* (2006) suggested that sl-sd is not suitable to explain sex determination. Under sl-sd high inbreeding levels due to sib-mating would lead to homozygosity at the sex-locus and thus to excessive diploid male production. Despite 10 generations of inbreeding Schrempf *et al.* (2006) detected only one diploid winged male in the seventh inbred generation out of 67 tested males using flow-cytometry. Although multi-locus sex determination with more than two loci cannot be excluded based on these data, inbreeding probably will favor the evolution of alternative mechanisms of sex determination similar to other inbred Hymenopterans (Wu *et al.*, 2005). As a first step to unravel sex determination in *C. obscurior*, we aimed to confirm involvement of the conserved genes *dsx* and *tra* in the *C. obscurior* sex determination cascade. While the presence of sex-specific isoforms of *dsx* was confirmed by this thesis (Chapter 3), the role of *tra* is not as clear. Similar to most other ant genomes the *C. obscurior* genome contains two copies of *tra*. These are located in close proximity to

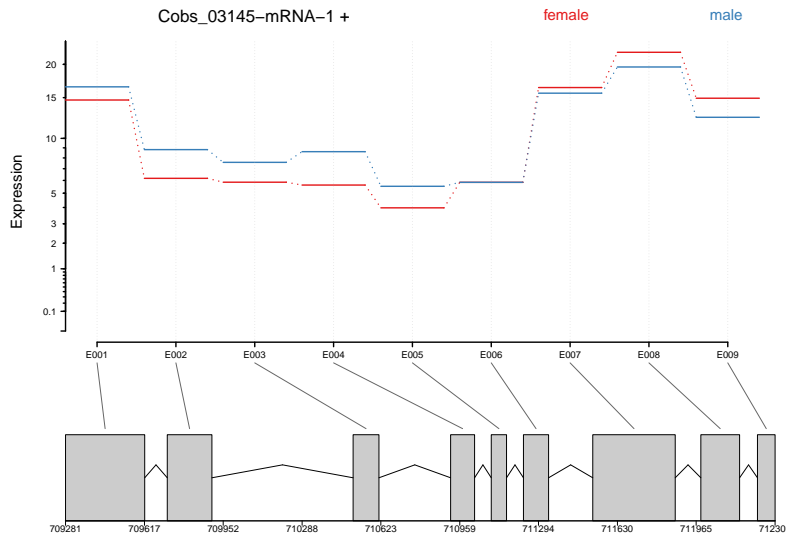




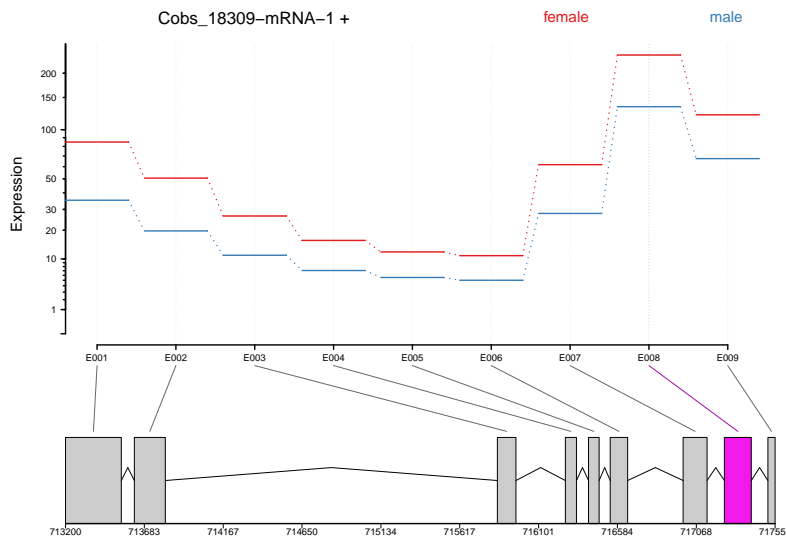
**Figure 5.3: From Koch *et al.* (2014). Two models for the evolutionary history of *fem* paralogous genes in ants and bees:** (a) repeated gene duplication and (b) concerted evolution. Points in (a) denote gene duplication events giving rise to two gene copies. Connecting lines in (b) between branches indicate concerted evolution events resulting from unequal crossing over and/or gene conversion.

each other (*tra1* and *tra2* are separated by only ~1 kb) and show extreme sequence similarity, which suggests that concerted evolution may have homogenized their sequences. Alternative splicing of *tra* seems to be a prevalent pattern across a wide range of insect species (e.g. Gempe *et al.*, 2000; Shukla & Palli, 2012b; Liu *et al.*, 2015a). However, in *C. obscurior* analyses of gene expression data of developing larvae (Schrader *et al.*, 2015) with DEXSeq (Anders *et al.*, 2012) showed that neither *tra1* nor *tra2* exhibit signs of differential splicing in males and females (Figures 5.4, 5.5), questioning the direct effect of *tra* splicing in the sex determination cascade of *C. obscurior*. Possibly, *tra* function could be mediated not by differential splicing but by different expression levels, as the gene expression data show higher expression levels in females compared to males in all exons of *tra2* (Figure 5.5). However, expression levels of *tra1* vary across the transcript (Figure 5.4). Differential expression levels without alternative splicing would point towards a mechanism of maternal imprinting similar to *Nasonia* (Verhulst *et al.*, 2010). Here, maternally derived *tra* prevents zygotic transcription in unfertilized eggs and leads to male development, whereas fertilization results in zygotic *tra* transcription and induces an autoregulatory loop that maintains the female pathway (Verhulst *et al.*, 2010). However, most likely sequence homology between *tra*-paralogs in *C. obscurior* has prevented correct mapping of RNAseq reads to the two different

copies and thus these results have to be treated with caution. Consequently, the next step in elucidating *C. obscurior* sex determination should be the modification of *tra* expression by RNAi or CRIPR/Cas and the analyses of phenotypic effects. This will help to reveal the role of *tra* in *C. obscurior* sex determination in the future.



**Figure 5.4:** Splicing pattern and expression differences between males (blue) and females (red) of *C. obscurior tra1* (=Cobs\_03145). In none of the exons significant differences of expression levels between males and females could be detected.



**Figure 5.5:** Splicing pattern and expression differences between males (blue) and females (red) of *C. obscurior tra2* (=Cobs\_18309). *tra2* exhibits higher expression levels in females vs. males over the entire transcript, but significant differences of expression levels could be only detected in one exon (pink).

### 5.3.2 Evolution of alternative sex determination mechanisms in *C. obscurior*

After Schrempf *et al.* (2006) have established that sl-sd is highly unlikely what alternative mechanisms could underlie sex determination in *C. obscurior*? For progressing towards this questions it will help to investigate the evolutionary forces shaping sex determination mechanisms. Sex determination is a fundamental process during development and thus should underlie strong purifying selection. However, the sex determination pathway evolves rapidly as demonstrated by frequent turnover of sex chromosomes (van Doorn & Kirkpatrick, 2007), transitions between genetic and environmental sex determination (Sarre *et al.*, 2004), the presence of different sex determination mechanisms even within a single species (Hamm *et al.*, 2015) or rapid molecular evolution of sex determination loci (Geuverink & Beukeboom, 2014; Eirín-López & Sánchez, 2015). Evolution of sex determination mechanisms seems to take place at the bottom of the cascade via recruitment of new upstream genes (Wilkins, 1995) while the downstream part of the cascade (*tra*, *dsx*) is subject to stronger pleiotropic constraints (nevertheless evidence for evolutionary potential in the conserved part of the cascade is present, see Chapters 1 and 3).

But how do new sex determination loci get involved in the cascade? Proximally, the recruitment of new genes, gene loss, gene duplications and neo-functionalizations provide potential for novelty. Prominent examples are *sex-lethal*, which is suggested to have evolved novel sex determining function exclusively in the genus *Drosophila* (Zhang *et al.*, 2014), *csd*, which emerged by gene duplication (Hasselmann *et al.*, 2008) or *womanizer*, which is suggested to be a novel locus recruited into the *Nasonia* sex determination cascade (Verhulst *et al.*, 2013). After a new locus has been introduced into the cascade by one of these mechanisms, selection will either lead to its maintenance or its extinction. Theory on the evolution of sex determination systems suggests that transitions may be non-adaptive, meaning that there exist selectively neutral intermediate states in which polymorphic sex determination loci determine sex (van Doorn, 2014). Evidence for this comes from species where sex is determined by multiple factors simultaneously (Orzack, *et al.*, 1980; Kozielska *et al.*, 2006), however these seem to be rare examples. It is much more likely that transitions between sex determination mechanisms are indeed adaptive. Positive selection acts either directly on the new sex determining locus or indirectly on linked loci and may lead to its fixation (Gempe *et al.*, 2010; van Doorn, 2014). Thus population genetic effects play an important role for evolution of sex determination systems. In highly inbred species, such as *C. obscurior*, selection for mechanisms other than sl-sd is expected. For example, sl-sd was shown

to be absent in the parasitoid wasp *Heterospilus prosopidis* (Braconidae) (Wu *et al.*, 2005). Under inbreeding, sl-sd will result in two identical alleles at the sex determining locus and hence will result in increased proportions of diploid males. In social insects this poses significant costs to the colony (Whitehorn *et al.*, 2009; van Wilgenburg *et al.*, 2006). The evolution of the ergatoid male morph in the genus *Cardiocondyla* was accompanied by the transition to high levels of inbreeding because of local mate competition (LMC) (Hamilton, 1967; Cremer & Heinze, 2002). While sl-sd could have been present before the evolution of the ergatoid males, excessive diploid male production through inbreeding has probably selected for mechanisms other than sl-sd during the evolution of the ergatoid male morph in the genus *Cardiocondyla*.

Another selective force may be the interaction of sex determination mechanisms with selfish genetic elements which can influence sex ratios (e.g. endosymbionts favoring female biased sex ratios). Sex ratio distorters may preserve selection on sex ratios under otherwise stable population dynamics and thus lead to counter-adaptations of the sex-determining system (Werren & Beukeboom, 1998). The high prevalence of maternally inherited bacteria throughout the insect clade could be one factor leading to rapid turnover of sex determination mechanisms in insects. The detection of prevalent bacterial symbionts in *C. obscurior* in this thesis adds additional support for high selection pressures on the *C. obscurior* sex determination system in the past.

Alternatively to direct selection pressures on the new sex determining locus itself, strong linkage with loci under strong positive selection can explain evolution of sex determination systems (van Doorn & Kirkpatrick, 2010). The *C. obscurior* genome is enriched with transposable elements (TEs) (Schrader *et al.*, 2014). TEs are candidates for generating novel genes (Kaessmann, 2010), e.g. via exon shuffling (Moran *et al.*, 1999) or hitchhiking of transcription factor binding sites (Wagner & Lynch, 2010) and could thus provide the raw material for new sex determination loci.

Contrary to direct selection pressures described above, indirect selection pressures may also drive evolution of sex determination mechanisms. In sexually dimorphic species, genes are predicted to have different fitness optima for males and females, e.g. females may benefit from slower development to build up lasting resources for reproduction, whereas males might benefit from a fast development. A gene with sexually-antagonistic fitness optima (e.g. a gene accelerating development) linked to a newly emerging sex determining locus (e.g. a masculinizing factor) might consequently lead to the fixation of the new masculinizing factor by providing higher fitness to the male compared to female carriers. This may lead eventually to the evolution of sex chromosomes. The evolution of a new male morph in *C. obscurior* (the ergatoid male) was

probably accompanied by the evolution of new sexually antagonistic loci. For example, ergatoid males benefit from a fast growth rate to emerge before their rivals, kill them and monopolize mating with virgin queens in the nest (Kugler, 1983; Heinze *et al.*, 1998). Selection pressures on a fast development were not the present in the ancestral winged male morph. Consequently, before the occurrence of ergatoid males, no antagonistic selection pressures on divergent developmental rates were present between males and females. Now that the ergatoid male morph is the prevalent male morph in *C. obscurior*, selection on a fast development in males contradicts a slow development in females for reproduction in queens and a strong work-force in workers. Thus the evolution of ergatoid males may have driven the recruitment of new sex determining genes via indirect selection pressures.

Theory on sex determination evolution is mainly restricted to transitions between genetic and environment sex determination systems or theory on sex chromosome evolution. Theory on evolution of the pathway in social species taking the queen/worker diphenism into account is scarce, if not completely lacking. Eusociality will influence evolution of sex determination, as female diphenism will lead to two distinct fitness optima in females (mating and reproduction in queens vs. foraging and nursing in workers), which is a major constraint for establishing genes beneficial for all females but not for males and vice versa.

In summary, the underlying mechanism of *C. obscurior* sex determination remain unclear. However, several new players which influence the evolution of sex determination in *C. obscurior* have been identified. The *C. obscurior* lifestyle may encourage genetic drift events that may allow random mutations to be established without fitness benefits. Most importantly, the evolution of the ergatoid male morph is probably a major precursor for a new sex determining system, as high levels of inbreeding will have selected for mechanisms other than sl-sd. Additionally, the *C. obscurior* hologenome harbors genomic elements (endosymbionts and TEs) that are known to accelerate evolution of sex determination pathways and that could have led to a rapid evolution of this important pathway.

## 5.4 Supressed recombination in supergenes

An important mechanism to generate evolutionary novelty by mixing of existing genetic material is homologous recombination during meiosis. Despite the twofold "cost of meiosis", which is on the one hand the reduction of parent-offspring relatedness and on the other hand the disruption of co-adapted loci, the benefits of sexual reproduction and associated recombination during meiosis obviously exceed its costs (Rice, 2002). Experimental evidence for the adaptive significance of sexual recombination comes from a diverse range of organisms (viruses to flies) and suggests that recombination is adaptive by decreasing the accumulation of harmful mutations as well as by increasing the accumulation of beneficial mutations (Rice, 2002). However, local recombination rates across genomes vary substantially, with hotspots showing elevated rates of recombination as well as coldspots where recombination is completely suppressed (e.g. Baudat *et al.*, 2010). The linkage map of *C. obscurior* has revealed large regions where recombination is completely absent. Analysis of paired-end sequencing data, which allows to reconstruct genomic architectures, suggests that chromosomal rearrangements between the Brazilian and Japanese *C. obscurior* populations could have impeded recombination in the interpopulational cross (Chapter 4).

Recently, more and more examples across a diversity of species for regions with suppressed recombination have been associated with accumulations of co-adapted gene complexes, so-called "*supergenes*", for example in plants, ants and butterflies (Chu *et al.*, 2011; Wang *et al.*, 2013b; Kunte, 2014; Purcell *et al.*, 2014). These supergenes comprise co-segregating loci which are inherited together and express a variety of phenotypic and ecological traits (Schwander *et al.*, 2014). For example in two ants species, *Solenopsis invicta* and *Formica selysi*, supergenes have been suggested to underlie population differences in social form, i.e. phenotypic differences in queen number, aggression, queen and worker body size and fecundity (Wang *et al.*, 2013b; Purcell *et al.*, 2014). The absence of homology in the underlying genetic architecture between *Solenopsis* and *Formica* suggests converging evolution rather than a common origin of these supergenes (Purcell *et al.*, 2014).

What leads to the repeated emergence of supergenes across species? Schwander *et al.* (2014) suggest three mechanism underlying the evolution of supergenes: mutations leading to beneficial interactions between loci, duplications generating novel genes that interact with nearby loci and translocations leading to an accumulation of interacting genes. To maintain linkage between the loci of a supergene reduced recombination is required. This can be realized either by a location near the centromers, by epigenetic

modifications or by structural differences, such as large-scale inversions (Schwander *et al.*, 2014). For example, in butterflies supergenes responsible for wing pattern mimicry are located in inversions across different butterfly species (Joron *et al.*, 2011; Kunte, 2014; Ito *et al.*, 2015). Inversions can be introduced by ectopic recombination events of non-homologous DNA sequence. Gene duplications are suitable candidates for the induction of ectopic recombination events. The fact that gene duplications on the one hand have been shown to be responsible for supergene-emergence (e.g. in case of the MHC complex, Schwander *et al.*, 2014), and on the other hand may also be responsible for generating chromosomal rearrangement via induction of ectopic recombination events, could explain why several supergenes with diverging genetic architecture between populations are found. Gene duplications are a common mechanism of genome evolution in ants, exemplified by duplications in important gene families like desaturases, vitellogenins and chemosensory protein genes (Kulmuni *et al.*, 2013; Morandin *et al.*, 2014; Helmkampf *et al.*, 2015). Moreover high rates of turnover of chromosome numbers indicate rapid chromosomal rearrangements in social insects (Ross *et al.*, 2015) and suggest that the two examples in *Solenopsis* and *Formica* will likely be followed by other cases.

In *C. obscurior*, it is unclear which genes are located in the regions of suppressed recombination identified in Chapter 4 and whether they play a role in phenotypic divergence between the Brazilian and the Japanese population. Several gene families have been identified that show significant divergence between the two populations, such as genes involved in fatty acid synthesis, possibly involved in producing the detected distinct patterns of cuticular hydrocarbons between the populations (Schrader *et al.*, 2014, Draxinger & Oettler, unpub.). Although the genome assembly Cobs 1.4 has located the vast majority of these genes in “TE-islands” (regions with increased TE content compared to the rest of the genome, Schrader *et al.*, 2014), their exact genomic localization and possible accumulation in supergenes is unclear at this stage. The linkage map could not confirm the integrity of the TE-islands, but instead suggests that the genome assembler faced substantial problems in assembling repeat-rich regions. Consequently, regions of low recombination reported in the linkage map provide candidates for putative supergenes in *C. obscurior*.

## 5.5 Major challenges for the future of *C. obscurior* research

The improvement of the *C. obscurior* genome using the linkage map and sequencing of long reads will concentrate on assembling the repeat-rich regions to further test the integrity of the TE-islands and to gain deeper insights into the mechanistic basis of adaptation. Furthermore, the study of structural variants between the BR and JP population will allow for identification of regions under selection using a population genomics approach. A major drawback for *C. obscurior* research is the lack of functional genomics which hinders testing gene function by silencing or overexpression. The establishment of methods like RNAi or the CRISPR/Cas-system will complete the genomic toolkit for the model *C. obscurior*.



## 6 Summary

It is generally discussed if evolutionary novelties are caused by novel genes or by co-option of existing genes and pathways. This thesis highlights *de novo* acquisition of novelty via endosymbiosis as well as novelty by combination of existing material through co-option of genes and by homologous recombination of the genome.

**Chapter 1** summarizes the present state of the art in the study of evolutionary novelty. The chapter introduces into the holobiont concept and the potential of endosymbiotic bacteria for host evolution. Moreover it gives an overview of sex determination systems in insects and an introduction into the conserved sex determination gene *doublesex*, which is investigated in Chapter 3 in detail. Furthermore, an overview of meiotic recombination and genetic linkage mapping is given, thereby providing an introduction for Chapter 4. The Chapter concludes with the introduction into the model organism *Cardiocondyla obscurior* and the ergatoid male morph, an evolutionary novelty of the genus *Cardiocondyla*.

In **Chapter 2** a novel endosymbiotic bacterium, "*Candidatus* Westeberhardia cardiocondylae", and its interaction with the host *C. obscurior* is described. Severe genome reduction and the localization in bacteriomes during the pupal stage strongly suggest a mutualistic nature of the symbiosis. Metabolic pathway analysis of the streamlined *Westerberhardia* genome together with increasing *Westerberhardia* titers in pupae suggest that *Westerberhardia* provides the host with the tyrosine-precursor 4-hydroxyphenylpyruvate, which may be important for cuticle buildup during development. Thus, *Westerberhardia* putatively provides novel traits to the host, thereby facilitating adaptation of *C. obscurior* to novel, arboreal and nutrient-poor habitats. The detection of a *Westerberhardia*-free host population will allow for elucidating the symbiosis further in the future.

The study presented in **Chapter 3** introduces the concept of co-option to explain the mechanistic basis of polyphenism in social insects. We show that *doublesex* and other sex-specifically expressed genes are co-opted into female caste and male morph differentiation of *C. obscurior*. We hypothesize that this principle is a general pattern

of caste-differentiation in eusocial insects and thus of one of the major transitions in evolution.

**Chapter 4** addresses homologous recombination in *C. obscurior*. Recombination of existing loci is able to generate new genotypes and thus new selectable phenotypes much more rapidly than mutations, due to Hill-Robertson-Interference. During this study it became clear that the interaction between homologous recombination and transposable elements could have led to chromosomal rearrangements within the two studied *C. obscurior* populations. While recombination is impeded in regions with high abundance of TEs, regions of high local recombination rates show high levels of genetic divergence between the Brazilian and the Japanese *C. obscurior* population. Thus recombination is an important player in shaping the *C. obscurior* genome and likely drives divergence between the two *C. obscurior* populations.

In **Chapter 5** the analysis of the *C. obscurior* holobiont is pursued and discussed in detail. In addition to *Westeberhardia*, the host *C. obscurior* also harbors *Wolbachia*. An overview of the current knowledge about *Wolbachia* in *C. obscurior* is given, the potential for possible symbiont-symbiont interaction in the host is discussed and the importance of *C. obscurior* endosymbionts in the study of the host species is pointed out. Furthermore, an overview of *C. obscurior* sex determination and the evolution of sex determination systems is given, with a special emphasis on social insects and *C. obscurior*. Lastly, the evolutionary significance of recombination in this species is discussed in the light of recent findings of supergenes (linked loci with low recombination) in other species. Finally, an outlook on future research on *C. obscurior* is provided.

## 7 Zusammenfassung

Eine wichtige Fragestellung der Evolutionsbiologie ist es, ob evolutionäre Neuerungen durch die Entstehung neuer Gene oder durch die Ko-option (Wiederverwendung) von existierenden Genen und Genkaskaden realisiert werden. Diese Arbeit stellt die Bedeutung beider Mechanismen anhand der Ameise *C. obscurior* dar. Einerseits bietet Endosymbiose mit Bakterien Potential für die Entstehung neuer Gene und Merkmale im Rahmen des Konzept des "Holobionten", während andererseits das Potential für Neuerungen durch die Koooption von existierenden Genen und durch die Durchmischung des Genoms durch homologe Rekombination aufgezeigt wird.

**Kapitel 1** gibt einen Überblick über den aktuellen Forschungsstand über evolutionäre Neuerungen. Das Kapitel stellt das "Holobiont"-Konzept vor und erklärt das Potential von endosymbiotischen Bakterien für die Evolution des Wirtes. Weiterhin wird ein Überblick über verschiedene Geschlechtsbestimmungssysteme in Insekten gegeben und das Gen *doublesex* vorgestellt, welches in Kapitel 3 im Detail analysiert wird. Darüber hinaus wird meiotische Rekombination und das Prinzip der Genkartierung beschrieben, welches in Kapitel 4 behandelt und angewandt wird. Die Einleitung endet mit einer Einführung in den Modelorganismus *Cardiocondyla obscurior* und die ergatoide Männchen Morphe, eine evolutionäre Neuerung in der Gattung *Cardiocondyla*.

In **Kapitel 2** wird ein neu entdecktes, endosymbiotisches Bakterium ("*Candidatus Westeberhardia cardiocondylae*") und dessen Symbiose mit dem Wirt *C. obscurior* beschrieben. Das stark reduzierte Genom von *Westerberhardia* und die Lokalisierung des Symbionten in sog. "Bakteriomen" während der Puppenphase sprechen für eine mutualistische Assoziation mit dem Wirt. Die Analyse der Stoffwechselwege des reduzierten *Westerberhardia* Genoms und ein Anstieg der Bakteriendichte während der Puppenphase deuten darauf hin, dass *Westerberhardia* den Wirt mit der Tyrosin Vorstufe 4-Hydroxyphenylpyruvat versorgen könnte, das für den Aufbau der Kutikula während der Puppenphase wichtig sein könnte. Folglich, könnte *Westerberhardia* es dem Wirt ermöglichen, neue Merkmale zu evolvieren und sich an neue, nährstoffarme Habitate anzupassen. Die Entdeckung einer *Westerberhardia*-freien Wirtspopulation bietet ideale

Voraussetzungen um in Zukunft die Symbiose zwischen *C. obscurior* und *Westeberhardia* genauer zu untersuchen.

Die in **Kapitel 3** präsentierte Studie führt das Konzept der "Koooption" zur Erklärung der Mechanismen bei der Entstehung von Polyphenismen bei sozialen Insekten ein. Die Studie zeigt, dass das Gen *doublesex* neben anderen geschlechtsspezifisch exprimierten Genen kooptiert wurde, um in *C. obscurior* die Kastendifferenzierung bei Weibchen und die Morphendifferenzierung der Männchen zu regulieren. Vermutlich könnte dieses Prinzip der Koooption ein genereller Mechanismus bei der Entstehung von Kasten bei sozialen Insekten sein, und damit an einer der wichtigsten evolutionären Neuerungen beteiligt sein.

**Kapitel 4** behandelt homologe Rekombination in *C. obscurior*. Durch die "Hill-Robertson-Interferenz" ermöglicht es Rekombination von existierendem genetischen Material viel schneller neue Genotypen und folglich neue selektierbare Phänotypen zu generieren als zufällige Mutationen. Während dieser Arbeit zeigte sich eine Interaktion von homologer Rekombination mit transposablen Elementen (TE). Diese Interaktion könnte zu chromosomalen Umlagerungen zwischen zwei *C. obscurior* Populationen aus Brasilien und Japan geführt haben. Während Rekombination in Regionen mit hohem TE Gehalt niedrig ist, findet sich in Regionen hoher Rekombination eine erhöhte genetische Divergenz zwischen den beiden *C. obscurior* Populationen aus Brasilien und Japan. Damit ist Rekombination ein wichtiger Faktor in der Genom-Evolution von *C. obscurior*.

In **Kapitel 5** wird die Analyse des *C. obscurior* Holobionten fortgesetzt und im Detail diskutiert. Zusätzlich zu *Westeberhardia* findet sich im Wirt *C. obscurior* das Bakterium *Wolbachia*. Nachdem ein Überblick über *Wolbachia* in *C. obscurior* gegeben wurde, wird das Potential für Interaktionen zwischen den beiden Symbionten in der Ameise diskutiert und die Bedeutung der Endosymbionten für das Studium der Wirts-Art herausgearbeitet. Weiterhin wird die Geschlechtsbestimmung in *C. obscurior* diskutiert, wobei die evolutionären Faktoren welche die Geschlechtsbestimmung von *C. obscurior* beeinflusst haben könnten, hervorgehoben werden. Schließlich wird die Bedeutung von Rekombination im Hinblick auf die Beschreibung von sog. "Supergenen" in anderen Arten diskutiert. Abschließend wird ein Ausblick auf die zukünftige Forschung am Modelorganismus *C. obscurior* gegeben.



## Supplementary Tables

**Table S2.1:** Pairwise T-tests using post-hoc Benjamini-Hochberg correction for intrapopulational comparisons of *Westeberhardia* titers measured by qPCR in the BR Una (2012) population (Fig. 2.3B).

	L	PP	PW
PP	0.5894	-	-
PW	0.2080	0.3647	-
PB	0.1031	0.0286	0.0033

**Table S2.2:** Pairwise Mann-Whitney U-tests using post-hoc Benjamini-Hochberg correction for morph dependence of *Westeberhardia* titers measured by qPCR in BR (Una 2012) individuals (Fig.2.3E).

	Q	W	WM
W	0.00012	-	-
WM	0.00014	0.75027	-
M	0.00014	0.21678	0.21678

**Table S2.3:** Pairwise t-tests using post-hoc Benjamini-Hochberg correction for *Westeberhardia* titers measured by qPCR in BR (Una 2012) workers (Fig. 2.3F).

	PW	PB	W2	W14
PB	0.1233	-	-	-
W2	0.0868	0.9135	-	-
W14	0.9135	0.0868	0.0575	-
W28	0.1969	0.0071	0.0046	0.1774

**Table S2.4:** Pairwise Mann-Whitney U-tests using post-hoc Benjamini-Hochberg correction for *Westeberhardia* titers measured by qPCR in BR (Una 2012) queens (Fig. 2.3F).

	PW	PB	Q2	Q14	Q28	V28
PB	0.0222	-	-	-	-	-
Q2	0.0082	0.3524	-	-	-	-
Q14	0.0047	0.0082	0.1192	-	-	-
Q28	0.0065	0.0098	0.0047	0.0029	-	-
V28	0.0082	0.0315	0.0188	0.0329	0.2029	-
Q48	0.0047	0.0065	0.0028	0.0017	0.0363	0.0047

**Table S2.5:** Raw data for intrapopulation infection densities (Figure2.3 C+D).

population	colony	Workers			Queens		
		number infected	number tested	proportion	number infected	number tested	proportion
BR Ilhéus 2013	1	1	10	0.10	10	10	1.00
	2	3	10	0.30	10	10	1.00
	3	4	10	0.40	10	10	1.00
	4	3	10	0.30	10	10	1.00
	5	6	10	0.60	10	10	1.00
	6	5	10	0.50	7	7	1.00
	7	8	10	0.80	6	6	1.00
	8	4	10	0.40	7	8	0.88
total		34	80	0.43	70	71	0.99
BR Ilhéus 2009	1	10	10	1.00	10	10	1.00
	2	10	10	1.00	10	10	1.00
	3	8	10	0.80	10	10	1.00
	4	8	10	0.80	9	10	1.00
	5	9	10	0.90	10	10	1.00
	6	10	10	1.00	10	10	1.00
	7	7	10	0.70	9	9	1.00
	8	7	9	0.78	10	10	1.00
total		69	79	0.87	78	79	0.99
BR Una 2012	1	3	10	0.30	10	10	1.00
	2	6	10	0.60	10	10	1.00
	3	5	10	0.50	10	10	1.00
	4	7	10	0.70	10	10	1.00
	5	5	10	0.50	10	10	1.00
	6	4	10	0.40	10	10	1.00
	7	5	10	0.50	10	10	1.00
	8	5	10	0.50	10	10	1.00
total		40	80	0.50	80	80	1.00
JP OypU 2013	1	1	10	0.10	10	10	1.00
	2	1	10	0.10	4	10	0.40
	3	3	10	0.30	10	10	1.00
	4	5	10	0.50	9	10	0.90
	5	10	10	1.00	9	10	0.90
	6	6	10	0.60	9	9	1.00
	7	9	10	0.90	10	10	1.00
	8	5	10	0.50	9	10	0.90
total		40	80	0.50	70	79	0.89
1		0	10	0.00	0	10	0.00

population	colony	Workers			Queens		
		number infected	number tested	proportion	number infected	number tested	proportion
JP OypB 2011	2	0	10	0.00	0	10	0.00
	3	0	10	0.00	1	10	0.10
	4	0	10	0.00	1	10	0.10
	5	0	10	0.00	0	10	0.00
	6	1	10	0.10	0	10	0.00
	<b>total</b>	<b>1</b>	<b>60</b>	<b>0.02</b>	<b>2</b>	<b>60</b>	<b>0.03</b>
JP OypC 2011	1	10	10	1.00	10	10	1.00
	2	9	10	0.90	10	10	1.00
	3	10	10	1.00	10	10	1.00
	4	10	10	1.00	10	10	1.00
	5	9	10	0.90	10	10	1.00
	6	10	10	1.00	8	8	1.00
	7	9	10	0.90	10	10	1.00
	8	10	10	1.00	N/A	N/A	N/A
	<b>total</b>	<b>77</b>	<b>80</b>	<b>0.96</b>	<b>68</b>	<b>68</b>	<b>1.00</b>
SP Ten 2012/13	1	7	9	0.78	10	10	1.00
	2	4	10	0.40	9	10	0.90
	3	2	10	0.20	10	10	1.00
	4	3	10	0.30	10	10	1.00
	5	6	10	0.60	10	10	1.00
	6	7	10	0.70	10	10	1.00
	7	3	10	0.30	10	10	1.00
	8	3	10	0.30	10	10	1.00
	<b>total</b>	<b>35</b>	<b>79</b>	<b>0.44</b>	<b>79</b>	<b>80</b>	<b>0.99</b>
<b>total</b>		<b>296</b>	<b>538</b>	<b>0.55</b>	<b>447</b>	<b>517</b>	<b>0.86</b>

## Supplementary Methods

### Phylogenomic reconstruction

For phylogenetic placement of *Westeberhardia*, we used translated CDS sequences of *Westeberhardia* and followed the approach of (Husník *et al.*, 2011). We constructed ortholog clusters using the proteomes from different Gammaproteobacteria (Table S2.6) on a standalone version of OrthoMCL v2.0.9 (Chen *et al.*, 2007) and identified 64 single-copy core protein clusters out of the 69 ones identified by the aforementioned study. These were then aligned using mafft v7.123b (Katoh & Standley, 2013). Alignments were refined using Gblocks v0.91b (Talavera & Castresana, 2007) (Supplementary file



protein\_sequences.fasta). Dayhoff6 recoding and phylogenetic reconstruction was done using PhyloBayes v3.3f (Lartillot *et al.*, 2009). The chains ran for 16243 generations and a burn in of 6000 was chosen. Both bipartition and summary variables were  $\leq 0.3$ , and all effective sizes of all summary variables were higher than 100.

**Table S2.6:** Table showing accession numbers of protein sequences used for phylogenetic tree reconstruction

abb.	species	strain	accession number
Xaxo	Xanthomonas axonopodis	Xac29-1	CP004399.1, CP004400.1, CP004401.1, CP004402.1
Paer	Pseudomonas aeruginosa	PAO1	AE004091.2
Vcho	Vibrio cholerae	O1 biovar eltor str. N16961	AE003852.1, AE003853.1
Pmul	Pasteurella multocida	subsp. multocida str. Pm70	AE004439.1
Hinf	Haemophilus influenzae	Rd KW20	L42023.1
Hduc	Haemophilus ducreyi	35000HP	AE017143.1
Eict	Edwardsiella ictaluri	93-146	CP001600.2
Etar	Edwardsiella tarda	EIB202	CP001135.1
Db11	Serratia marcescens marcescens	Db11	HG326223.1
SsCt	Serratia symbiotica	SCt-VLC	FR904230.1, FR904231.1, FR904232.1, FR904233.1, FR904234.1, FR904235.1, FR904236.1, FR904237.1, FR904238.1, FR904239.1, FR904240.1, FR904241.1, FR904242.1, FR904243.1, FR904244.1, FR904245.1, FR904246.1, FR904247.1, FR904248.1, HG934887.1, HG934888.1, HG934889.1
SsCc	Serratia symbiotica	Cinara cedri	CP002295.1
SAS9	Serratia plymuthica	AS9	CP002773.1
S568	Serratia proteamaculans	568	CP000826.1, CP000827.1
C092	Yersinia pestis	C092	AL590842.1, AL109969.1, AL117189.1, AL117211.1
RLSR	Regiella insecticola	LSR1	GL379589.1:GL379760.1, CM000957.1
H5AT	Hamiltonella defensa	5AT	CP001277.1, CP001278.1
Plau	Photorhabdus luminescens	subsp. laumondii TTO1	BX470251.1
Pasy	Photorhabdus asymbiotica	ATCC43949	FM162591.1, FM162592.1
Xnem	Xenorhabdus nematophila	ATCC 19061	FN667742.1, FN667743.1
Xbov	Xenorhabdus bovienii	SS-2004	FN667741.1
Pmir	Proteus mirabilis	HI4320	AM942759.1, AM942760.1
Anas	Arsenophonus nasoniae	DSM 15247	AUCC01000001.1:AUCC01000302.1
Rped	Riesia pediculicola	USDA	CP001085.1, CP001086.1
Pstu	Providencia stuartii	MRSN 2154	CP003488.1
Pret	Providencia rettgeri	Dmell	AJSB01000001.1:AJSB01000009.1, CM001856.1
SgMo	Sodalis glossinidius	morsitans	AP008232.1, AP008233.1, AP008234.1, AP008235.1
SHS1	Sodalis sp.	HS1	CP006569.1, CP006570.1
SSOP	Sodalis pierantonius	SOPE	CP006568.1
BcHc	Baumannia cicadellinicola	Hc	CP000238.1
Bpen	Blochmannia pennsylvanicus	BPEN	CP000016.1
Bflo	Blochmannia floridanus	NULL	BX248583.1

**Table S2.6: (continued)** Table showing accession numbers of protein sequences used for phylogenetic tree reconstruction

abb.	species	strain	accession number
Wglo	Wigglesworthia glossinidia	brevipalpis	BA000021.3, AB063523.1
Wcar	Westeberhardia cardiocondylae	STRAINUNK	ACCNUM
Pcar	Pectobacterium carotovorum	subsp. carotovorum PC1	CP001657.1
Dzea	Dickeya zeae	Ech1591	CP001655.1
Ddad	Dickeya dadantii	Ech703	CP001654.1
Csak	Cronobacter sakazakii	ATCC BAA-894	CP000783.1, CP000784.1, CP000785.1
Ctur	Cronobacter turicensis	z3032	FN543093.2, FN543094.1, FN543095.1, FN543096.1
MG16	Escherichia coli	K-12 substr. MG1655	U00096.3
Ckos	Citrobacter koseri	ATCC BAA-895	CP000822.1, CP000823.1, CP000824.1
Sent	Salmonella enterica subsp.	LT2	AE006468.1, AE006471.1
Sent	enterica serovar Typhimurium str.	LT2	AE006468.1, AE006471.1
E638	Enterobacter sp.	638	CP000653.1, CP000654.1
Eclo	Enterobacter cloacae	subsp. cloacae ATCC 13047	CP000654.1, CP001918.1, CP001919.1, CP001920.1
Kpne	Klebsiella pneumoniae	subsp. pneumoniae HS11286	CP003200.1, CP003223.1, CP003224.1, CP003225.1, CP003226.1, CP003227.1, CP003228.1
Kvar	Klebsiella variicola	At-22	CP001891.1
Icap	Ishikawaella capsulata	Mpkobe	AP010872.1, AP010873.1
BAsps	Buchnera aphidicola	APS	A000003.2, AP001071.1, AP001070.1
BSgr	Buchnera aphidicola	Sg	AE013218.1
BBpi	Buchnera aphidicola	Bp	AE016826.1, AF492591.1
BCce	Buchnera aphidicola	BCc	CP000263.1, AY438025.1
Pana	Pantoea ananatis	PA13	CP003085.1, CP003086.1
Ebil	Erwinia billingiae	Eb661	FP236843.1, FP236826.1, FP236830.1
Etas	Erwinia tasmaniensis	ET1/99	CU468135.1, CU468128.1, CU468130.1, CU468131.1, CU468132.1, CU468133.1
Epyr	Erwinia pyrifoliae	Ep1/96	FP236842.1, FP928999.1, FP236827.1 FP236828.1, FP236829.1

## PCR assay on *nrdB1* on worldwide collected samples and sequencing of 16S rDNA gene

To assess infection presence of *Westeberhardia* in *C. obscurior* across different populations, we performed a diagnostic PCR assay on material collected worldwide (Table 2.1). We used DNA material from a previous study (Oettler *et al.*, 2010), and extracted DNA from additional samples using a chloroform-based method (Sambrook & Russell, 2001). We performed PCR on the *nrdB1* (*ribonucleoside-diphosphate reductase 1 subunit beta*) gene of *Westeberhardia* (WEOB\_403) (*nrdB1*for: 5'-GGAAGGAGT-CCTAATGTTGCG-3' and *nrdB1*rev: 5'-ACCAGAAATATCTTTTGCACGTT-3'), using the ant housekeeping gene *elongation factor 1-alpha 1* (Cobs\_01649) (*EF1*for:

5'-TCACTGGTACCTCGCAAGCCGA-3', *EF1* rev: 5'-AGCGTGCTCACGAGTTTGTCCG-3') as a control. PCRs were performed in 10 µl reactions with BIO-X-ACT™ Short Mix (Bioline) on an Eppendorf Cyclor using the following protocol: 94 °C 4 min, followed by 39 cycles 94 °C 30 s, 60 °C 30 s, 72 °C 30 s and 72 °C 10 min final elongation. PCR products were checked visually on 1.5 % agarose TAE-gels. PCRs on a 917 bp fragment of the 16S rDNA of *Westeberhardia* gene were performed for one individual each of the three BR lineages (Ilhéus 2009, Una 2012, Ilhéus 2013), the infected JP lineages (OypC, OypU), the SP population and *C. wroughtonii* (PCR protocol: 94 °C 4 min; 39 x (94 °C 30 s; 50 °C 30 s; 72 °C 120 s); 72 °C 10 min). PCR products were purified using the Nucleo Spin Kit (Machery-Nagel) and Sanger sequenced (LGC Genomics, Germany).

### **Diagnostic PCR and qPCR assays on the *nrdB1* gene to assess inter- and intrapopulation variation**

To screen for intra- and interpopulation variation in *Westeberhardia* infection, we performed a diagnostic PCR assay for 538 workers and a diagnostic real-time quantitative PCR (qPCR) assay for 517 queens on *nrdB1*. Worker DNA was extracted with a rapid hot shot method (Alasaad *et al.*, 2008) and PCRs were performed in 10 µl reactions, with 5 µl BIO-X-ACT™ Short Mix (Bioline), 0.3 µl 10 µM forward and reverse primer each, 0.1 µl MgCl<sub>2</sub>, 3.3 µl H<sub>2</sub>O and 1 µl template DNA (PCR protocol: 94 °C 4 min; 39 x (94 °C 30 s; 60 °C 30 s; 72 °C 30 s); 72 °C 10 min). Subsequently, PCR products were checked on 1.5 % agarose gels with TAE buffer. Successful DNA extraction protocol was confirmed by amplification of *C. obscurior EF1* in PCR assays. Queen DNA was extracted using the NucleoSpin®Tissue XS Kit (Machery-Nagel) and real-time quantitative PCR was performed on a CFX Connect™ Real-Time PCR Detection System (BioRad) with 5 µl KAPA SYBR FAST Universal (peqlab), 1 µl template DNA, 2 µl H<sub>2</sub>O and 2 µl 2 µM forward and reverse primer each in 10 µl reactions. For each queen, we amplified the *nrdB1* fragment and the housekeeping gene *EF1* using the following protocol: 95 °C 4 min; 41 x (95 °C 10 s; 60 °C 30 s) followed by a melting curve with 0.5 °C temperature reduction every 5 s from 95 °C to 65 °C. Results of each assay were checked via amplification and melt curve analyses. A queen was ranked as infected, if the PCR on *nrdB* produced a single amplicon with the expected melting temperature (75.5 °C). A queen was ranked as not infected, if the PCR on *nrdB1* produced no amplicon, but the PCR on *EF1* was successful (amplicon with melting temperature of 82.5 °C).

## Rearing of individuals of defined age

To analyse *Westeberhardia*-abundance in dependence of age in adult ants, worker, queen and male pupae from BR (Una, 2012) were transferred from stock colonies to breeding colonies, which were screened daily to record exact hatching dates. Following hatching, breeding colonies were checked regularly twice a week to prevent additional individuals from hatching and to check for success of mating in queen breeding colonies (i.e. queen dealation and presence of eggs). For males and mated queens, breeding colonies consisted of 20 workers, few eggs and larvae, a single male pupa and a single queen pupa. To rear virgin queens, no male pupae were added to the colonies. To rear workers of the same age, late worker pupae were collected from stock colonies. After 24 h, workers that had emerged from the collected pupae were transferred to a new nest, together with few eggs and larvae from the stock colony. Individuals were sampled at the desired age and stored at  $-70^{\circ}\text{C}$  until DNA extraction.

## Real-time quantitative PCR to quantify *Westeberhardia* titers

We used real-time quantitative PCR to compare *Westeberhardia* infection levels between single individuals. This was done on the one hand for larvae, prepupae, early pupae, and late pupae of the OypB population (JP, 2011) and the BR (Una, 2012) population, respectively (Figure 2.3 B). All samples were sampled directly from stock colonies. Larvae and prepupa were of unknown sex and caste, whereas pupae were females (queens or workers). On the other hand this was done for age-controlled adult individuals (see above) from BR (Una, 2012) to infer dependency of *Westeberhardia* titer on morph and age of an individual (Figure 2.3 E,F). DNA of single ants was extracted using a chloroform-based method (Sambrook & Russell, 2001), quantified fluorometrically with Qubit<sup>®</sup> 2.0 Fluorometer (Life technologies) and diluted to 2.5 ng/ $\mu\text{l}$ . For each sample, we amplified a 204 bp fragment of the *nrdB* (*ribonucleoside-diphosphate reductase 1 subunit beta*) gene of *Westeberhardia* (*WEOB\_403*) using the primers *nrdB1*for and *nrdB1*rev (primer sequences see above). Additionally, for each sample two single-copy ant housekeeper genes were amplified: *EF1* (primer sequences see above; 104 bp fragment) and *Actin-4* (*Cobs\_04257*) (*Actin4*for: 5'-TGCCAAC-ACCGTTCTGTCTG-3', *Actin4*rev: 5'-GACGCGAGAATAGATCCGCC-3', 162 bp fragment). All real time qPCR reactions were performed in triplicates and each plate included no-template-controls for all primer pairs as well as triplets of an interplate calibrator (pool of 15 adult queens from Brazil (Una 2012)) for each of the two house-keeping genes. For each reaction 200 nM of forward and reverse primer were provided,

and a master mix including 5 ng DNA, 5 µl KAPA SYBR FAST Universal (peqlab) and H<sub>2</sub>O up to 10 µl was added. Hard-Shell<sup>®</sup> Low-Profile Thin-Wall 96-Well Skirted PCR Plates (BioRad) sealed with adhesive, optically clear MicroSeal seals (BioRad) were used and reactions were performed on a CFX Connect<sup>™</sup> Real-Time PCR Detection System (BioRad) with the following protocol: 95 °C 4 min; 41 x (95 °C 10 s; 60 °C 30 s) followed by a melting curve with 0.5 °C temperature reduction every 5 s from 95 °C to 65 °C. Analyses of C<sub>q</sub> values were performed with a modified protocol of the 2<sup>-ΔΔC<sub>q</sub></sup> protocol (Livak & Schmittgen, 2001). First, the mean of the three technical replicates was calculated. The resulting means were calibrated with the interplate calibrator, to normalize across the different plates. Calibrated means of target samples were then normalized by subtracting the geometric mean (GM) (Vandesompele *et al.*, 2002) of both housekeepers for the corresponding sample, giving the Δ C<sub>q</sub> value. To compare between all samples, an artificial calibrator with a C<sub>q</sub> value of 40 was used to normalize all samples. The Δ C<sub>q</sub> value for the artificial calibrator was determined by subtracting the mean of all geometrical means (mean=21.38; SD=0.24) from C<sub>q</sub>(calibrator)=40, yielding ΔC<sub>q</sub>(calibrator)=18.72. For each target gene ΔΔC<sub>q</sub> was calculated by subtracting ΔC<sub>q</sub>(target) - ΔC<sub>q</sub>(calibrator). Statistical tests were carried out with the ln transformed 2<sup>-ΔΔC<sub>q</sub></sup> values in R (version 3.0.2, R Core Team, 2013). We used Shapiro-Wilk tests to test for normal distribution of the data and Bartlett's tests to test for homogeneity of variances. For parametric data, we used Student's t-tests or ANOVAs followed by pairwise t-tests with post-hoc Benjamini-Hochberg correction for multiple testing (Benjamini & Hochberg, 1995). For non-parametric data we used Kruskal-Wallis tests or pairwise Mann-Whitney U-tests with post-hoc Benjamini-Hochberg correction. The boxplots (Fig. 2.3 B, E, F) show ln transformed 2<sup>-ΔΔC<sub>q</sub></sup> values.

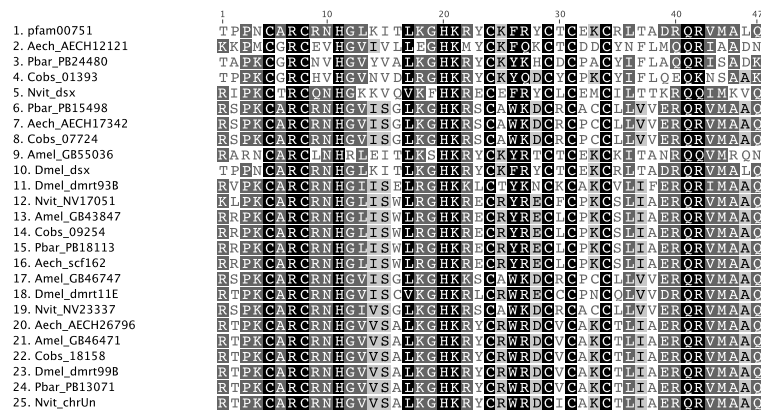
## Fluorescence *in situ* hybridization

Ants were fixated in 4 % paraformaldehyde in PBS and embedded in cold-polymerizing resin (Technovit 8100, Heraeus Kulzer, Germany). Longitudinal sections (5 µm) through the abdomen were obtained with a Microm HM355S microtome (Thermo Fisher Scientific, Germany) and mounted on microscope slides coated with poly-L-lysine (Kindler, Germany). Tissue sections were incubated for 90 min at 50 °C in hybridization buffer (0.9 M NaCl, 20 mM Tris/HCl pH=8.0, 0.01 % SDS), containing 0.5 µM of the general eubacterial probe *EUB338* (5'-GCTGCCTCCCGTAGGAGT-3') (Amann *et al.*, 1990) and one of the *Westeberhardia*-specific probes *Wcard1* (5'-ATCAGTTTCGAACGCC-ATTC-3') and *Wcard2* (5'-CGGAAGCCACAATTCAAGAT-3'). Probes were labeled

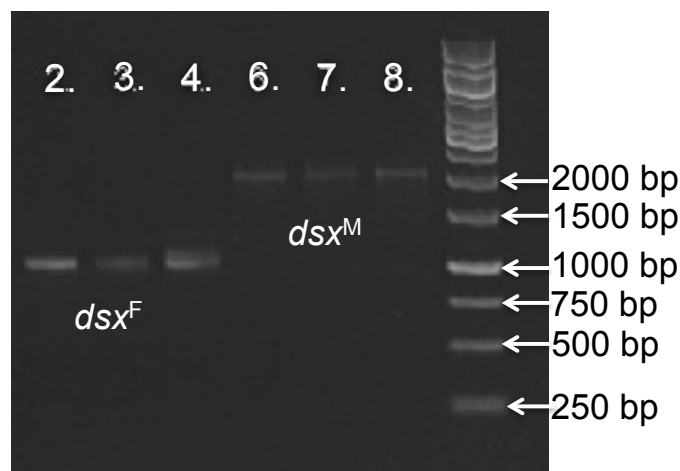
with Cy3 or Cy5, and samples were counterstained with 5 µg/ml DAPI (4',6-diamidino-2-phenylindole). After hybridization, samples were washed once with pre-warmed wash buffer (0.1 M NaCl, 20 mM Tris/HCl pH=8.0, 5 mM EDTA, 0.01 % SDS), incubated in the same buffer for 20 min at 50 °C, washed twice in ddH<sub>2</sub>O, air-dried, and finally covered with VectaShield<sup>®</sup> (Vector Laboratories, Burlingame, CA, USA). Images were acquired on an AxioImager.Z1 epifluorescence microscope (Carl Zeiss, Jena, Germany), using the mosaic tool and the z-stack option to combine different focus planes.

# Supplement Chapter 3

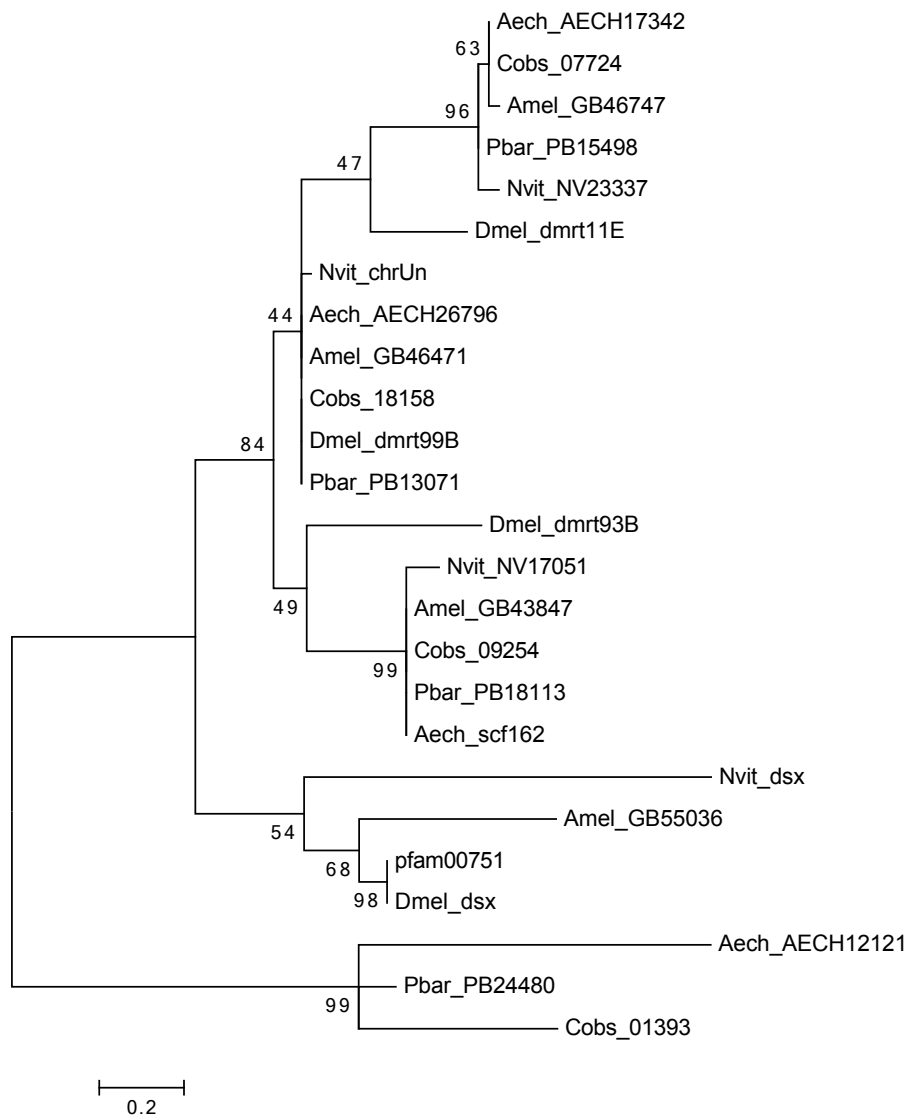
## Supplementary Figures



**Figure S3.1:** Alignment of the amino acid sequences of the DM domain (pfam00751) for DM domain-containing proteins of *Drosophila melanogaster* (Dmel), *Apis mellifera* (Amel), *Nasonia vitripennis* (Nvit), *Acromyrmex echinator* (Aech), *Pogonomyrmex barbatus* (Pbar) and *Cardiocondyla obscurior* (Cobs).

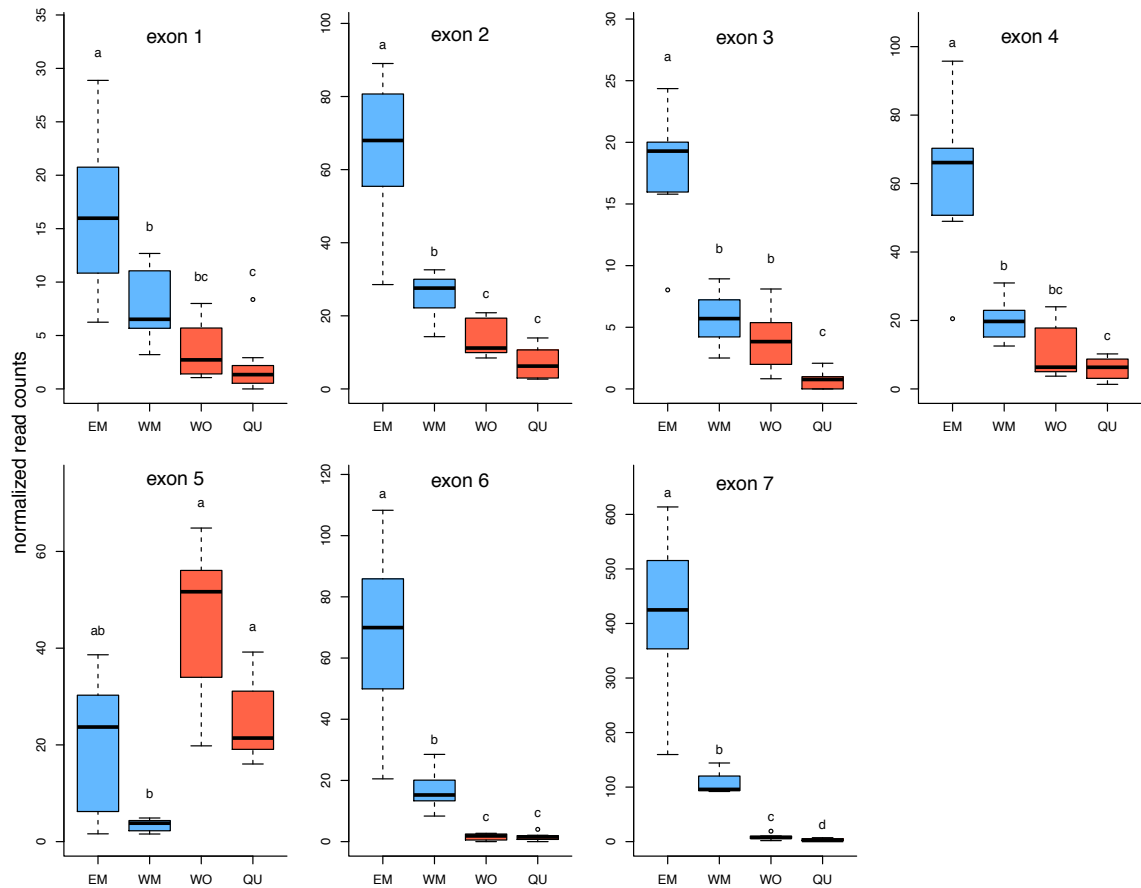


**Figure S3.2:** Gel image of sex-specific dsx splicing variants amplified by 3' RACE (2-4: *dsx<sup>F</sup>* in females, 6-8: *dsx<sup>M</sup>* in males).

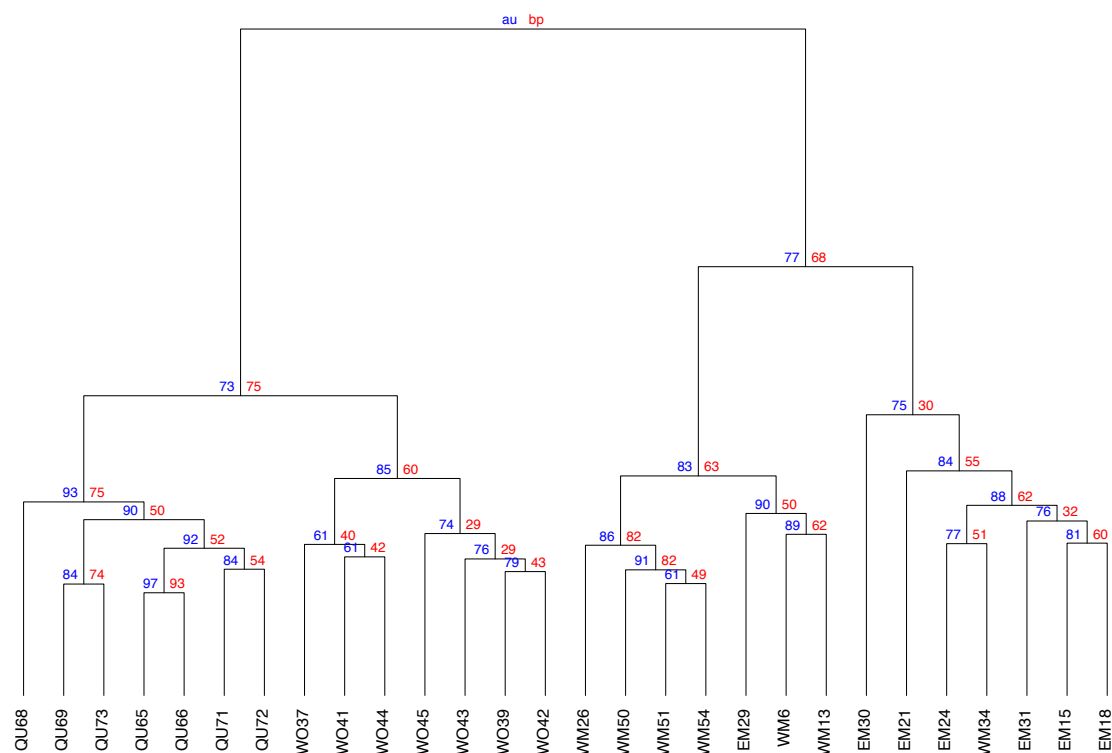


**Figure S3.3:** Phylogenetic tree based on amino acid sequence of the DM domain of four paralogous genes each of *Drosophila melanogaster* (Dmel), *Apis mellifera* (Amel), *Nasonia vitripennis* (Nvit), *Acromyrmex echinator* (Aech), *Pogonomyrmex barbatus* (Pbar) and *Cardiocondyla obscurior* (Cobs). Supplementary Table 1 shows an overview of the used sequences. For phylogenetic tree reconstruction we used MEGA (Hall, 2013), and applied a WAG+G+I phylogenetic model and bootstrap resampling with 1,000 replicates. Numbers show bootstrap support values.

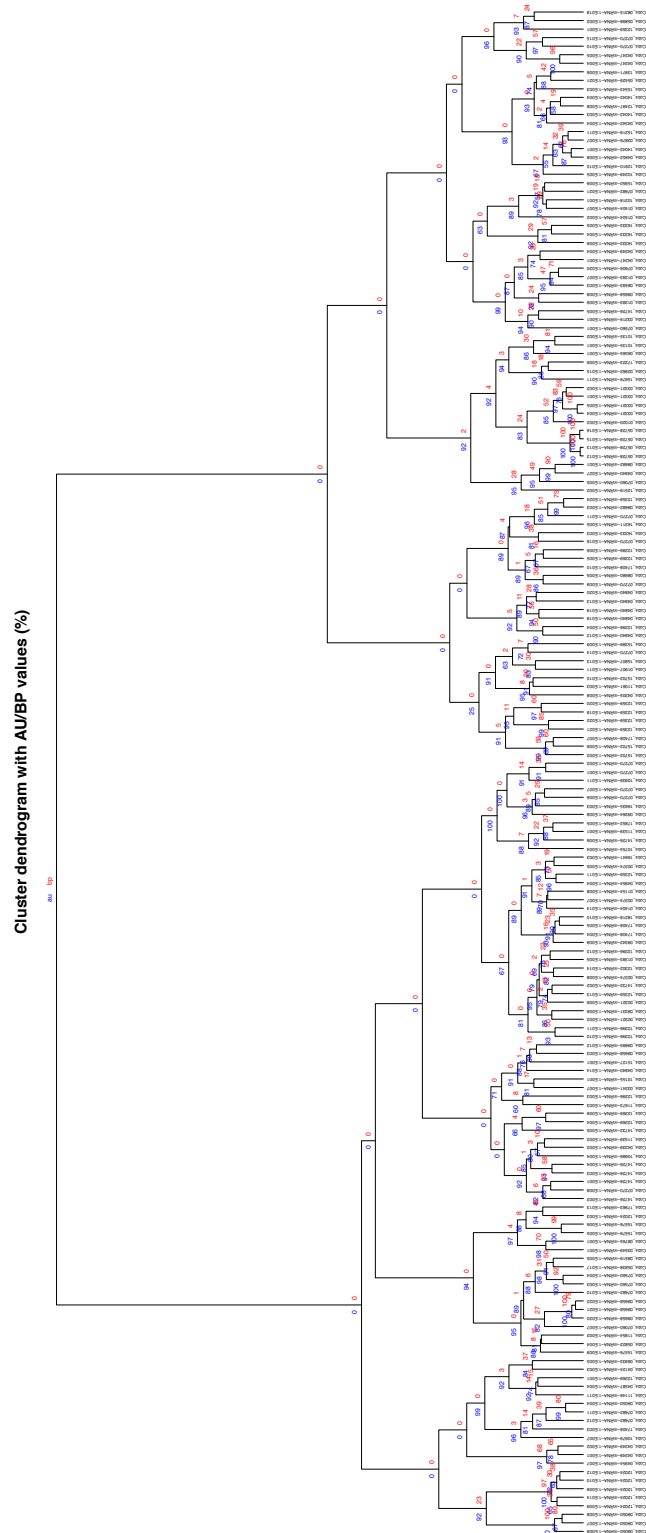




**Figure S3.4:** Normalized read counts mapping against *dsx* exons. RNAseq data of larvae of all four morphs (EM = wingless, ergatoid males, WM = winged males, WO = workers, QU = queens, N=7 each) previously published in (Schrader *et al.*, 2015) were reanalyzed using DEXseq (Anders *et al.*, 2012). Female castes show significant differences in expression of exons 3 and 7, male morphs show significant differences in all exons except exon 5. Letters indicate significant differences based on pairwise Wilcoxon Tests with Benjamini-Hochberg correction for multiple testing (see Supplementary Table 7).



**Figure S3.5:** Cluster dendrogram of larvae based on expression of 179 sex-specifically expressed exons. Red values show bootstrap probabilities (bp); blue values show approximately unbiased p-values (au). Workers (WO) and queens (QU) are well separated, whereas one winged male (WM) clusters within the wingless, ergatoid males (EM) and one ergatoid male clusters within the winged males.



**Figure S3.6:** Cluster dendrogram of 179 sex-specifically expressed exons. Red values show bootstrap probabilities (bp); blue values show approximately unbiased p-values (au).

## Supplementary Tables

**Table S3.1:** Overview over DM domain-containing proteins used for phylogenetic tree reconstruction. *Apis mellifera* (Amel), *Nasonia vitripennis* (Nvit), *Cardiocondyla obscurior* (Cobs), *Acromyrmex echinator* (Aech) and *Pogonomyrmex barbatus* (Pbar) predicted proteins were accessed by BLASTp and tBLASTn analyses on <http://hymenoptera.genome.org/>. *Drosophila melanogaster* (Dmel) proteins were accessed on <http://flybase.org/blast/>.

species (database)	proteins
Dmel (Flybase)	dsx (FBgn0000504) dmrt99B (FBgn0039683) dmrt93B (FBgn0038851) dmrt11E (FBgn0030477)
Amel (Amel_4.5_OGSv3.2_pep)	GB55036 GB46471 GB43847 GB46747
Nvit (Nasonia Official Gene Set v1.2 Protein)	dsx: Genbank accession: NP_001155990 NV17051 NV23337 no predicted transcript; chrUn: positions 3,401,206 – 3,401,096
Cobs (Cardiocondyla__ obscurior__1.4__ proteins)	Cobs_09254 (scf0029) Cobs_18158 (scf0002) Cobs_07724 (scf0049) Cobs_01393 (scf0005)
Aech (Aech_OGSv3.8__ proteins)	AECH26796 AECH17342 AECH12121 no predicted transcript; scf162: Positions 728,652 – 728,804
Pbar (proteins 1.2)	PB13071 PB18113 PB24480

**Table S3.2:** Sequences and targets of primers used in this study, based on *C. obscurior* genome version 1.4 (Schrader *et al.*, 2014). Primers were used either for 3' Rapid amplification of cDNA ends (RACE), in real-time quantitative PCR (qPCR) as housekeepers (HK) or targets (T), or for microsatellite analyses (MS).

primer name	Sequence (5' → 3')	target (transcript-ID / scaffold, positions)	usage
RACE outer = dsx4_for4	TTGATCTCAGAGGACACAAG	Cobs_01393 (dsx), exon 2	RACE + qPCR (T)
RACE inner = Co_dsx_p3_for	GATGAACGAAATTTGCCCTCGGTC	Cobs_01393 (dsx), exon 3	RACE
dsx4_rev1	GGATAGGAAACCAATGATGAT	Cobs_01393 (dsx), exon 3	qPCR (T)
4for	CCTCGGACTTGAGACTAAACG	Cobs_01393 (dsx), exon 4	qPCR (T)
F5rev	GGCGAATATTTTCAAAATGACGAG	Cobs_01393 (dsx), exon 5	qPCR (T)
M5rev	GTTCTCACAACCATGATGATG	Cobs_01393 (dsx), exon 6	qPCR (T)
RPL32_for	TCGCAGGGCGTTTAAAGGGCCA	Cobs_10346 (60S ribosomal protein L32)	qPCR (HK)
RPL32_rev	CTCCGAACGCAAGCGTGCACTA		
RPS2_new_for	AAGCCATTCTGCGATGGCC	Cobs_18295 (DNA-directed RNA polymerase II subunit RPB1, multi-copy gene)	qPCR (HK)
RPS2_new_rev	TCGAAGCCAAACATGCTTAGCG		
Y45F10D_JO1_for	CATCGGCGCGACGTCCAAGA	Cobs_04843 (iron-sulfur cluster assembly enzyme ISCU, mitochondrial)	qPCR (HK)
Y45F10D_JO1_rev	GCCCCCACCAGACCTGTTC		
Cobs_1.1_for-FAM	GGATCCCGAAATCAGCTAAATAAA	scf0002: 7,823,131 – 7,823,155	MS
Cobs_1.1_rev	CGTATCAACTGGAAATTTTAGA	scf0002: 7,823,288 – 7,823,267	
Cobs_8.3_for-FAM	TAACAGAGTTCCGTAGGTTT	scf0008: 2,887,283 – 2,887,302	MS
Cobs_8.3_rev	TGATGTTATCTCAATACTGGTCT	scf0008: 2,887,416 – 2,887,394	
Cobs_8.4_for-HEX	GCTAAACGCAGTCACAGTCT	scf0008: 3,226,875 – 3,226,894	MS
Cobs_8.4_rev	TGTTATATCTCTTAAAAATTTGCAT	scf0008: 3,227,032 – 3,227,008	

**Table S3.3:** Gene expression of the four DM-containing genes of *C. obscurior*. We used previously published RNAseq data (Schrader *et al.*, 2015) and analyzed expression in females vs. males with DESeq2 (Love *et al.*, 2014). Cobs\_01393 is the only paralog differentially expressed between the sexes.

transcript	base Mean	log2FoldChange	padj
Cobs_01393	53.22	1.74	< 0.001
Cobs_07724	17.83	0.09	0.83
Cobs_09254	11.49	-0.11	0.81
Cobs_18158	46.63	0.04	0.90

**Table S3.4:** Gene structure of *C. obscurior doublesex*. Positions are based on genome version 1.4 (Schrader *et al.*, 2014).

exon	positions on scf0005	size in bp
1	3,962,553 – 3,962,614	62
2	3,962,850 – 3,963,207	358
3	3,963,485 – 3,963,524	40
4	3,973,929 – 3,974,117	189
5 (female-specific)	3,974,295 – 3,975,124	830
6 (male-specific)	3,989,920 – 3,990,340	421
7 (male-specific)	3,999,828 – 4,004,104	4277

**Table S3.5:** Results of microsatellite analyses of F1 individuals emerging from interpopulational crosses. For each family, a Japanese (JP) *C. obscurior* queen was mated with a Brazilian (BR) male. Emerging F1 individuals were genotyped using three population-specific microsatellite markers. This showed that all F1 males (EM = ergatoid, wingless males, WM = winged males) and one gynandromorph (GY) exclusively carried the maternal (JP) allele, whereas emerging females (QU = queens, WO = workers) carried both parental alleles (JP+BR). Sample sizes are given in parenthesis.

family	F1-EM	F1-WM	F1-QU	F1-WO	F1-GY
sd5	JP (4)	-	JP+BR (5)	JP+BR (4)	-
sd9	JP (4)	JP (1)	JP+BR (4)	JP+BR (4)	JP (1)
sd11	JP (4)	-	JP+BR (4)	JP+BR (4)	-
sd12	JP (5)	JP (2)	JP+BR (4)	JP+BR (4)	-
sd18	JP (6)	-	JP+BR (5)	JP+BR (6)	-
sum	23	3	22	22	1

**Table S3.6:** Morphological description of *C. obscurior* sex-mosaics sampled during the course of this study. Sex mosaics are classified as ergatandromorph (E, intersex worker (WO) / ergatoid male (EM)) or gynandromorph (G, intersex queen (QU) / winged male (WM)). Morphological descriptions mostly rely on the head, as here morphological differences are easily identifiable.

id	G/E	life stage	morphological description
# 4*	G	pupa	left half: QU, right half: WM
# 6*	G	pupa	left half: QU + male antennae, right half: WM + QU antennae
# 8*	E	pupa	left half: EM, right half: WO
# 10*	G	pupa	left half: WM, right half: QU
# 7	G	adult	laterally separated in QU and WM
# 9	E	adult	laterally separated in WO and EM
# 11	G	adult	laterally separated in QU and WM
# 12	G	adult	both eyes big like in WM; both antennae short scapus like in QU
# 14	G	adult	both eyes big like in WM; right antenna short scapus like in QU, left antennae long scapus like in WM, left antennae
# 15	G	adult	both eyes big like in WM; right antenna long scapus like in WM, left antennae short scapus like in QU, left antenna long scapus like in WM
# 16	G	adult	laterally separated in QU and WM
# 17	E	pupa	eyes WO-like, saber-shaped mandibels like in EM
# 18	G	adult	both eyes big like in WM, both antennae short scapus like in QU
# 19	G	adult	head of WM, but wings were shed, meaning that thorax is probably of QU origin
# 26	G	adult	laterally separated in QU and WM
# 27	G	adult	laterally separated in QU and WM
# 34	E	pupa	laterally separated in WO and EM
* = used in qPCR			

**Table S3.7:** Results of statistical tests (Kruskal-Wallis rank sum tests and pairwise Wilcoxon-Test with Benjamini-Hochberg correction) for normalized count data. RNAseq data (2) was used to generate per exon count tables for the corrected *dsx* gene model for 3rd instar larvae.

exon	kruskal test across all morphs	pairwise wilcox test					
		EM- WM	EM- WO	EM- QU	WM- WO	WM- QU	WO- QU
exon1	X2=16.13, df=3, p=0.011	0.0390	0.0150	0.0150	0.0640	0.0210	0.2500
exon2	X2=22.71, df=3, p<0.001	0.0035	0.0012	0.0012	0.0049	0.0012	0.0530
exon3	X2=21.68, df=3, p<0.001	0.0031	0.0031	0.0031	0.2086	0.0031	0.0123
exon4	X2=19.06, df=3, p<0.001	0.0061	0.0047	0.0017	0.1168	0.0017	0.3176
exon5	X2=16.49, df=3, p<0.001	0.0568	0.0524	0.6200	0.0017	0.0017	0.0636
exon6	X2=22.70, df=3, p<0.001	0.0026	0.0026	0.0026	0.0026	0.0026	0.7961
exon7	X2=24.10, df=3, p<0.001	0.0007	0.0007	0.0007	0.0007	0.0007	0.0262

**Table S3.8:** Loadings of 179 exons on the first four principal components (bold numbers indicate loadings that fell in the 10 % or 90 % quantiles for PC1, PC2 & PC4. The 179 sex-biased exons were extracted with DEXseq using a false discovery rate of 0.005.)

Exon ID	PC1	PC2	PC3	PC4
Cobs_00321.mRNA.1.E002	0.0562382873	-0.0469522101	-0.0522020161	-0.0248998208
Cobs_00321.mRNA.1.E006	0.0209059028	-0.0535784210	-0.0663160451	-0.0033977391
Cobs_00374.mRNA.1.E005	0.0988034890	0.0781461284	0.0241897146	0.0186604289
Cobs_00374.mRNA.1.E007	0.0776478991	0.0879914611	0.0077047346	0.0538194057
Cobs_00374.mRNA.1.E008	0.0777561199	0.0063557308	-0.0490479993	0.0042016531
Cobs_01020.mRNA.1.E002	0.0570320853	-0.1371031360	-0.0289168756	0.0582219468
Cobs_01154.mRNA.1.E008	0.0259214593	0.0978799515	0.0871926940	0.0875834806
Cobs_01393.mRNA.1.E005	0.0752227118	-0.0272423911	-0.0863315256	0.1199763586
Cobs_01393.mRNA.1.E006	-0.0983936018	0.0270897173	-0.0421485645	0.1370003723
Cobs_01393.mRNA.1.E007	-0.1026979831	0.0263004114	-0.0524285714	0.1306402712
Cobs_01404.mRNA.1.E007	-0.0735259289	0.0610078176	0.0554938947	-0.1422011784
Cobs_01404.mRNA.1.E014	0.0517963573	0.1515488442	0.0732645364	-0.0218614922
Cobs_01504.mRNA.1.E002	0.0016959771	0.1402418831	-0.0135906012	-0.0855509310
Cobs_01907.mRNA.1.E011	-0.0891935029	-0.0946207759	-0.0025935604	-0.0439723528
Cobs_02962.mRNA.1.E015	-0.0006626646	-0.1499880879	-0.0123599697	-0.0860383784
Cobs_03219.mRNA.1.E001	-0.0770700977	0.0872844848	-0.0793515483	-0.0085188897
Cobs_03241.mRNA.1.E007	-0.0895224483	0.0348534133	0.0660123650	0.0117203295
Cobs_03321.mRNA.1.E001	0.0627406550	-0.1492376292	-0.0363229466	-0.0658655640
Cobs_03321.mRNA.1.E002	0.0704399383	-0.1496325800	-0.0357914249	-0.0260744246
Cobs_03321.mRNA.1.E004	0.0936470138	-0.1123395150	-0.0604334007	0.0209392194
Cobs_03321.mRNA.1.E005	0.0928920791	-0.1175679389	-0.0504467044	0.0350537294
Cobs_03549.mRNA.1.E001	0.0872420519	-0.0564780017	-0.0214028332	-0.0448670244
Cobs_03876.mRNA.1.E007	-0.1057160446	0.0006825815	-0.0217198809	-0.0939389563



**Table S3.8: (continued)** Loadings of 179 exons on the first four principal components (bold numbers indicate loadings that fell in the 10 % or 90 % quantiles for PC1, PC2 & PC4. The 179 sex-biased exons were extracted with DEXseq using a false discovery rate of 0.005.)

Exon ID	PC1	PC2	PC3	PC4
Cobs_04124.mRNA.1.E003	0.0056562913	0.0811835272	-0.1426600756	-0.0226364374
Cobs_04205.mRNA.1.E008	0.0503024972	-0.0879869261	0.0214201339	0.0774399590
Cobs_04247.mRNA.1.E001	-0.0059591657	-0.0225328679	-0.0144477939	-0.0164814469
Cobs_04247.mRNA.1.E004	-0.0865993406	-0.0286615794	0.1045264707	-0.0109143738
Cobs_04247.mRNA.1.E005	-0.0912547440	-0.0345202184	0.1076812888	-0.0632772076
Cobs_04249.mRNA.1.E001	0.0356037902	0.1351604504	0.1222119190	-0.0336030054
Cobs_04249.mRNA.1.E002	0.0395380636	0.1377974774	0.1202601310	-0.0374530368
Cobs_04249.mRNA.1.E004	-0.0939417760	0.0227188679	-0.1104284859	-0.0692015158
Cobs_04339.mRNA.1.E002	-0.0253937630	0.0424600334	-0.0594897474	0.1833834904
Cobs_04342.mRNA.1.E004	-0.0920855018	-0.0589812343	-0.0787523569	-0.0330124227
Cobs_04587.mRNA.1.E004	0.0629235699	0.0528072046	-0.1639305528	-0.0318304980
Cobs_04652.mRNA.1.E008	0.1000144284	0.0743249095	0.0280407205	-0.0004933052
Cobs_04840.mRNA.1.E012	-0.1097961436	-0.0511257274	-0.0437184486	-0.0199934947
Cobs_04840.mRNA.1.E013	-0.1054513384	0.0279910100	0.0238100727	-0.0962623918
Cobs_04840.mRNA.1.E014	-0.0992775999	-0.0090532787	-0.0118330020	-0.0554953936
Cobs_04840.mRNA.1.E018	-0.0997497449	0.0125108153	-0.0648554888	-0.0654761828
Cobs_04840.mRNA.1.E019	-0.0857149882	-0.0390511265	-0.0581807849	-0.1266666118
Cobs_04840.mRNA.1.E020	-0.0931041770	-0.0416563358	-0.0503802388	-0.1142545550
Cobs_04840.mRNA.1.E027	0.0663718368	-0.1121115144	-0.0051625831	-0.0061544746
Cobs_04954.mRNA.1.E004	0.0659148888	0.1215326744	0.0975162350	0.0078514808
Cobs_04954.mRNA.1.E007	0.0338963657	0.1343001572	0.1386968379	-0.0376432495
Cobs_05429.mRNA.1.E021	0.0335194626	-0.1107957920	0.0866388582	0.0742242668
Cobs_05728.mRNA.1.E012	0.0849394796	-0.0743109246	-0.0891492315	-0.1217624595
Cobs_05728.mRNA.1.E013	0.0811590157	-0.0689587947	-0.0900994456	-0.1344718007
Cobs_05728.mRNA.1.E015	0.1053835877	-0.0440799648	-0.0861253515	-0.0953617875
Cobs_05728.mRNA.1.E016	0.1092376107	-0.0395369286	-0.0697426526	-0.0861952490
Cobs_05822.mRNA.1.E004	0.0710459452	-0.1250571023	-0.0493063265	0.0370597079
Cobs_05838.mRNA.1.E017	0.0930996779	-0.0085289707	-0.1483049666	-0.0405078520
Cobs_05898.mRNA.1.E002	0.0684530125	0.0883671166	0.0837715341	-0.0607667855
Cobs_06565.mRNA.1.E001	0.0417674481	-0.1102425688	-0.1090104584	-0.0449167273
Cobs_06593.mRNA.1.E003	-0.1048160252	0.0123267175	-0.0195574944	0.1054393843
Cobs_06933.mRNA.1.E005	0.0764959126	0.0451070667	-0.1243731488	-0.0035664523
Cobs_07060.mRNA.1.E005	0.0562128432	-0.1098553340	-0.0327398337	-0.0967354323
Cobs_07060.mRNA.1.E007	0.0696803503	-0.0796269641	-0.0691106431	-0.0856100089
Cobs_07270.mRNA.1.E001	0.0176370090	0.0050354935	0.0411620002	-0.0481000547
Cobs_07270.mRNA.1.E002	0.0036710805	0.0282291310	0.0886117710	0.0695161889
Cobs_07270.mRNA.1.E006	-0.0317105271	0.0141466398	0.0663072679	-0.0086866251
Cobs_07270.mRNA.1.E007	-0.0193510669	0.0176931113	0.0353681526	0.0566708853
Cobs_07270.mRNA.1.E008	0.0247357353	0.0526113501	-0.0395991067	0.0551707820
Cobs_07270.mRNA.1.E009	-0.0792805534	-0.0584768615	0.0795268584	-0.1037892954

**Table S3.8: (continued)** Loadings of 179 exons on the first four principal components (bold numbers indicate loadings that fell in the 10 % or 90 % quantiles for PC1, PC2 & PC4. The 179 sex-biased exons were extracted with DEXseq using a false discovery rate of 0.005.)

Exon ID	PC1	PC2	PC3	PC4
Cobs_07270.mRNA.1.E010	-0.0831577975	-0.0588943538	0.0798053896	-0.0700194650
Cobs_07270.mRNA.1.E011	-0.0925733224	-0.0403000965	0.0965028452	-0.0601201052
Cobs_07270.mRNA.1.E014	-0.0811809833	-0.0569213253	0.0640064847	-0.1500016866
Cobs_07270.mRNA.1.E015	-0.0728829201	-0.0582242860	0.0822015410	-0.1350003875
Cobs_07270.mRNA.1.E016	-0.0818668733	-0.0603957907	0.0900087708	-0.0916281460
Cobs_07560.mRNA.1.E001	-0.0962700875	0.0290439890	-0.0587213684	0.0393211818
Cobs_07560.mRNA.1.E003	0.0921253848	-0.0182648789	-0.1032532346	-0.0479360658
Cobs_07560.mRNA.1.E004	0.0797172627	-0.0002137049	-0.1114493102	-0.0982133335
Cobs_07606.mRNA.1.E020	-0.1118297296	0.0024240695	-0.1107470635	0.0300095574
Cobs_07682.mRNA.1.E010	0.0582640932	0.0802506622	0.0488494809	-0.1094219505
Cobs_07682.mRNA.1.E011	0.0713310860	0.0815219093	0.0609168669	-0.1094844429
Cobs_07682.mRNA.1.E012	0.0602911509	0.0854491656	0.0621900794	-0.1270765090
Cobs_07682.mRNA.1.E021	-0.0560209098	0.0724286551	0.1317773037	-0.1127351340
Cobs_08315.mRNA.1.E018	-0.0459039203	-0.0688403733	0.0704164252	-0.1415938024
Cobs_08519.mRNA.1.E005	0.0887047414	-0.0422317974	-0.1292681106	-0.0404120195
Cobs_08680.mRNA.1.E005	-0.0687866380	0.0051932459	-0.0363867875	-0.0125091250
Cobs_08682.mRNA.1.E001	0.0733209189	-0.1216389725	-0.0174783357	-0.0009056975
Cobs_08682.mRNA.1.E003	-0.1022600015	0.0111618175	-0.0808747737	0.1139248490
Cobs_08765.mRNA.1.E001	0.0616058993	-0.0576938929	0.0398078055	0.0051611451
Cobs_09050.mRNA.1.E004	0.0674403708	0.0630585960	-0.0138253133	-0.1076868685
Cobs_09050.mRNA.1.E007	-0.0362849706	0.0952325307	-0.1017677965	-0.1250339366
Cobs_09050.mRNA.1.E008	-0.0441826812	0.0888968170	-0.0895999365	-0.1400029457
Cobs_09050.mRNA.1.E009	-0.0324334885	0.0985614853	-0.1073047609	-0.1550594271
Cobs_09264.mRNA.1.E008	0.0584143988	0.0620081748	-0.0202413285	0.0473076855
Cobs_09331.mRNA.1.E006	0.0691173809	-0.0530090865	-0.0635604435	0.0239948263
Cobs_09342.mRNA.1.E008	0.0979050052	0.0709239011	-0.0781798985	-0.0288999326
Cobs_09658.mRNA.1.E003	-0.0676272380	0.0777055469	-0.0129724379	-0.0018916338
Cobs_09658.mRNA.1.E008	-0.0425239715	0.0781639748	-0.1407176502	-0.0241373541
Cobs_09658.mRNA.1.E020	0.1036047503	-0.0378552027	-0.1210111711	-0.0251442178
Cobs_09658.mRNA.1.E021	0.1068589058	-0.0436906731	-0.1112571209	-0.0301942771
Cobs_09658.mRNA.1.E022	0.1061528794	-0.0348364113	-0.1149623731	-0.0487857966
Cobs_09895.mRNA.1.E012	-0.0686270412	0.0831334236	-0.0736872057	-0.0349327826
Cobs_10135.mRNA.1.E001	-0.0422443780	-0.0438499936	-0.0607409909	-0.1723455107
Cobs_10135.mRNA.1.E002	0.0432025775	-0.0567729370	-0.1054888378	-0.0774635115
Cobs_10155.mRNA.1.E001	-0.1006603208	-0.0146498091	-0.0298257150	0.0584648543
Cobs_10249.mRNA.1.E005	-0.0508352634	0.0555190408	0.0142071490	-0.0655331308
Cobs_10318.mRNA.1.E001	-0.0713221665	0.0918495331	0.0364389756	-0.0863922601
Cobs_10679.mRNA.1.E007	0.0990014265	-0.0040951762	-0.0321325993	-0.0338906578
Cobs_10686.mRNA.1.E004	-0.0951010503	0.0240402354	-0.0698104309	0.1239063399
Cobs_10755.mRNA.1.E004	0.0669928780	0.0570598320	-0.0463348160	0.1202727480

**Table S3.8: (continued)** Loadings of 179 exons on the first four principal components (bold numbers indicate loadings that fell in the 10 % or 90 % quantiles for PC1, PC2 & PC4. The 179 sex-biased exons were extracted with DEXseq using a false discovery rate of 0.005.)

Exon ID	PC1	PC2	PC3	PC4
Cobs_10939.mRNA.1.E011	-0.0429223561	-0.0054908365	0.1093785715	-0.0618795340
Cobs_11061.mRNA.1.E003	-0.0090875798	-0.1314355719	0.0812395202	0.0913591000
Cobs_11146.mRNA.1.E011	0.0690400597	0.1051375547	0.0209936049	-0.0196831326
Cobs_11539.mRNA.1.E001	0.0715162494	0.1023505479	0.0284929727	0.1083570685
Cobs_11539.mRNA.1.E002	-0.0640259976	-0.0099304180	-0.0329083163	0.1524632964
Cobs_11673.mRNA.1.E003	-0.0480845499	-0.0046769280	-0.0400688482	-0.0729317711
Cobs_11854.mRNA.1.E002	0.0667928997	-0.1136061746	-0.0063148972	-0.0478166717
Cobs_12024.mRNA.1.E003	0.0973610344	0.0097371701	-0.0752641859	-0.0739590929
Cobs_12024.mRNA.1.E006	-0.0656009304	0.1275013470	-0.0900714800	-0.0288808270
Cobs_12024.mRNA.1.E009	-0.0740325530	0.0998158800	-0.1125229548	-0.0397625456
Cobs_12024.mRNA.1.E010	-0.0616677625	0.1307255933	-0.0792081979	-0.0356211696
Cobs_12024.mRNA.1.E012	-0.0652069616	0.1315692975	-0.0713111765	-0.0256511763
Cobs_12024.mRNA.1.E014	-0.0653527769	0.1076954742	-0.0998658606	-0.0167308590
Cobs_12269.mRNA.1.E001	0.0762840000	0.0855582439	-0.0542601085	-0.0514465016
Cobs_12269.mRNA.1.E004	-0.0726983968	0.1002327288	-0.1006608206	0.0637172228
Cobs_12269.mRNA.1.E005	-0.0759348103	0.0770418683	-0.0588225035	0.0147178905
Cobs_12269.mRNA.1.E006	-0.0760003143	0.0970345300	-0.0792093452	0.0848421752
Cobs_12296.mRNA.1.E003	-0.1116698382	0.0057954363	-0.0641288785	-0.0337302535
Cobs_12296.mRNA.1.E004	-0.1077903141	-0.0167899405	-0.1095983859	0.0135853723
Cobs_12296.mRNA.1.E006	-0.1048772824	0.0059773416	-0.0669533787	0.0733045471
Cobs_12296.mRNA.1.E010	0.0426252350	-0.0560399482	0.0277342049	0.1568083507
Cobs_12296.mRNA.1.E011	0.0566102870	-0.0607572525	0.0496290344	0.1574800578
Cobs_12296.mRNA.1.E012	0.0376749395	-0.0670643660	-0.0106826181	0.1304610266
Cobs_12302.mRNA.1.E014	0.0403836525	-0.0138352663	0.0028241888	-0.0119398657
Cobs_12359.mRNA.1.E001	0.0195337865	0.0708711805	0.1233542800	-0.0080550379
Cobs_12359.mRNA.1.E011	0.0580935484	0.0836772361	0.1151622522	-0.0340260625
Cobs_12359.mRNA.1.E013	0.0824857727	0.0670909829	0.0198038890	0.0166588643
Cobs_12359.mRNA.1.E018	-0.1062291699	-0.0534953766	-0.0225090992	-0.0348680520
Cobs_12359.mRNA.1.E020	-0.0977898959	-0.0483641059	0.0214106886	-0.0257368707
Cobs_12359.mRNA.1.E021	-0.1087821404	-0.0327849447	0.0802310272	0.0090158166
Cobs_12359.mRNA.1.E022	-0.1115455126	-0.0367064469	0.0652272465	0.0301763646
Cobs_12359.mRNA.1.E024	-0.1081852780	-0.0139146636	-0.0397573699	-0.0698419724
Cobs_12497.mRNA.1.E008	-0.0909576464	-0.0561074740	0.1202726163	0.0767377182
Cobs_12519.mRNA.1.E002	0.0926462189	0.0165413616	-0.0714085109	-0.0804290907
Cobs_12910.mRNA.1.E010	0.0414125749	0.1358258365	0.0718225411	-0.0782091007
Cobs_13971.mRNA.1.E006	-0.0055882286	-0.0283775684	0.0568499902	0.0503206621
Cobs_14042.mRNA.1.E001	0.1079279128	0.0826460606	0.0235494851	-0.0244945073
Cobs_14042.mRNA.1.E003	-0.0350497375	-0.1168862838	-0.0273618692	-0.0084846726
Cobs_14042.mRNA.1.E004	-0.0389630641	-0.1125577642	0.0189555622	-0.0677494098
Cobs_14726.mRNA.1.E001	-0.0358643610	0.0771303818	-0.1632061911	0.0528253378

**Table S3.8: (continued)** Loadings of 179 exons on the first four principal components (bold numbers indicate loadings that fell in the 10 % or 90 % quantiles for PC1, PC2 & PC4. The 179 sex-biased exons were extracted with DEXseq using a false discovery rate of 0.005.)

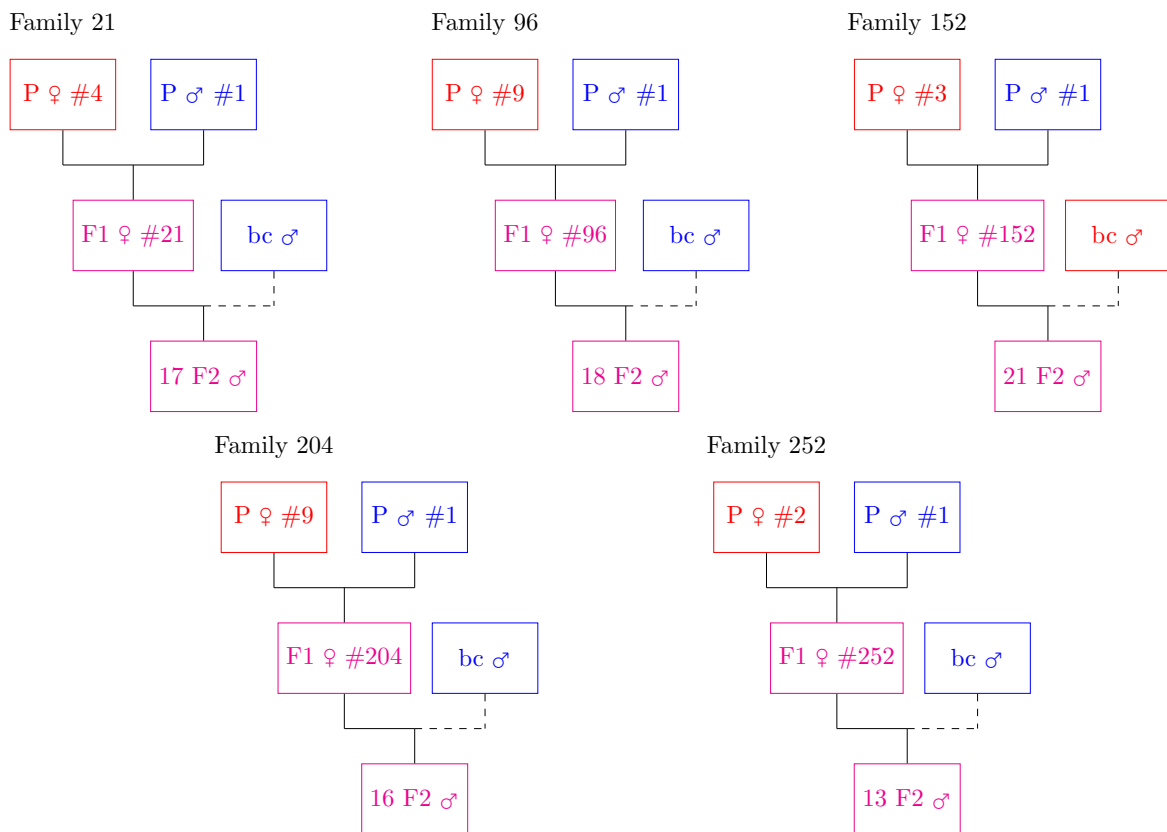
Exon ID	PC1	PC2	PC3	PC4
Cobs_14726.mRNA.1.E002	-0.0465444829	0.0786262098	-0.1410842356	0.0663365579
Cobs_14726.mRNA.1.E003	-0.0524339633	0.0609795740	-0.1456727656	0.0573440049
Cobs_14726.mRNA.1.E004	-0.0230923960	0.0583610072	-0.1239663597	0.0487567124
Cobs_14726.mRNA.1.E005	-0.0821946259	0.0744758323	-0.1365683415	0.0136492522
Cobs_14726.mRNA.1.E006	0.0998324945	0.0687706939	-0.0232245446	0.0851115591
Cobs_14732.mRNA.1.E002	0.0986371407	0.0059237625	-0.0314611866	0.0476048891
Cobs_14732.mRNA.1.E005	-0.0940764916	0.0050943488	-0.0978083118	0.0083284816
Cobs_15127.mRNA.1.E001	-0.0887239946	-0.0443620782	0.0256253602	-0.0026329274
Cobs_15219.mRNA.1.E011	0.0128700122	0.0891599127	0.0497866967	-0.1349689379
Cobs_15399.mRNA.1.E009	-0.0878909571	-0.0486980806	0.0815449116	0.0082752767
Cobs_15543.mRNA.1.E003	-0.0067243603	-0.0766635768	-0.0047438577	-0.0014739423
Cobs_15576.mRNA.1.E005	0.0425674173	-0.0976250101	-0.0085942852	-0.0690836411
Cobs_15576.mRNA.1.E006	0.0506267813	-0.0893476115	-0.0024390048	-0.0673246445
Cobs_15576.mRNA.1.E009	0.0923302600	-0.0640820382	-0.0280944917	0.0154278095
Cobs_15676.mRNA.1.E011	-0.0569739426	-0.1196721816	-0.0208955217	-0.1170061469
Cobs_15702.mRNA.1.E002	-0.1082999595	-0.0393847267	-0.0916959409	0.0727991730
Cobs_15702.mRNA.1.E006	-0.1147517773	-0.0230736623	-0.0645811567	0.0604098802
Cobs_15702.mRNA.1.E010	0.0283154402	-0.0659191487	-0.0635084916	0.1234883797
Cobs_15807.mRNA.1.E013	-0.0753479453	-0.0381186652	0.0992682890	-0.0221494212
Cobs_15950.mRNA.1.E006	0.0508592781	0.1375654070	0.0513814211	-0.0652501140
Cobs_16311.mRNA.1.E005	-0.0738887814	0.0108580241	-0.0115083660	-0.0228163669
Cobs_16333.mRNA.1.E003	-0.0419501532	-0.0092349047	-0.0747171140	-0.0656606478
Cobs_16333.mRNA.1.E004	-0.0475158788	-0.0117214955	-0.0606550780	-0.0621935787
Cobs_16333.mRNA.1.E005	-0.0389374725	-0.0119939643	-0.0632989341	-0.0793198297
Cobs_16333.mRNA.1.E006	-0.0438658929	-0.0159963561	-0.0601022561	-0.0670369262
Cobs_16635.mRNA.1.E003	0.0654048947	0.0324086691	-0.0606894379	-0.0046155695
Cobs_16641.mRNA.1.E003	0.0769809357	0.0456196250	0.0954157888	0.0460784660
Cobs_17223.mRNA.1.E006	-0.0250521090	-0.1364130330	0.0649810234	-0.0756715369
Cobs_17408.mRNA.1.E003	0.0900575275	0.0669644348	0.0340965154	-0.0322146250
Cobs_17408.mRNA.1.E004	0.0612223571	0.1108630300	-0.0415525324	-0.0822979169
Cobs_17408.mRNA.1.E005	0.0566992016	0.1111408883	-0.0006394764	-0.0269494382
Cobs_17408.mRNA.1.E007	-0.1235932661	-0.0116278414	-0.0542794849	0.0499657212
Cobs_17408.mRNA.1.E010	-0.1123484872	-0.0646543068	-0.0628456217	0.0062875970
Cobs_17852.mRNA.1.E005	0.0842531564	0.0077489791	0.0052408443	0.1328759123
Cobs_17963.mRNA.1.E013	0.0968118245	-0.0270663682	-0.0302395941	-0.0423850511
Cobs_18218.mRNA.1.E010	0.0905119225	0.0967837180	-0.0423275381	0.0385708616

**Table S3.9:** Gene Ontology (GO) terms enriched in candidate genes for co-option into female or male morph differentiation, as revealed by topGO (Alexa & Rahnenfuhrer, 2010).

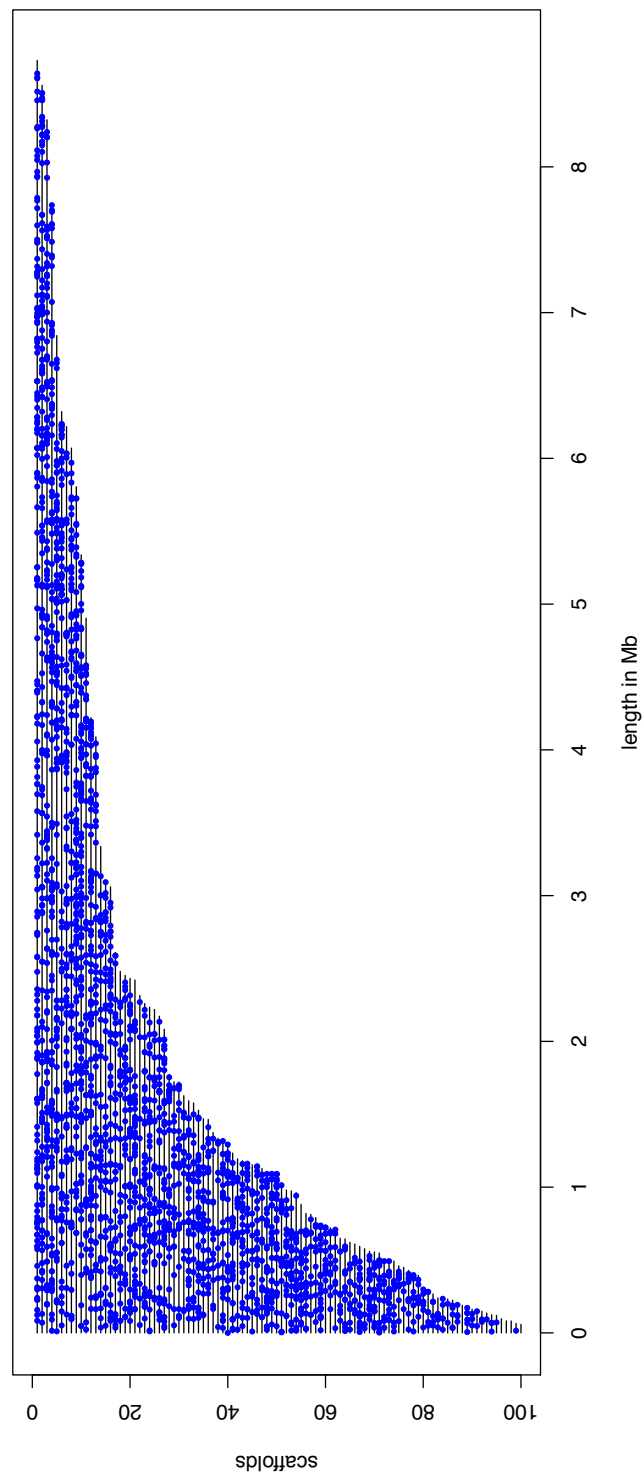
GO-ID	Term	Annotated	Significant	Expected	p-Value
GO:0036125	fatty acid beta-oxidation multienzyme complex	1	1	0	0.0015
GO:0051301	cell division	176	3	0.19	0.0019
GO:0016507	mitochondrial fatty acid beta-oxidation multienzyme complex	1	1	0	0.0061
GO:0051642	centrosome localization	1	1	0	0.0119
GO:0051383	kinetochore organization	3	1	0	0.0157
GO:0030037	actin filament reorganization involved in cell cycle	2	1	0	0.0178
GO:0048132	female germ-line stem cell asymmetric division	3	1	0	0.0191
GO:0030723	ovarian fusome organization	5	1	0.01	0.0192
GO:0045478	fusome organization	7	1	0.01	0.0192
GO:0007282	cystoblast division	4	1	0	0.0204
GO:0032403	protein complex binding	65	1	0.06	0.025
GO:0090527	actin filament reorganization	2	1	0	0.029
GO:0042537	benzene-containing compound metabolic process	13	1	0.01	0.0296
GO:0022414	reproductive process	277	2	0.3	0.0336
GO:0006807	nitrogen compound metabolic process	2585	6	2.81	0.0341
GO:0003006	developmental process involved in reproduction	211	2	0.23	0.0347
GO:0034440	lipid oxidation	10	1	0.01	0.0355
GO:0051293	establishment of spindle localization	11	1	0.01	0.0369
GO:0051295	establishment of meiotic spindle localization	2	1	0	0.0377
GO:0019482	beta-alanine metabolic process	11	1	0.01	0.0399
GO:0030261	chromosome condensation	16	1	0.02	0.0399
GO:0042078	germ-line stem cell division	8	1	0.01	0.0406
GO:0000922	spindle pole	13	1	0.02	0.0411
GO:0045298	tubulin complex	30	1	0.04	0.0428
GO:0022402	cell cycle process	224	2	0.24	0.0463

# Supplement Chapter 4

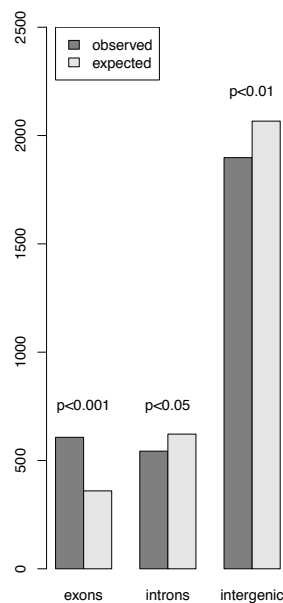
## Supplementary Figures



**Figure S4.1: Multi-family design of the cross used for generating the F2 male mapping population for linkage mapping.** F2 males emerging from five different F1 queens, which were daughters of four different parental queens, were used for linkage mapping. Parental queens were daughters of one highly inbred Japanese (JP, red) population and all of them were mated to the same single male male from Brazil (BR, blue). F1 queens and F2 males are hybrid JP-BR, indicated by the magenta colouring. The following individuals were included in the analysis: P♀#4, P♂#1, F1♀#96, F1♀#252, and 85 F2 males: 17 of family 21, 18 of family 96, 21 of family 152, 16 of family 204, and 13 of family 252.

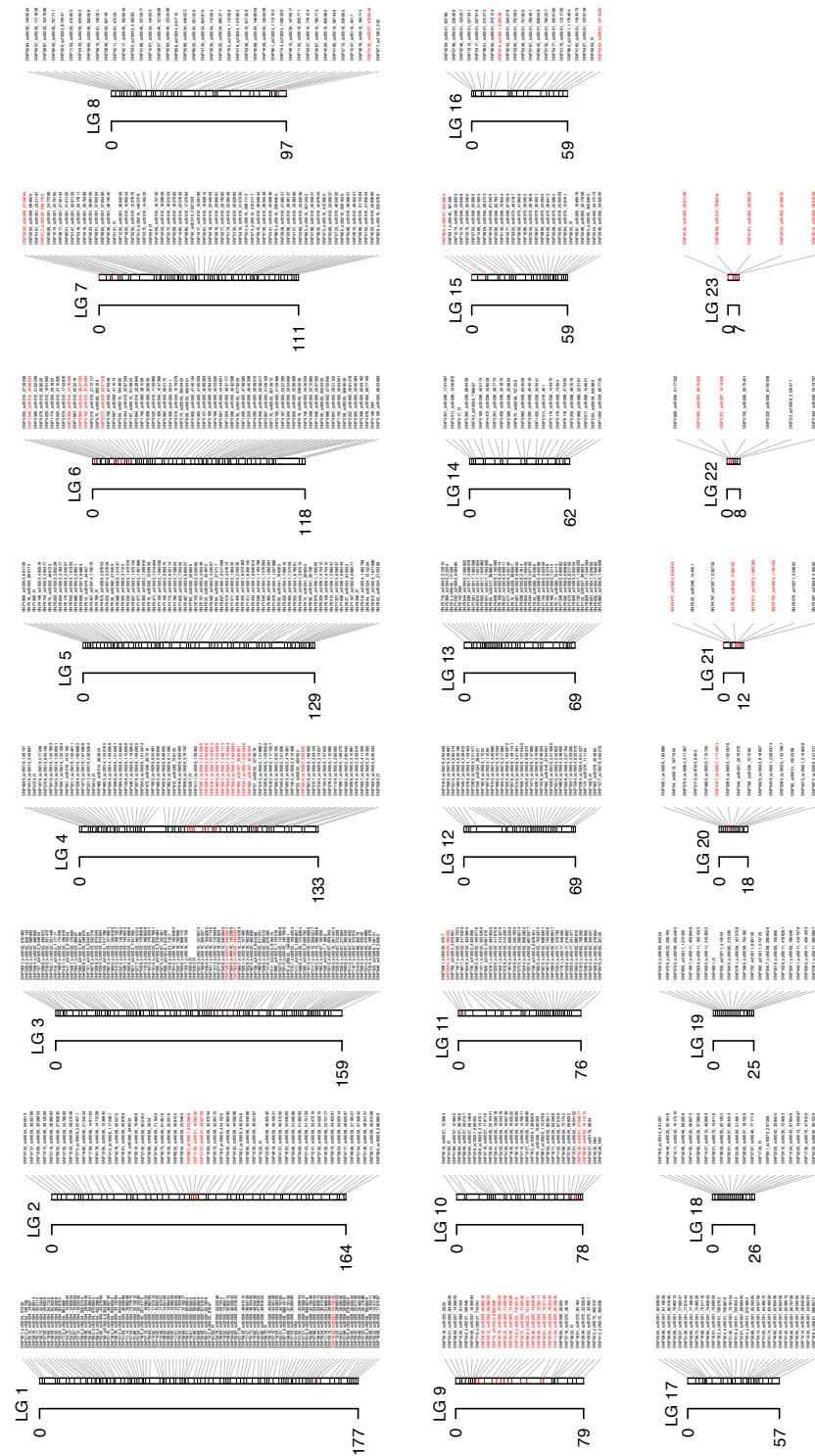


**Figure S4.2:** Distribution of markers over the 100 largest scaffolds of the *C. obscurior* genome assembly version Cobs 1.4 (Schrader *et al.*, 2014). Each blue point represents one marker of the linkage map. The plot shows that all of the 95 largest scaffolds are equally covered with markers.

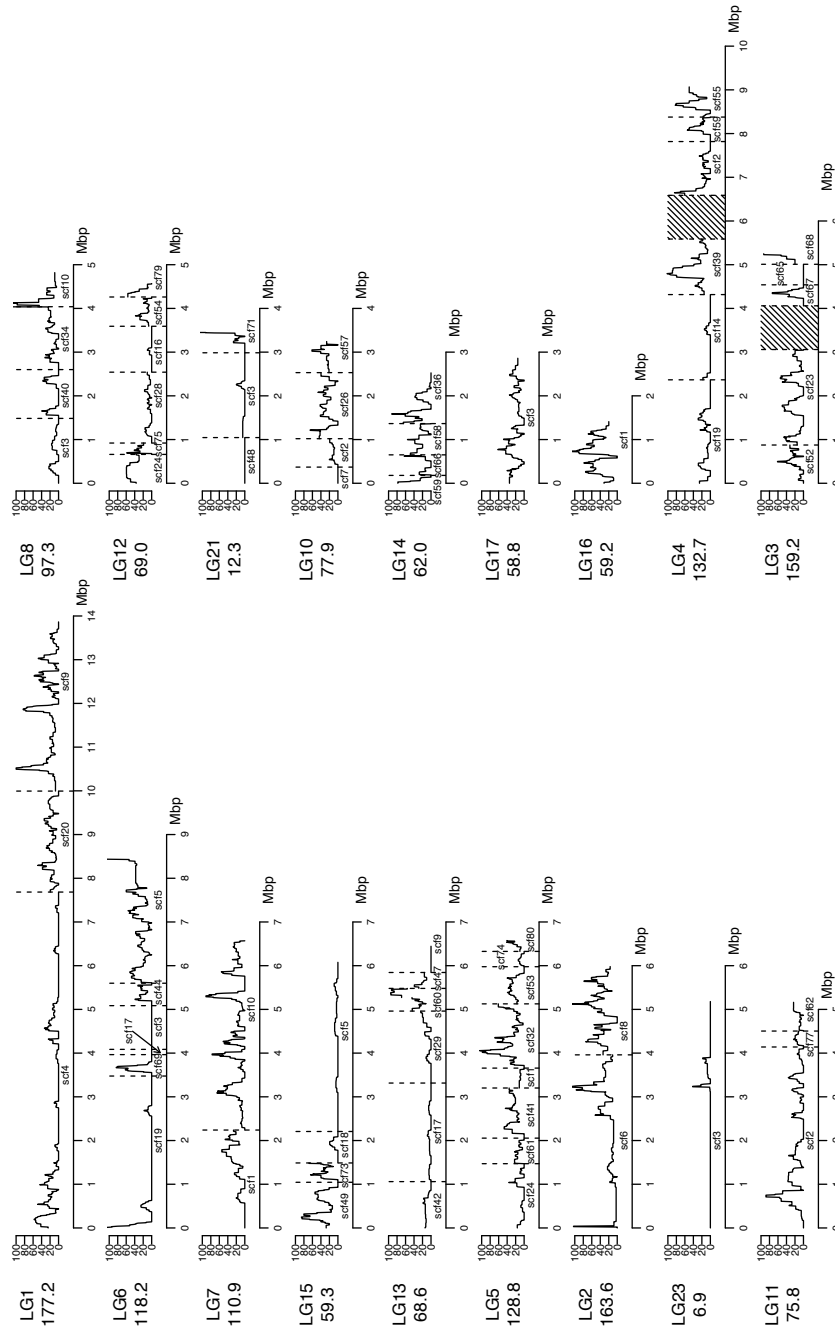


**Figure S4.3: Barplot of the observed and expected number of markers in exons, introns and intergenic regions.** Significant differences between observed and expected valued were calculated with Chi-square tests.





**Figure S4.4: Linkage Map with marker names.** Y-axes show size of each linkage group (LG) in cM. LGs are sorted by genetic size. Red markers indicate markers in TE islands as defined by (Schrader *et al.*, 2014). Linkage groups which are completely inconsistent with the physical assembly are indicated by “NR”, linkage groups which are partly inconsistent with the physical assembly are indicated by “(NR)”.



**Figure S4.5: Local recombination rates plotted along those linkage groups for which genetic map and physical assembly were consistent.** Linkage groups (LG) are ordered by physical size in Mb. Numbers below LGs indicate size in cM. Dashed lines on LG 3 and LG 4 indicate regions where recombination rates could not be estimated because genetic map and physical assembly were inconsistent. Hence the physical size of these linkage groups is unknown.

## Supplementary Tables

**Table S4.1:** Observed and expected number of markers in exons, introns, and intergenic regions.

	percentage in genome	number of expected markers (e)	number of observed markers (o)	o/e
exons	11.8 %	360	607	1.69
introns	20.4 %	622	543	0.87
intergenic	67.8 %	2067	1898	0.92

Table S4.2: Regions used for calculation of local recombination rates.

LG	no. of discarded markers	no. of markers before/after	length of LG in cM before/after	scaffolds	no. of markers	physical position of first and last marker	physical size in Mb	mean no. of markers per window
1	11	279/268	177.2/178.8	scf0004	135	14,627 - 7,737,124	7.7	2.71
				scf0009	83	1,541,734 - 5,324,348	3.8	3.09
				scf0020	50	81,897 - 2,315,775	2.2	3.03
2	16	93/77	163.6/163.6	scf0006	39	51,222 - 3,961,220	3.9	1.93
				scf0008	38	4,034,137 - 5,969,806	1.9	2.77
3	166	279/113	159.2/84.5	scf0052	22	115,046 - 910,365	0.8	3.22
				scf0023	49	98,967 - 2,216,484	2.1	3.12
				middle part not supported by physical assembly				
				scf0067	25	6,246 - 489,880	0.5	4.92
				scf0065	12	23,349 - 470,888	0.4	3.04
				scf0068	5	30,579 - 245,749	0.2	1.88
4	10	221/211	132.7/157.3	scf0019	44	109,746 - 2,404,655	2.3	2.75
				scf0014	27	100,664 - 1,965,349	1.9	2.28
				scf0039	28	33,724 - 1,316,391	1.3	2.96
				middle part not supported by physical assembly				
				scf0002	25	130,229 - 1,276,678	1.1	2.85
				scf0059	20	18,095 - 567,794	0.5	4.16
5	18	172/154	128.8/133.5	scf0055	20	57,440 - 695,533	0.6	3.54
				scf0024	29	741,206 - 2,129,388	1.4	2.82
				scf0061	10	162,560 - 693,300	0.5	2.27
				scf0041	32	20,758 - 1,165,788	1.1	3.50
				scf0001	6	8,273,383 - 8,640,814	0.4	1.85

Table S4.2: (continued) Regions used for calculation of local recombination rates.

LG	no. of discarded markers	no. of markers before/after	length of LG in cM before/after	scaffolds	no. of markers	physical position of first and last marker	physical size in Mb	mean no. of markers per window
6	14	208/194	118.2/118.2	scf0032	36	97,973 – 1,473,319	1.4	3.30
				scf0080	7	37,150 – 297,853	0.3	2.67
				scf0074	9	9,631 – 356,503	0.3	2.72
				scf0053	24	15,766 – 855,362	0.8	3.49
7	26	167/141	110.9/127.8	scf0012	77	775,980 – 4,194,084	3.4	3.13
				scf0069	16	91,877 – 527,913	0.4	3.64
				scf0017	5	35,214 – 128,756	0.1	2.92
				scf0003	19	5,371,556 – 6,280,657	0.9	2.68
				scf0044	12	698,879 – 1,160,875	0.5	2.75
				scf0005	65	3,911,708 – 6,677,193	2.8	3.20
7	26	167/141	110.9/127.8	scf0001	36	1,775,554 – 3,929,085	2.2	2.51
				scf0010	105	1,026,886 – 5,283,660	4.3	3.36
8	16	104/88	97.3/69.0	scf0033	22	95,749 – 1,499,534	1.4	2.33
				scf0040	24	229,180 – 1,292,235	1.1	2.88
				scf0034	34	103,189 – 1,476,517	1.4	3.18
				scf0010	8	189,715 – 877,359	0.7	1.76
9				not supported by physical assembly				
10	22	80/58	77.9/75.3	scf0007	7	2527865 – 2807078	0.3	2.32
				scf0002	18	684,423 – 2,004,894	1.3	3.42
				scf0026	23	688,295 – 2,134,791	1.4	2.35
				scf0057	10	123,684 – 778,716	0.7	2.07

Table S4.2: (continued) Regions used for calculation of local recombination rates.

LG	no. of discarded markers	no. of markers before/after	length of LG in cM before/after	scaffolds	no. of markers	physical position of first and last marker	physical size in Mb	mean no. of markers per window
11	10	112/102	75.8/81.0	scf0002 scf0062 scf0077	82 10 10	3,565,185 – 7,613,105 32,120 – 708,290 23,178 – 404,791	4.0 0.7 0.4	2.93 2.13 2.95
12	4	103/99	60.0/73.3	scf0024 scf0075 scf0028 scf0016 scf0054 scf0079	11 9 29 23 19 8	11,084 – 665,214 223,837 – 420,666 182,598 – 1,753,775 1,997,335 – 2,950,889 14,923 – 679,843 111,194 – 365,213	0.7 0.2 1.6 1.0 0.7 0.3	2.22 3.33 2.56 2.98 3.31 2.71
13	9	103/94	68.6/74.6	scf0042 scf0017 scf0029 scf0060 scf0047 scf0009	15 30 21 15 6 7	22,039 – 1,064,372 388,809 – 2,588,378 51,814 – 1,678,505 287,183 – 724,113 784,683 – 1,086,011 979,022 – 1,484,608	1.0 2.2 1.6 0.4 0.3 0.5	2.21 2.22 2.13 2.90 2.00 1.82
14	8	54/46	62.0/61.3	scf0036 scf0058 scf0059 scf0066	23 10 5 8	269,894 – 1,331,109 74,354 – 740,636 607,720 – 736,513 132,258 – 515,198	1.1 0.7 0.1 0.4	2.79 2.07 2.30 2.25
15	10	130/120	59.3/62.5	scf0005 scf0018 scf0049	51 17 32	10,639 – 3,875,743 1,635,046 – 2,267,455 57,944 – 1,091,394	3.9 0.6 1.0	2.26 2.61 3.73

Table S4.2: (continued) Regions used for calculation of local recombination rates.

LG	no. of discarded markers	no. of markers before/after	length of LG in cM before/after	scaffolds	no. of markers	physical position of first and last marker	physical size in Mb	mean no. of markers per window
				scf0073	20	89,175 – 492,187	0.4	4.58
16	10	47/37	59.2/65.3	scf0001	37	83,796 – 1,414,280	1.3	3.46
17	0	59/59	58.8/58.9	scf0001	59	5,490,361 – 8,263,805	2.8	2.99
18				not supported by physical assembly				
19				not supported by physical assembly				
20				not supported by physical assembly				
21	9	95/86	12.3/12.6	scf0048	35	20,682 – 1,090,721	1.1	3.91
				scf0003	32	684,423 – 2,537,902	1.9	2.54
				scf0071	19	2,451 - 498,293	0.5	3.98
22				not supported by physical assembly				
23	16	54/38	6.9/9.7	scf0003	38	195,040 – 5,286,652	5.1	1.69

**Table S4.3:** Gene ontology (GO) terms enriched in “hotspots” (recombination rate > 50 cM/Mb) as revealed by topGO (Alexa & Rahnenfuhrer, 2010) on a FDR of 0.01.

GO.ID	Term	Annotated	Significant	Expected	p-Value
GO:0048634	regulation of muscle organ development	10	4	0.22	0.00012
GO:1901861	regulation of muscle tissue development	8	3	0.18	0.00046
GO:1901862	negative regulation of muscle tissue development	2	2	0.04	0.00049
GO:0061321	garland nephrocyte differentiation	3	2	0.07	0.00105
GO:0017157	regulation of exocytosis	8	3	0.18	0.00155
GO:0048635	negative regulation of muscle organ development	2	2	0.04	0.00183
GO:0061061	muscle structure development	104	8	2.33	0.00199
GO:0007517	muscle organ development	74	7	1.66	0.00346
GO:0003014	renal system process	5	2	0.11	0.00497
GO:0048137	spermatocyte division	4	2	0.09	0.00499
GO:0007523	larval visceral muscle development	3	2	0.07	0.005
GO:0061318	renal filtration cell differentiation	6	2	0.13	0.00753
GO:0045843	negative regulation of striated muscle tissue development	2	2	0.04	0.00923
GO:0000038	very long-chain fatty acid metabolic processes	4	2	0.09	0.00931

**Table S4.4:** Gene ontology (GO) terms enriched in “coldspots” (recombination rate = 0 cM/Mb) as revealed by topGO (Alexa & Rahnenfuhrer, 2010) on a FDR of 0.01.

GO.ID	Term	Annotated	Significant	Expected	p-Value
GO:0050767	regulation of neurogenesis	71	23	12.02	0.00076
GO:0048839	inner ear development	11	7	1.86	0.00116
GO:0045664	regulation of neuron differentiation	59	21	9.99	0.0012
GO:0060602	branch elongation of an epithelium	3	3	0.51	0.00151
GO:0043583	ear development	11	7	1.86	0.00237
GO:0006357	regulation of transcription from RNA polymerase II promoter	188	47	31.84	0.00251
GO:0071599	otic vesicle development	7	5	1.19	0.00312
GO:0006366	transcription from RNA polymerase II promoter	229	52	38.78	0.00357



**Table S4.4: (continued)** Gene ontology (GO) terms enriched in “coldspots” (recombination rate = 0 cM/Mb) as revealed by topGO (Alexa & Rahnenfuhrer, 2010) on a FDR of 0.01.

GO.ID	Term	Annotated	Significant	Expected	p-Value
GO:0061323	cell proliferation involved in heart morphogenesis	5	4	0.85	0.00359
GO:0045665	negative regulation of neuron differentiation	10	6	1.69	0.00428
GO:0050680	negative regulation of epithelial cell proliferation	6	4	1.02	0.00453
GO:0022612	gland morphogenesis	42	15	7.11	0.00477
GO:0060560	developmental growth involved in morphogenesis	34	13	5.76	0.00478
GO:0035725	sodium ion transmembrane transport	9	5	1.52	0.00479
GO:0090288	negative regulation of cellular response to growth factor stimulus	3	3	0.51	0.00485
GO:0021511	spinal cord patterning	3	3	0.51	0.00503
GO:0060284	regulation of cell development	96	27	16.26	0.00554
GO:0051960	regulation of nervous system development	104	29	17.61	0.00588
GO:0030916	otic vesicle formation	3	3	0.51	0.00605
GO:0001658	branching involved in ureteric bud morphogenesis	4	3	0.68	0.00657
GO:0021513	spinal cord dorsal/ventral patterning	3	3	0.51	0.00677
GO:2000136	regulation of cell proliferation involved in heart morphogenesis	5	4	0.85	0.00682
GO:0031016	pancreas development	8	5	1.35	0.00683
GO:0009889	regulation of biosynthetic process	752	142	127.36	0.00719
GO:0048709	oligodendrocyte differentiation	5	4	0.85	0.00748
GO:0090190	positive regulation of branching involved in ureteric bud morphogenesis	3	3	0.51	0.00756
GO:0060512	prostate gland morphogenesis	3	3	0.51	0.00759
GO:0035272	exocrine system development	45	15	7.62	0.00788

**Table S4.4: (continued)** Gene ontology (GO) terms enriched in “coldspots” (recombination rate = 0 cM/Mb) as revealed by topGO (Alexa & Rahnenfuhrer, 2010) on a FDR of 0.01.

GO.ID	Term	Annotated	Significant	Expected	p-Value
GO:0031326	regulation of cellular biosynthetic process	749	141	126.85	0.00788
GO:0042490	mechanoreceptor differentiation	3	3	0.51	0.008
GO:0042472	inner ear morphogenesis	10	6	1.69	0.00808
GO:0009612	response to mechanical stimulus	11	6	1.86	0.00851
GO:0035212	cell competition in a multicellular organism	6	4	1.02	0.00869
GO:0015672	monovalent inorganic cation transport	145	28	24.56	0.00879
GO:0070371	ERK1 and ERK2 cascade	3	3	0.51	0.00887
GO:0098602	single organism cell adhesion	23	9	3.9	0.00894
GO:0048708	astrocyte differentiation	3	3	0.51	0.00903
GO:0048568	embryonic organ development	53	17	8.98	0.00958
GO:0090183	regulation of kidney development	3	3	0.51	0.00978

**Table S4.5: Results of correlations between estimated recombination rates and different genomic features as revealed by generalized linear models (GLM) of the quasipoisson family.** GC content was calculated with the R package seqinr Charif:2007cr. For genomic elements (genes, exons, introns, repeats, SNVs) the number of elements and the number of bases per 100 kb windows was counted and correlated to recombination rate in these windows with GLMs of the quasipoisson family. Asterices indicate significances based on the following FDRs: (\*): 0.1, \*: 0.05, \*\*: 0.01, \*\*\*: 0.001.

	genomic features	overall					LDRs						
		df	dev.	res.df	res.dev.	F	p	df	dev.	res.df	F	p	
GC content	GC content overall	1	7.373	987	20533	0.3215	0.571	1	2.233	961	19939	0.0983	0.754
	GC content in exons	1	0.015	1001	5.479	2.7766	0.095 (*)	1	0.007	952	4.7382	1.4267	0.232
	GC content in introns	1	0.002	1000	7.3667	0.2752	0.600	1	<0.001	951	6.8694	0.0058	0.939
	GC content in intergenic regions	1	34.236	1009	22024	1.3155	0.252	1	4.935	960	19923	0.2173	0.641
gene content	gene number / 100 kb	1	6.042	977	20376	0.2635	0.608	1	8.230	951	19771	0.3629	0.547
	content bases / 100 kb	1	6.013	977	20376	0.2604	0.610	1	0.650	951	19779	0.0285	0.866
	exon number/ 100 kb	1	11.899	978	20370	0.5154	0.473	1	4.063	952	19775	0.1781	0.673
	content bases / 100 kb	1	18.349	978	20364	0.7944	0.373	1	10.948	952	19768	0.4794	0.489
	intron number / 100 kb	1	14.947	977	20338	0.6475	0.421	1	5.425	951	19745	0.2379	0.626
content bases / 100 kb	1	2.2893	977	20351	0.0995	0.753	1	0.238	951	19750	0.0105	0.919	
transposable element & repeat content	repeats number / 100 kb	1	173.940	987	20366	7.6453	<0.01 **	1	158.980	961	19782	7.0546	<0.01 **
	overall bases / 100 kb	1	228.850	987	20311	10.088	<0.01 **	1	1.776	961	19939	0.0781	0.780
	classI number / 100 kb	1	0.542	987	20540	0.0236	0.878	1	43.501	961	19898	1.9113	0.167
	bases / 100 kb	1	373.770	987	20166	16.603	<0.001 ***	1	49.965	961	19891	2.2029	0.138
	LINE number / 100 kb	1	32.923	987	20507	1.4316	0.232	1	86.385	961	19855	3.7934	0.0518 (*)
	bases / 100 kb	1	130.530	987	20410	5.7322	<0.05 *	1	4.030	961	19937	0.1773	0.674
	LTR number / 100 kb	1	3.026	987	20537	0.1317	0.717	1	15.308	961	19926	0.6731	0.412
	bases / 100 kb	1	374.390	987	20166	16.549	<0.001 ***	1	52.503	961	19889	2.3121	0.129
	classII number / 100 kb	1	25.104	987	20515	1.0981	0.295	1	1.638	961	19940	0.0721	0.788
	bases / 100 kb	1	145.130	987	20395	6.4104	<0.05 *	1	12.480	961	19929	0.5508	0.458

**Table S4.5: (continued) Results of correlations between estimated recombination rates and different genomic features as revealed by generalized linear models (GLM) of the quasipoisson family.** GC content was calculated with the R package seqinr Charif:2007cr. For genomic elements (genes, exons, introns, repeats, SNVs) the number of elements and the number of bases per 100 kb windows was counted and correlated to recombination rate in these windows with GLMs of the quasipoisson family. Asterices indicate significances based on the following FDRs: (\*): 0.1, \*: 0.05, \*\*: 0.01, \*\*\*, 0.001.

genomic features		overall					LDRs						
		df	dev.	res.df	res.dev.	F	p	df	dev.	res.df	res.dev.	F	p
TIR	number / 100 kb	1	3.930	987	20536	0.1713	0.679	1	1.537	961	19940	0.0676	0.795
	bases / 100 kb	1	47.592	987	20493	2.0818	0.149	1	4.097	961	19937	0.1804	0.671
SSR	number / 100 kb	1	428.620	987	20112	19.158	<0.001 ***	1	236.580	961	19704	10.59	<0.01 **
	bases / 100 kb	1	100.900	987	20439	4.4418	<0.05 *	1	139.740	961	19801	6.2299	<0.05 *
SNV content	number of SNPs / 100 kb	1	1.256	987	20539	0.0547	0.815	1	279.460	961	19662	12.518	<0.001 ***
	number of homozygous SNPs / 100 kb	1	556.990	987	19983	24.521	<0.001 ***	1	485.320	961	19456	21.540	<0.001 ***
	number of heterozygous SNPs / 100 kb	1	62.297	987	20478	2.7007	0.101	1	37.355	961	19904	1.6553	0.199
	number of Indels / 100 kb	1	902.64	987	19638	41.433	<0.001 ***	1	911.200	961	19030	42.321	<0.001 ***

**Table S4.6:** Gene ontology (GO) terms enriched in non-supported regions (NR) as revealed by topGO (Alexa & Rahnenfuhrer, 2010) on a FDR of 0.01.

GO.ID	Term	Annotated	Significant	Expected	p-Value
GO:0098590	plasma membrane region	34	6	0.89	0.00048
GO:0016324	apical plasma membrane	15	4	0.39	0.00088
GO:0045202	synapse	146	11	3.84	0.00147
GO:0050896	response to stimulus	1897	66	48.38	0.0019
GO:0005856	cytoskeleton	388	14	10.21	0.00214
GO:0030054	cell junction	116	9	3.05	0.00325
GO:0030029	actin filament-based process	112	9	2.86	0.0035
GO:0022803	passive transmembrane transporter activity	183	11	4.9	0.0045
GO:0000003	reproduction	277	15	7.07	0.0045
GO:0005515	protein binding	2581	89	69.09	0.0053
GO:0022412	cellular process involved in reproduction in multicellular organisms	191	12	4.87	0.0059
GO:0005216	ion channel activity	182	11	4.87	0.0067
GO:0035265	organ growth	28	4	0.71	0.0069
GO:0030036	actin cytoskeleton organization	111	9	2.83	0.0069
GO:0023052	signaling	1565	54	39.92	0.0074
GO:0007010	cytoskeleton organization	250	13	6.38	0.0087
GO:0003006	developmental process involved in reproduction	211	12	5.38	0.0088
GO:0045254	pyruvate dehydrogenase complex	7	2	0.18	0.00938
GO:0051716	cellular response to stimulus	1612	56	41.11	0.0095

**Table S4.7:** Observed (obs.) and expected (exp.) number of structural variants falling within supported regions (SR) and not supported regions (NR).

	inversion			deletions			duplications		
	obs.	exp.	o/e	obs.	exp.	o/e	obs.	exp.	o/e
NR	221	232	0.95	344	263	1.31	296	237	1.25
SR	608	597	1.02	596	677	0.88	552	611	0.90

# Bibliography

- Abdul Rahman N., Parks D. H., Willner D. L., Engelbrektson A. L., Goffredi S. K., Warnecke F., Scheffrahn R. H. & Hugenholtz P. (2015). A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes. *Microbiome* **3**(1): 5.
- Abrusán G. & Krambeck H.-J. (2006). The distribution of L1 and Alu retroelements in relation to GC content on human sex chromosomes is consistent with the ectopic recombination model. *J. Mol. Evol.* **63**(4): 484–492.
- Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M. & Aksoy S. (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat. Genet.* **32**(3): 402–407.
- Aksoy S. (2000). Tsetse - A haven for microorganisms. *Parasitol. Today* **16**(3): 114–118.
- Aksoy S. & Rio R. V. M. (2005). Interactions among multiple genomes: Tsetse, its symbionts and trypanosomes. *Insect Biochem. Mol. Biol.* **35**(7): 691–698.
- Alasaad S., Rossi L., Maione S., Sartore S., Soriguer R. C., Pérez J. M., Rasero R., Zhu X. Q. & Soglia D. (2008). HotSHOT Plus ThermalSHOCK, a new and efficient technique for preparation of PCR-quality mite genomic DNA. *Parasitol. Res.* **103**(6): 1455–1457.
- Alexa A. & Rahnenfuhrer J. (2010) topGO: Enrichment analysis for Gene Ontology. R package version 2.18.0.
- Alexander R. D. & Sherman P. W. (1977). Local mate competition and parental investment in social insects. *Science* **196**(4289): 494–500.
- Alvarado S., Rajakumar R., Abouheif E. & Szyf M. (2015). Epigenetic variation in the *Egfr* gene generates quantitative variation in a complex trait in ants. *Nat. Comms.* **6**: 6513.
- Alves J. M. P., Klein C. C., da Silva F. M., Costa-Martins A. G., Serrano M. G., Buck G. A., Vasconcelos A. T. R., Sagot M.-F., Teixeira M. M. G., Motta M. C. M. & Camargo E. P. (2013). Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers. *BMC Evol. Biol.* **13**(1): 190.

- Amann R. I., Binder B. J., Olson R. J., Chisholm S. W., Devereux R. & Stahl D. A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56**(6): 1919–1925.
- Anders S., Reyes A. & Huber W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**(10): 2008–2017.
- Anders S., Pyl P. T. & Huber W. (2015). HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**(2): 166–169.
- Andersen S. O. (2012). Cuticular sclerotization and tanning. In: Gilbert L. I. (ed), *Insect molecular biology and biochemistry*. Academic Press: Waltham, MA, USA, pp. 167–192.
- Andersen S. B., Boye M., Nash D. R. & Boomsma J. J. (2012). Dynamic *Wolbachia* prevalence in *Acromyrmex* leaf-cutting ants: potential for a nutritional symbiosis. *J. Evol. Biol.* **25**(7), 1340–1350.
- Arbeithuber B., Betancourt A. J., Ebner T. & Tiemann-Boege I. (2015). Crossovers are associated with mutation and biased gene conversion at recombination hotspots. *Proc. Natl. Acad. Sci. USA* **112**(7): 2109–2114.
- Baker B. S. & Wolfner M. F. (1988) A molecular analysis of *doublesex*, a bifunctional gene that controls both male and female sexual differentiation in *Drosophila melanogaster*. *Genes Dev.* **2**: 477–489.
- Baldo L., Dunning Hotopp J. C., Jolley K. A., Bordenstein S. R., Biber S. A., Choudhury R. R., Hayashi C., Maiden M. C. J., Tettelin H. & Werren J. H. (2006). Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl. Environ. Microbiol.* **72**(11): 7098–7110.
- Balmand S., Lohs C., Aksoy S. & Heddi A. (2013). Tissue distribution and transmission routes for the tsetse fly endosymbionts. *J. Invertebr. Pathol.* **112**(Suppl): S116–S122.
- Bartolomé C., Maside X. & Charlesworth B. (2002). On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol. Biol. Evol.* **19**(6): 926–937.
- Baudat F., Buard J., Grey C., Fledel-Alon A., Ober C., Przeworski M., Coop G. & de Massy B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**: 836–840.
- Baumann P. (2005). Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu. Rev. Microbiol.* **59**(1): 155–189.
- Benevolenskaya E. V., Frolov M. V. & Birchler J. A. (2000). *Krüppel* homolog (*Kr h*) is a dosage-dependent modifier of gene expression in *Drosophila*. *Genet. Res.* **75**(2): 137–142.

- Ben-Yosef M., Pasternak Z., Jurkevitch E. & Yuval B. (2014). Symbiotic bacteria enable olive flies (*Bactrocera oleae*) to exploit intractable sources of nitrogen. *J. Evol. Biol.* **27(12)**: 2695–2705.
- Benjamini Y. & Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57(1)**: 289–300.
- Beye M., Hunt G. J., Page R. E., Fondrk M. K., Grohmann L. & Moritz R. F. (1999). Unusually high recombination rate detected in the sex locus region of the honey bee (*Apis mellifera*). *Genetics* **153(4)**: 1701–1708.
- Beye M., Hasselmann M., Fondrk M. K., Page R. E. & Omholt S. W. (2003). The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. *Cell* **114**: 419–429.
- Beye M., Gattermeier I., Hasselmann M., Gempe T., Schioett M., Baines J. F., Schlipalius D., Mougél F., Emore C., Rueppell O., Sirvioö A., Guzmán-Novoa E., Hunt G., Solignac M. & Page R. E. Jr. (2006). Exceptionally high levels of recombination across the honey bee genome. *Genome Res.* **16(11)**: 1339–1344.
- Boetzer M., Henkel C. V., Jansen H. J., Butler D. & Pirovano W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27(4)**: 578–579.
- Boissinot S., Entezam A. & Furano A. V. (2001). Selection against deleterious LINE-1-containing loci in the human lineage. *Mol. Biol. Evol.* **18(6)**: 926–935.
- Bolger A. M., Lohse M. & Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30(15)**: 2114–2120.
- Bopp D., Saccone G. & Beye M. (2014). Sex determination in insects: variations on a common theme. *Sex. Dev.* **8(1-3)**: 20–28.
- Bordenstein S. R., O’Hara F. P. & Werren J. H. (2001). *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature* **409(6821)**: 707–710.
- Bourke A. F. G. (2011) Principles of social evolution. Oxford series in ecology and evolution. Harvey P. H., May R. M., Godfray C. H. & Dunne J. A. (eds). Oxford University Press, Oxford.
- Brigandt I. & Love A. C. (2012). Conceptualizing evolutionary novelty: Moving beyond definitional debates. *J. Exp. Zool. B Mol. Dev. Evol.* **318(6)**: 417–427.
- Broman K. W., Wu H., Sen S. & Churchill G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19(7)**: 889–890.
- Brucker R. M. & Bordenstein S. R. (2012). Speciation by symbiosis. *Trends Ecol. Evol.* **27(8)**: 443–451.
- Brucker R. M. & Bordenstein S. R. (2013). The hologenomic basis of speciation: Gut bacteria cause hybrid lethality in the genus *Nasonia*. *Science* **341(6146)**: 667–669.



- Brune A. (2014). Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* **12**: 168.
- Buchner P. (1965). Endosymbiosis of animals with plant microorganisms. New York: Interscience.
- Burge S. W., Daub J., Eberhardt R., Tate J., Barquist L., Nawrocki E. P., Eddy S. R., Gardner P. P. & Bateman A. (2013). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* **41**: D226–D232.
- Burtis K. C. & Baker B. S. (1989). *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* **56(6)**: 997–1010.
- Caspi R., Altman T., Dreher K., Fulcher C. A., Subhraveti P., Keseler I. M., Kothari A., Krummenacker M., Latendresse M., Mueller L. A., Ong Q., Paley S., Pujar A., Shearer A. G., Travers M., Weerasinghe D., Zhang P. & Karp P. D. (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **40**: D742–D753.
- Chan, A. H., Jenkins, P. A., & Song, Y. S. (2012). Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLoS Genet.* **8(12)**: e1003090.
- Charif D., & Lobry J. R. (2007). SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Structural approaches to sequence evolution (pp. 207–232). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Charlesworth B. (2009). Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10(3)**: 195–205.
- Charlesworth D. & Charlesworth B. (1995). Transposable elements in inbreeding and outbreeding populations. *Genetics* **140(1)**: 415–417.
- Charlesworth B., Sniegowski P. & Stephan W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371(6494)**: 215–220.
- Chatterjee S. S., Uppendahl L. D., Chowdhury M. A., Ip P.-L. & Siegal M. L. (2011). The female-specific *Doublesex* isoform regulates pleiotropic transcription factors to pattern genital development in *Drosophila*. *Development* **138(6)**: 1099–1109.
- Chen F., Mackey A. J., Vermunt J. K. & Roos D. S. (2007). Assessing performance of orthology detection strategies applied to eukaryotic genomes. *PLoS ONE* **2(4)**: e383.
- Chevreur B., Wetter T. & Suhai S. (1999). Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology. Proceedings of the German Conference on Bioinformatics (GCB)*: 45–56.

- Cho S., Huang Z. Y. & Zhang J. (2007). Sex-specific splicing of the honeybee *doublesex* gene reveals 300 million years of evolution at the bottom of the insect sex-determination pathway. *Genetics* **177**(3): 1733–1741.
- Chu H.Y., Wegel E. & Osbourn A. (2011). From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. *Plant J.* **66**: 66–79.
- Clayton A. L., Oakeson K. F., Gutin M., Pontes A., Dunn D. M., von Niederhausern A. C., Weiss R. B., Fisher M. & Dale C. (2012). A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect–bacterial symbioses. *PLoS Genet.* **8**(11): e1002990.
- Clément Y. & Arndt P. F. (2013). Meiotic recombination strongly influences GC-content evolution in short regions in the mouse genome. *Mol. Biol. Evol.* **30**(12): 2612–2618.
- Clough E., Jimenez E., Kim Y.-A., Whitworth C., Neville M. C., Hempel L. U., Pavlou H. J., Chen Z.-X., Sturgill D., Dale R. K., Smith H. E., Przytycka T. M., Goodwin S. F., Van Doren M. & Oliver B. (2014). Sex- and tissue-specific functions of *Drosophila doublesex* transcription factor target genes. *Dev. Cell* **31**(6): 761–773.
- Cremer S. & Heinze J. (2002). Adaptive production of fighter males: queens of the ant *Cardiocondyla* adjust the sex ratio under local mate competition. *Proc. Biol. Sci.* **269**(1489): 417–422.
- Cremer S. & Heinze J. (2003). Stress grows wings: environmental induction of winged dispersal males in *Cardiocondyla* ants. *Curr. Biol.* **13**(3): 219–223.
- Crozier R. H. (1969). Chromosome number polymorphism in an Australian ponerine ant. *Canadian Journal of Genetics and Cytology.* **11**(2): 333–339.
- Cook J. M. (1993). Sex determination in the Hymenoptera: a review of models and evidence. *Heredity* **71**: 421–435.
- Cordaux R., Bouchon D. & Gréve P. (2011). The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends Genet.* **27**(8): 332–341.
- Dale C. & Welburn S. C. (2001). The endosymbionts of tsetse flies: manipulating host-parasite interactions. *Int. J. Parasitol.* **31**(5-6): 628–631.
- Damiani C., Ricci I., Crotti E., Rossi P., Rizzi A., Scuppa P., Esposito F., Bandi C., Daffonchio D. & Favia G. (2008). Paternal transmission of symbiotic bacteria in malaria vectors. *Curr. Biol.* **18**(23): R1087–8.
- Darwin C. (1872). The origin of species. 6th ed. London: John Murray.
- Davis L. & Smith G. R. (2001). Meiotic recombination and chromosome segregation in *Schizosaccharomyces pombe*. *Proc. Natl. Acad. Sci.* **98**(15): 8395–8402.

- Desjardins C. A., Gadau J., Lopez J. A., Niehuis O., Avery A. R., Loehlin D. W., Richards S., Colbourne J. K. & Werren J. H. (2013). Fine-scale mapping of the *Nasonia* genome to chromosomes using a high-density genotyping microarray. *G3* **3**(2): 205–215.
- Dobin A., Davis C. A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M. & Gingeras T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**(1): 15–21.
- Dohlen C. D. von, Kohler S., Alsop S. T. & McManus W. R. (2001). Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. *Nature* **412**(6845): 433–436.
- Dolgin E. S. & Charlesworth B. (2008). The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* **178**(4): 2169–2177.
- Domselaar G. H. Van, Stothard P., Shrivastava S., Cruz J. A., Guo A., Dong X., Lu P., Szafron D., Greiner R. & Wishart D. S. (2005). BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Res.* **33**: W455–W459.
- Doorn G. S. van (2014). Evolutionary transitions between sex-determining mechanisms: A review of theory. *Sex. Dev.* **8**(1-3): 7–19.
- Doorn G. S. van & Kirkpatrick M. (2007). Turnover of sex chromosomes induced by sexual conflict. *Nature* **449**(7164): 909–912.
- Doorn G. S. van & Kirkpatrick M. (2010). Transitions between male and female heterogamety caused by sex-antagonistic selection. *Genetics* **186**(2): 629–645.
- Douglas A. E. (2006). Phloem-sap feeding by animals: problems and solutions. *J. Exp. Bot.* **57**(4): 747–754.
- Douglas A. E. (2009). The microbial dimension in insect nutritional ecology. *Functional Ecology* **23**(1): 38–47.
- Douglas A. E. (2010). The symbiotic habit. New Jersey: Princeton University Press.
- Dumont B. L. & Payseur B. A. (2011). Genetic analysis of genome-scale recombination rate evolution in house mice. *PLoS Genet.* **7**(6): e1002116.
- Duret L., & Arndt P. F. (2008). The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet.* **4**(5): e1000071.
- Duret L., Marais G. & Biémont C. (2000). Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*. *Genetics* **156**(4): 1661–1669.
- Edgar R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**(5): 1792–1797.
- Engel P. & Moran N. A. (2013). The gut microbiota of insects - diversity in structure and function. *FEMS Microbiol. Rev.* **37**(5): 699–735.

- Engelstädter J. & Hurst G. D. D. (2006). Can maternally transmitted endosymbionts facilitate the evolution of haplodiploidy? *J. Evol. Biol.*, **19**(1): 194–202.
- Engelstädter J. & Telschow A. (2014). Cytoplasmic incompatibility and host population structure. *Heredity* **103**(3): 196–207.
- Erickson J. W. & Quintero J. J. (2007). Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in *Drosophila*. *PLoS Biol.* **5**: e332.
- Erwin D. H. (2012). Novelties that change carrying capacity. *J. Exp. Zool. B Mol. Dev. Evol.* **318**(6): 460–465.
- Eirín-López J. M. & Sánchez L. (2015). The comparative study of five sex-determining proteins across insects unveils high rates of evolution at basal components of the sex determination cascade. *Dev. Genes Evol.* **225**(1): 23–30.
- Feldhaar H., Straka J., Krischke M., Berthold K., Stoll S., Mueller M. J. & Gross R. (2007). Nutritional upgrading for omnivorous carpenter ants by the endosymbiont *Blochmannia*. *BMC Biol.* **5**: 48.
- Feldhaar H. (2011). Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecological Entomology* **36**(5): 533–543.
- Fischer O. & Schmid-Hempel P. (2005). Selection by parasites may increase host recombination frequency. *Biol. Lett.* **1**(2): 193–195.
- Fujii T. & Shimada T. (2007). Sex determination in the silkworm, *Bombyx mori*: A female determinant on the W chromosome and the sex-determining gene cascade. *Semin. Cell. Dev. Biol.* **18**(3): 379–388.
- Fullerton S. M., Bernardo Carvalho A. & Clark A. G. (2001). Local rates of recombination are positively correlated with GC content in the human genome. *Mol. Biol. Evol.* **18**(6): 1139–1142.
- Gadau J. (2009). DNA isolation from ants. *Cold Spring Harbor Protocols* **7**: 5245.
- Gadau J., Gerloff C. U., Krüger N., Chan H., Schmid-Hempel P., Wille A. & Page R. E. (2001). A linkage analysis of sex determination in *Bombus terrestris* (L.) (Hymenoptera: Apidae). *Heredity* **87**(2): 234–242.
- Gempe T. & Beye, M. (2010). Function and evolution of sex determination mechanisms, genes and pathways in insects. *BioEssays* **33**(1): 52–60.
- Gempe T., Hasselmann M., Schiøtt M., Hause G., Otte M. & Beye M. (2009). Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway. *PLoS Biol.* **7**(10): e1000222.
- Gerhart J. & Kirschner M. (2007). The theory of facilitated variation. *Proc. Natl. Acad. Sci. USA* **104**(1): 8582–8589.

- Geuverink E. & Beukeboom L. W. (2014). Phylogenetic distribution and evolutionary dynamics of the sex determination genes *doublesex* and *transformer* in insects. *Sex. Dev.* **8**(1-3): 38–49.
- Gil R., Sabater-Munoz B., Latorre A., Silva F. J. & Moya, A. (2002). Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. *Proc. Natl. Acad. Sci. USA* **99**(7): 4454–4458.
- Gil R., Silva F. J., Zientz E., Delmotte F., González-Candela F., Latorre A., Rausell C., Kamerbeek J., Gadau J., Hölldobler B., van Ham R. C., Gross R. & Moya A. (2003). The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proc. Natl. Acad. Sci. USA* **100**(16): 9388–9393.
- Gil R., Silva F. J., Peretó J. & Moya A. (2004). Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* **68**(3): 518–537.
- Gil R., Latorre A. & Moya A. (2010). Evolution of prokaryote-animal symbiosis from a genomics perspective. *Microbiology Monographs* **19**: 207–233.
- Giraut L., Falque M., Drouaud J., Pereira L., Martin O. C. & Mézard C. (2011). Genome-wide crossover distribution in *Arabidopsis thaliana* meiosis reveals sex-specific patterns along chromosomes. *PLoS Genet.* **7**(11): e1002354.
- Gómez-Valero L., Soriano-Navarro M., Pérez-Brocal V., Heddi A., Moya A., García-Verdugo J. M. & Latorre A. (2004). Coexistence of *Wolbachia* with *Buchnera aphidicola* and a secondary symbiont in the aphid *Cinara cedri*. *J. Bacteriol.* **186**(19): 6626–6633.
- Gotoh T., Noda H. & Ito S. (2006). *Cardinium* symbionts cause cytoplasmic incompatibility in spider mites. *Heredity* **98**(1): 13–20.
- Gotoh H., Miyakawa H., Ishikawa A., Ishikawa Y., Sugime Y., Emlen D. J., Lavine L. C. & Miura T. (2014). Developmental link between sex and nutrition; *doublesex* regulates sex-specific mandible growth via juvenile hormone signaling in stag beetles. *PLoS Genet.* **10**(1): e1004098.
- Grozinger C. M., Fan Y., Hoover S. E. R. & Winston M. L. (2007). Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Mol. Ecol.* **16**: 4837–4848.
- Hall B. G. (2013). Building phylogenetic trees from molecular data with MEGA. *Mol. Biol. Evol.* **30**(5): 1229–1235.
- Hallgrímsson B., Jamniczky H. A., Young N. M., Rolian C., Schmidt-Ott U. & Marcucio R. S. (2012). The generation of variation and the developmental basis for evolutionary novelty. *J. Exp. Zool. B Mol. Dev. Evol.* **318**(6): 501–517.
- Hamilton W. D. (1964). The genetical evolution of social behaviour. I. *J. Theor. Biol.* **7**(1): 1–16.
- Hamilton W. D. (1967) Extraordinary sex ratios. *Science* **156**: 477–488.

- Hamm R. L., Meisel R. P. & Scott J. G. (2015). The evolving puzzle of autosomal versus Y-linked male determination in *Musca domestica*. *G3* **5**(3): 371–384.
- Hansen A. K. & Moran N. A. (2014). The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol. Ecol.* **23**(6): 1473–1496.
- Hasselmann M., Gempe T., Schjøtt M., Nunes-Silva C. G., Otte M. & Beye M. (2008). Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature* **454**(7203): 519–522.
- Heath B. D., Butcher R. D. J., Whitfield W. G. F. & Hubbard, S. F. (1999). Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. *Curr. Biol.* **9**(6): 313–316.
- Heddi A., Grenier A. M., Khatchadourian C., Charles H. & Nardon P. (1999). Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and *Wolbachia*. *Proc. Natl. Acad. Sci. USA* **96**(12): 6814–6819.
- Heimpel G. E. & de Boer J. G. (2008). Sex determination in the Hymenoptera. *Annu. Rev. Entomol.* **53**(1): 209–230.
- Heinze J. & Hölldobler B. (1993). Fighting for a harem of queens: physiology of reproduction in *Cardiocondyla* male ants. *Proc. Natl. Acad. Sci. USA* **90**(18): 8412–8414.
- Heinze J. & Trenkle S. (1997). Male polymorphism and gynandromorphs in the ant *Cardiocondyla emeryi*. *Naturwissenschaften* **84**(3): 129–131.
- Heinze J., Hölldobler B. & Yamauchi K. (1998). Male competition in *Cardiocondyla* ants. *Behav. Ecol. Sociobiol.* **42**(4): 239–246.
- Heinze J., Cremer S., Eckl N. & Schrempf A. (2006). Stealthy invaders: the biology of *Cardiocondyla* tramp ants. *Insectes Sociaux* **53**(1): 1–7.
- Helmkampf M., Cash E. & Gadau J. (2015). Evolution of the insect desaturase gene family with an emphasis on social Hymenoptera. *Mol. Biol. Evol.* **32**(2): 456–471.
- Hilgenboecker K., Hammerstein P., Schlattmann P., Telschow A. & Werren J. H. (2008). How many species are infected with *Wolbachia*? - a statistical analysis of current data. *FEMS Microbiol. Lett.* **281**(2): 215–220.
- Hill W. G. & Robertson A. (1966) The effect of linkage on limits to artificial selection. *Genet. Res.* **8**: 269–294.
- Holloway A. K., Strand M. R., Black W. C. & Antolin M. F. (2000). Linkage analysis of sex determination in *Bracon* sp. near *hebetor* (Hymenoptera: Braconidae). *Genetics* **154**(1): 205–212.
- Hopkins T. L. & Kramer K. J. (1992). Insect cuticle sclerotization. *Annu. Rev. Entomol.* **37**: 273–302.

- Hosokawa T., Koga R., Kikuchi Y., Meng X.-Y. & Fukatsu T. (2010). *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc. Natl. Acad. Sci. USA* **107**(2): 769–774.
- Hotopp J. C. D., Clark M. E., Oliveira D. C. S. G., Foster J. M., Fischer P., Torres M. C. M., Giebel J. D., Kumar N., Ishmael N., Wang S., Ingram J., Nene R. V., Shepard J., Tomkins J., Richards S., Spiro D. J., Ghedin E., Slatko B. E., Tettelin H. & Werren J. H. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **317**(5845): 1753–1756.
- Hu Y., Lukasik P., Moreau C. S. & Russell J. A. (2014). Correlates of gut community composition across an ant species (*Cephalotes varians*) elucidate causes and consequences of symbiotic variability. *Mol. Ecol.* **23**(6): 1284–1300.
- Husník F., Chrudimský T. & Hypša V. (2011). Multiple origins of endosymbiosis within the Enterobacteriaceae ( $\gamma$ -Proteobacteria): convergence of complex phylogenetic approaches. *BMC Biol.* **9**(1): 87.
- Husník F., Nikoh N., Koga R., Ross L., Duncan R. P., Fujie M., Tanaka M., Satoh N., Bachtrog D., Wilson A. C.C., von Dohlen C. D., Fukatsu T. & McCutcheon J. P. (2013). Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* **153**(7): 1567–1578.
- Hyatt D., Chen G.-L., LoCascio P. F., Land M. L., Larimer F. W. & Hauser L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**: 119.
- Imai H. & Taylor R. W. (1989). Chromosomal polymorphisms involving telomere fusion, centromeric inactivation and centromere shift in the ant *Myrmecia (pilosula) n=1*. *Chromosoma* **98**: 456–460.
- Ito K., Katsuma S., Kuwazaki S., Jouraku A., Fujimoto T., Sahara K., Yasukochi Y., Yamamoto K., Tabunoki H., Yokoyama T., Kadono-Okuda K. & Shimada T. (2015). Mapping and recombination analysis of two moth colour mutations, Black moth and Wild wing spot, in the silkworm *Bombyx mori*. *Heredity* Epub ahead of print.
- Jiménez N., González-Candelas F. & Silva F. J. (2000). Prephenate dehydratase from the aphid endosymbiont (*Buchnera*) displays changes in the regulatory domain that suggest its desensitization to inhibition by phenylalanine. *J. Bacteriol.* **182**(10): 2967–2969.
- Joron M., Frezal L., Jones R.T., Chamberlain N.L., Lee S.F., Haag C.R., Whibley A., Becuwe M., Baxter S.W., Ferguson L., Wilkinson P. A., Salazar C., Davidson C., Clark R., Quail M. A., Beasley H., Glithero R., Lloyd C., Sims S., Jones M. C., Rogers J., Jiggins C. D. & French-Constant R. H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**: 203–206.
- Kaessmann H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**(10): 1313–1326.

- Kamakura M. (2011). Royalactin induces queen differentiation in honeybees. *Nature* **473**(7348): 478–483.
- Kamping A., Katju V., Beukeboom L. W. & Werren J. H. (2007). Inheritance of gynandromorphism in the parasitic wasp *Nasonia vitripennis*. *Genetics* **175**(3): 1321–1333.
- Kaltenpoth M., Gottler W., Herzner G. & Strohm E. (2005). Symbiotic bacteria protect wasp larvae from fungal infestation. *Curr. Biol.* **15**(5): 475–479.
- Kaltenpoth M., Yildirim E., Gürbüz M. F. Herzner, G. & Strohm E. (2012). Refining the roots of the beewolf-*Streptomyces* symbiosis: antennal symbionts in the rare genus *Philanthinus* (Hymenoptera, Crabronidae). *Appl. Environ. Microbiol.* **78**(3): 822–827.
- Kaltenpoth M., Roeser-Mueller K., Koehler S., Peterson A., Nechitaylo T. Y., Stubblefield J. W., Herzner G., Segerd J. & Strohm E. (2014). Partner choice and fidelity stabilize coevolution in a Cretaceous-age defensive symbiosis. *Proc. Natl. Acad. Sci.* **111**(17): 6359–6364.
- Karp P. D., Paley S. M., Krummenacker M., Latendresse M., Dale J. M., Lee T. J., Kaipa P., Gilham F., Spaulding A., Popescu L., Altman T., Paulsen I., Keseler I. M. & Caspi R. (2010). Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics* **11**(1): 40–79.
- Katoh K. & Standley D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**: 772–780.
- Kent C. F., Minaei S., Harpur B. A. & Zayed A. (2012). Recombination is associated with the evolution of genome structure and worker behavior in honey bees. *Proc. Natl. Acad. Sci. USA* **109**(44): 18012–18017.
- Keseler I. M., Mackie A., Peralta-Gil M., Santos-Zavaleta A., Gama-Castro S., Bonavides-Martínez C., Fulcher C., Huerta A. M., Kothari A., Krummenacker M., Latendresse M., Muñiz-Rascado L., Ong Q., Paley S., Schröder I., Shearer A. G. Subhraveti P., Travers M., Weerasinghe D., Weiss V., Collado-Vides J., Gunsalus R. P., Paulsen I. & Karp P. D. (2013). EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* **41**: D605–D612.
- Kijimoto T., Moczek A. P. & Andrews J. (2012). Diversification of *doublesex* function underlies morph-, sex-, and species-specific development of beetle horns. *Proc. Natl. Acad. Sci. USA* **109**(50): 20526–20531.
- Kinomura K. & Yamauchi K. (1987). Fighting and mating behaviors of dimorphic males in the ant *Cardiocondyla wroughtoni*. *J. Ethol.* **5**(1): 75–81.
- Klein A., Schrader L., Gil R., Manzano-Marín A., Flórez L., Wheeler D., Werren J. H., Latorre A., Heinze J., Kaltenpoth M., Moya A. & Oettler J. (2015). A novel intracellular mutualistic bacterium in the invasive ant *Cardiocondyla obscurior*. *ISME J.* Epub ahead of print.



- Koch H. & Schmid-Hempel P. (2011). Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proc. Natl. Acad. Sci. USA* **108**(48): 19288–19292.
- Koch V., Nissen I., Schmitt B. D. & Beye M. (2014). Independent evolutionary origin of fem paralogous genes and complementary sex determination in hymenopteran insects. *PLoS ONE* **9**(4): e91883.
- Koga R. & Moran N. A. (2014). Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. *ISME J.* **8**(6): 1237–1246.
- Koga R., Meng X. Y., Tsuchida T. & Fukatsu T. (2012). Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proc. Natl. Acad. Sci. USA* **109**(20): E1230–E1237.
- Kopp A. (2012). *Dmrt* genes in the development and evolution of sexual dimorphism. *Trends Genet.* **28**(4): 175–184.
- Kosambi D. D. (1944). The estimation of map distances from recombination value. *Ann. Eugen.* **12**: 172–175.
- Kozielska M., Pen I., Beukeboom L. W. & Weissing F. J. (2006). Sex ratio selection and multi-factorial sex determination in the housefly: a dynamic model. *J. Evol. Biol.* **19**: 879–888.
- Krzywinski M., Schein J., Birol I., Connors J., Gascoyne R., Horsman D., Jones S. J. & Marra M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**(9): 1639–1645.
- Kugler J. (1983). The males of *Cardiocondyla emery* (Hymenoptera: Formicidae) with the description of the winged male of *Cardiocondyla wroughtonii* (Forel). *Israel J. Entomol.* **17**: 1–21.
- Kuijper B. & Pen I. (2010). The evolution of haplodiploidy by male-killing endosymbionts: importance of population structure and endosymbiont mutualisms. *J. Evol. Biol.* **23**(1): 40–52.
- Kulmuni J., Wurm Y. & Pamilo P. (2013). Comparative genomics of chemosensory protein genes reveals rapid evolution and positive selection in ant-specific duplicates. *Heredity* **110**(6): 538–547.
- Kunte K., Zhang W., Tenger-Trolander A., Palmer D. H., Martin A., Reed R. D., Mullen S. P. & Kronforst M. R. (2014). *doublesex* is a mimicry supergene. *Nature* **507**(7491): 229–232.
- Lai C. Y., Baumann L. & Baumann P. (1994). Amplification of *trpEG*: adaptation of *Buchnera aphidicola* to an endosymbiotic association with aphids. *Proc. Natl. Acad. Sci. USA* **91**(9): 3819–3823.

- Lapidus A., LaButti K., Foster B., Lowry S., Trong S. & Goltsman E. (2008). POLISHER: An effective tool for using ultra short reads in microbial genome assembly and finishing. AGBT, Marco Island, FL, USA.
- Lartillot N., Lepage T. & Blanquart S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**(17): 2286–2288.
- Laslett D. & Canback B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**(1): 11–16.
- Leniaud L., Darras H., Boulay R. & Aron S. (2012). Social hybridogenesis in the clonal ant *Cataglyphis hispanica*. *Curr. Biol.* **22**(13): 1188–1193.
- Lesecque Y., Mouchiroud D. & Duret L. (2013). GC-biased gene conversion in yeast is specifically associated with crossovers: molecular mechanisms and evolutionary significance. *Mol. Biol. Evol.* **30**(6): 1409–1419.
- Li H. & Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754–1760.
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N. Marth G., Abecasis G., Durbin R. & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16): 2078–2079.
- Li H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**(21): 2987–2993.
- Liu G., Wu Q., Li J. Zhang G. & Wan F. (2015). RNAi-mediated knock-down of *transformer* and *transformer 2* to generate male-only progeny in the oriental fruit fly, *Bactrocera dorsalis* (Hendel). *PLoS ONE* **10**(6): e0128892.
- Liu H., Zhang X., Huang J., Chen J.-Q., Tian D., Hurst L. D. & Yang S. (2015). Causes and consequences of crossing-over evidenced via a high-resolution recombinational landscape of the honey bee. *Genome Biol.* **16**(1): 15.
- Livak K. J. & Schmittgen T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_T}$  method. *Methods* **25**(4): 402–408.
- Loehlin D. W., Oliveira D. C. S. G., Edwards R., Giebel J. D., Clark M. E., Cattani M. V., van de Zande L., Verhulst E. C., Beukeboom L. W., Muñoz-Torres M. & Werren J. H. (2010). Non-coding changes cause sex-specific wing size differences between closely related species of *Nasonia*. *PLoS Genet.* **6**(1): e1000821.
- Love M. I., Huber W. & Anders S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**(12): 550.
- Lowe T. M. & Eddy S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**(5): 955–964.

- Luo R., Liu B., Xie Y., Li Z., Huang W., Yuan J., He G., Chen Y., Pan Q., Liu Y., Tang J., Wu G., Zhang H., Shi Y., Liu Y., Yu C., Wang B., Lu Y., Han C., Cheung D. W., Yiu S.-M., Peng S., Xiaoqian Z., Liu G., Liao X., Li Y., Yang H., Wang J., Lam T.-W. & Wang J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* **1**(1): 18.
- Mahajan-Miklos S. & Cooley L. (1994). Intercellular cytoplasm transport during *Drosophila* oogenesis. *Dev. Biol.* **165**(2): 336–351.
- Manzano-Marín A. & Latorre A. (2014). Settling down: The genome of *Serratia symbiotica* from the aphid *Cinara tujaefilina* zooms in on the process of accommodation to a cooperative intracellular life. *Genome Biol. Evol.* **6**(7): 1683–1698.
- Marais G., Charlesworth B. & Wright S. I. (2004). Recombination and base composition: the case of the highly self-fertilizing plant *Arabidopsis thaliana*. *Genome Biol.* **5**(7): R45.
- Margulis L. (1993) Symbiosis in cell evolution. Microbial communities in the archaean and proterozoic eons. 2nd ed. W. H. Freeman and Co., New York.
- Marsolier-Kergoat M.-C. & Yeramian E. (2009). GC content and recombination: reassessing the causal effects for the *Saccharomyces cerevisiae* genome. *Genetics* **183**(1): 31–38.
- Matson C. K. & Zarkower D. (2012). Sex and the singular DM domain: insights into sexual regulation, evolution and plasticity. *Nat. Rev. Genet.* **13**(3): 163–174.
- Maynard Smith J. & Szathmáry E. (1997). The major transitions in evolution. Oxford University Press: Oxford, UK.
- Mayr E. (1954). Change of genetic environment and evolution. In: Julian Huxley. Evolution as a process. London: George Allen & Unwin.
- McCutcheon J. P. & Moran N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**(1): 13–26.
- McGaugh S. E. & Noor M. A. F. (2012). Genomic impacts of chromosomal inversions in parapatric *Drosophila* species. *Phil. Trans. R. Soc. B* **367**(1587): 422–429.
- Meznar E. R., Gadau J., Koeniger N. & Rueppell O. (2010). Comparative linkage mapping suggests a high recombination rate in all honeybees. *J. Hered.* **101**(Suppl 1): S118–26.
- Michalkova V., Benoit J. B., Weiss B. L., Attardo G. M. & Aksoy S. (2014). Vitamin B<sub>6</sub> generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies. *Appl. Environ. Microbiol.* **80**(18): 5844–5853.
- Michez D., Rasmont P., Terzo M. & Vereecken N. J. (2009). A synthesis of gynandromorphy among wild bees (Hymenoptera: Apoidea), with an annotated description of several new cases. *Annales De La Société Entomologique De France* **45**(3): 365–375.

- Miller M. R., Atwood T. S., Eames B. F., Eberhart J. K., Yan Y.-L., Postlethwait J. H. & Johnson E. A. (2007). RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biol.* **8(6)**: R105.
- Minakuchi C., Zhou X. & Riddiford L. M. (2008). *Krüppel homolog 1 (Kr-h1)* mediates juvenile hormone action during metamorphosis of *Drosophila melanogaster*. *Mech. Dev.* **125(1-2)**: 91–105.
- Minakuchi C., Namiki T. & Shinoda T. (2009). *Krüppel homolog 1*, an early juvenile hormone-response gene downstream of *Methoprene-tolerant*, mediates its anti-metamorphic action in the red flour beetle *Tribolium castaneum*. *Dev. Biol.* **325(2)**: 341–350.
- Moczek A. P. (2005). The evolution and development of novel traits, or how beetles got their horns. *BioScience* **55(11)**: 937–951.
- Moczek A. P. & Emlen D. J. (2000) Male horn dimorphism in the scarab beetle, *Onthophagus taurus*: Do alternative reproductive tactics favour alternative phenotypes? *Anim. Behav.* **59(2)**: 459–466.
- Moran J. V., DeBerardinis R. J. & Kazazian H. H. (1999). Exon shuffling by L1 retrotransposition. *Science* **283(5407)**: 1530–1534.
- Moran N. A., Dunbar H. E. & Wilcox J. L. (2005). Regulation of transcription in a reduced bacterial genome: nutrient-provisioning genes of the obligate symbiont *Buchnera aphidicola*. *J. Bacteriol.* **187(12)**: 4229–4237.
- Moran N. A. & Dunbar H. E. (2006). Sexual acquisition of beneficial symbionts in aphids. *Proc. Natl. Acad. Sci. USA* **103(34)**: 12803–12806.
- Morandin C., Havukainen H., Kulmuni J., Dhaygude K., Trontti K. & Helanterä H. (2014). Not only for egg yolk - functional and evolutionary insights from expression, selection, and structural analyses of *Formica* ant vitellogenins. *Mol. Biol. Evol.* **31(8)**: 2181–2193.
- Morandin C., Dhaygude K., Pavia J., Trontti K., Wheat C. & Helanterä H. (2015). Caste-biases in gene expression are specific to developmental stage in the ant *Formica exsecta*. *J. Evol. Biol.* Epub ahead of print.
- Morgan T. H. (1911). Random segregation versus coupling in Mendelian inheritance. *Science* **34(873)**: 384–384.
- Morgan T. H. (1912) Complete linkage in the second chromosome of the male of *Drosophila*. *Science* **36**: 719–20
- Morgan T. H., Sturtevant A. H., Muller H. J. & Bridges C. B. (1915). The mechanism of mendelian heredity. New York: Henry Holt and Company.
- Moya A., Peretó, J., Gil R. & Latorre A. (2008). Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* **9(3)**: 218–229.

- Müller G. B. & G. P. Wagner (1991). Novelty in evolution: Restructuring the concept. *Annu. Rev. Ecol. Systemat.* **22**: 229–256.
- Murray R. G. & Stackebrandt E. (1995). Taxonomic note: implementation of the provisional status Candidatus for incompletely described procaryotes. *Int. J. Syst. Bacteriol.* **45**(1): 186–187.
- Murtagh F. & Legendre P. (2014). Ward’s hierarchical agglomerative clustering method: which algorithms implement ward’s criterion? *Journal of Classification* **31**(3): 274–295.
- Nakabachi A., Shigenobu S., Sakazume N., Shiraki T., Hayashizaki Y., Carninci P., Ishikawa H., Kudo T. & Fukatsu T. (2005). Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera*. *Proc. Natl. Acad. Sci. USA* **102**(15): 5477–5482.
- Nawrocki E. P. & Eddy S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**(22): 2933–2935.
- Niehuis O., Gibson J. D., Rosenberg M. S., Pannebakker B. A., Koevoets T., Judson A. K., Desjardins C. A., Kennedy K., Duggan D., Beukeboom L. W., van de Zande L., Shuker D. W., Werren J. H. & Gadau J. (2010). Recombination and its impact on the genome of the haplodiploid parasitoid wasp *Nasonia*. *PLoS ONE* **5**(1): e8597.
- Nikoh N., McCutcheon J. P., Kudo T., Miyagishima S.-Y., Moran N. A. & Nakabachi A. (2010). Bacterial genes in the aphid genome: absence of functional gene transfer from *Buchnera* to its host. *PLoS Genet.* **6**(2): e1000827.
- Nipitwattanaphon M., Wang J., Ross K. G., Riba-Grognuz O., Wurm Y., Khurewathanakul C. & Keller L. (2014). Effects of ploidy and sex-locus genotype on gene expression patterns in the fire ant *Solenopsis invicta*. *Proc. Biol. Soc.* **281**(1797): 20141776.
- Ohbayashi F.M., Suzuki M.G., Mita K., Okano K. & Shimada T. (2001) A homologue of the *Drosophila doublesex* gene is transcribed into sex-specific mRNA isoforms in the silkworm, *Bombyx mori*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **1**: 145–158.
- Oettler J., Suefuji M. & Heinze J. (2010). The evolution of alternative reproductive tactics in male *Cardiocondyla* ants. *Evolution* **64**(11): 3310–3317.
- Ogata H., Goto S., Sato K., Fujibuchi W., Bono H. & Kanehisa M. (1999). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**(1): 29–34.
- Ohno S. (1970). Evolution by gene duplication. London: George Alien & Unwin Ltd; Berlin, Heidelberg and New York: Springer-Verlag.
- Oliveira D. C. S. G., Werren J. H., Verhulst E. C., Giebel J. D., Kamping A., Beukeboom L. W. & van de Zande L. (2009). Identification and characterization of the *doublesex* gene of *Nasonia*. *Insect Mol. Biol.* **18**(3): 315–324.

- Oliver K. M., Russell J. A., Moran N. A. & Hunter M. S. (2003). Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc. Natl. Acad. Sci. USA* **100**(4): 1803–1807.
- O'Reilly C., Turner P. D., Smith-Keary P. F. & McConnell D. J. (1984). Molecular cloning of genes involved in purine biosynthetic and salvage pathways of *Salmonella typhimurium*. *Molecular & General Genetics* **196**(1): 152–157.
- Orzack S. H., Sohn J. J., Kallman K. D., Levin S. A. & Johnston R. (1980). Maintenance of the three sex chromosome polymorphism in the platyfish *Xiphophorus maculatus*. *Evolution* **34**: 663–672.
- Patel A., Fondrk M. K., Kaftanoglu O., Emore C., Hunt G., Frederick K. & Amdam G. V. (2007). The making of a queen: TOR pathway is a key player in diphenic caste development. *PLoS ONE* **2**(6): e509.
- Patino-Navarrete R., Piulachs M.-D., Belles X., Moya A., Latorre A. & Peretó J. (2014). The cockroach *Blattella germanica* obtains nitrogen from uric acid through a metabolic pathway shared with its bacterial endosymbiont. *Biol. Lett.* **10**(7): 20140407.
- Pérez-Brocal V., Latorre A. & Moya A. (2013) Symbionts and pathogens: what is the difference? *Curr. Top. Microbiol. Immunol.* **358**: 215-43.
- Pérez-Brocal V., Gil R., Ramos S., Lamelas A., Postigo M., Michelena J. M., Silva F. J., Moya A. & Latorre A. (2006). A small microbial genome: the end of a long symbiotic relationship? *Science* **314**(5797): 312.
- Pérez-Cobas A. E., Maiques E., Angelova A., Carrasco P., Moya A. & Latorre A. (2015). Diet shapes the gut microbiota of the omnivorous cockroach *Blattella germanica*. *FEMS Microbiol. Ecol.* **91**(4): Epub ahead of print.
- Petersen T. N., Brunak S., von Heijne G. & Nielsen H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**(10): 785–786.
- Petrov D. A., Aminetzach Y. T., Davis J. C., Bensasson D. & Hirsh A. E. (2003). Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol. Biol. Evol.* **20**(6): 880–892.
- Pigliucci M. (2012). What, if Anything, Is an Evolutionary Novelty? *Philosophy of Science* **75**: 887-898.
- Prado S. S., Golden M., Follett P. A., Daugherty M. P. & Almeida R. P. P. (2009). Demography of gut symbiotic and aposymbiotic *Nezara viridula* L. (Hemiptera: Pentatomidae). *Environ. Entomol.* **38**(1): 103–109.
- Privman E., Wurm Y. & Keller L. (2013). Duplication and concerted evolution in a master sex determiner under balancing selection. *Proc. R. Soc. B* **280**(1758): 20122968.

- Prum R. O. (2005). Evolution of the morphological innovations of feathers. *J. Exp. Zool. Mol. Dev. Evol.* **304**(6): 570–579.
- Purcell J., Brelsford A., Wurm Y., Perrin N. & Chapuisat M. (2014). Convergent genetic architecture underlies social organization in ants. *Curr. Biol.* **24**(22): 2728–2732.
- Quinlan A. R. & Hall I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**(6): 841–842.
- Ratzka C., Gross R. & Feldhaar H. (2013). Gene expression analysis of the endosymbiont-bearing midgut tissue during ontogeny of the carpenter ant *Camponotus floridanus*. *J. Insect Physiol.* **59**(6): 611–623.
- R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>.
- Reuter M., Pedersen J. S. & Keller L. (2005). Loss of *Wolbachia* infection during colonisation in the invasive Argentine ant *Linepithema humile*. *Heredity* **94**(3): 364–369.
- Rheindt F. E., Strehl C. P. & Gadau J. (2005). A genetic component in the determination of worker polymorphism in the Florida harvester ant *Pogonomyrmex badius*. *Insectes Sociaux* **52**(2): 163–168.
- Rice W. R. (2002). Experimental tests of the adaptive significance of sexual recombination. *Nat. Rev. Genet.* **3**(4): 241–251.
- Rizzon C., Marais G., Gouy M. & Biémont C. (2002). Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome Res.* **12**(3): 400–407.
- Rodgers-Melnick E., Bradbury P. J., Elshire R. J., Glaubitz J. C., Acharya C. B., Mitchell S. E., Lic C., Lic Y. & Buckler E. S. (2015). Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proc. Natl. Acad. Sci. USA* **112**(12): 3823–3828.
- Romiguier J., Lourenco J., Gayral P., Faivre N., Weinert L. A., Ravel S., Ballenghien M., Cahais V., Bernard A., Loire E., Keller L. & Galtier N. (2014). Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *J. Evol. Biol.* **27**: 593–603.
- Rosengaus R. B., Schultheis K. F., Yalonetskaya A., Bulmer M. S., DuComb W. S., Benson R. W., Thottam J. P. & Godoy-Carter, V. (2014). Symbiont-derived  $\beta$ -1,3-glucanases in a social insect: mutualism beyond nutrition. *Front. Microbiol.* **5**: 607.
- Ross L., Gardner A., Hardy N. & West S. A. (2013). Ecology, not the genetics of sex determination, determines who helps in eusocial populations. *Curr. Biol.* **23**(23): 2383–2387.

- Ross L., Blackmon H., Lorite P., Gokhman V. E. & Hardy N. B. (2015). Recombination, chromosome number and eusociality in the Hymenoptera. *J. Evol. Biol.* **28**(1): 105–116.
- Roze D. & Barton N. H. (2006). The Hill-Robertson effect and the evolution of recombination. *Genetics* **173**(3): 1793–1811.
- Rueppell O., Meier S. & Deutsch R. (2012). Multiple mating but not recombination causes quantitative increase in offspring genetic diversity for varying genetic architectures. *PLoS ONE* **7**(10): e47220.
- Russell J. A. & Moran N. A. (2006). Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. *Proc. R. Soc. B* **273**(1586): 603–610.
- Russell J. A., Goldman-Huertas B., Moreau C. S., Baldo L., Stahlhut J. K., Werren J. H. & Pierce N. E. (2009). Specialization and geographic isolation among *Wolbachia* symbionts from ants and lycaenid butterflies. *Evolution* **63**(3): 624–640.
- Russell J. A., Moreau C. S., Goldman-Huertas B., Fujiwara M., Lohman D. J. & Pierce N. E. (2009). Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. *Proc. Natl. Sci. USA* **106**(50): 21236–21241.
- Russello M. A., Waterhouse M. D., Etter P. D. & Johnson E. A. (2015). From promise to practice: pairing non-invasive sampling with genomics in conservation. *PeerJ* **3**(3): e1106–18.
- Rutherford K., Parkhill J., Crook J., Horsnell T., Rice P., Rajandream M.-A. & Barrell B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**(10): 944–945.
- Sambrook J. & Russell D. W. (2001). Molecular cloning: A laboratory manual. New York: Cold Spring Harbor Laboratory Press.
- Sanders J. G., Powell S., Kronauer D. J. C., Vasconcelos H. L., Frederickson M. E. & Pierce N. E. (2014). Stability and phylogenetic correlation in gut microbiota: lessons from ants and apes. *Mol. Ecol.* **23**(6): 1268–1283.
- Santana M. F., Silva J. C. F., Mizubuti E. S. G., Araújo E. F., Condon B. J., Turgeon B. G. & Queiroz M. V. (2014). Characterization and potential evolutionary impact of transposable elements in the genome of *Cochliobolus heterostrophus*. *BMC Genomics* **15**(1): 536.
- Sarre S. D., Georges A. & Quinn A. (2004). The ends of a continuum: genetic and temperature-dependent sex determination in reptiles. *BioEssays* **26**(6): 639–645.
- Sauer C., Stackebrandt E., Gadau J., Hölldobler B. & Gross R. (2000). Systematic relationships and cospeciation of bacterial endosymbionts and their carpenter ant host species: proposal of the new taxon *Candidatus Blochmannia* gen. nov. *Int. J. Syst. Evol. Microbiol.* **50**: 1877–1886.



- Scheer M., Grote A., Chang A., Schomburg I., Munaretto C., Rother M., Söhngen C., Stelzer M., Thiele J. & Schomburg D. (2011). BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.* **39**: D670–D676.
- Schmieder S., Colinet D. & Poirié M. (2012). Tracing back the nascence of a new sex-determination pathway to the ancestor of bees and ants. *Nat. Comms.* **3**: 895.
- Schrader L., Kim J. W., Ence D., Zimin A., Klein A., Wyschetzki K., Weichselgartner T., Kemena C., Stökl J., Schultner E., Wurm Y., Smith C. D., Yandell M., Heinze J., Gadau J. & Oettler J. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nat. Comms.* **5**: 5495.
- Schrader L., Simola D. F., Heinze J. & Oettler J. (2015). Sphingolipids, transcription factors, and conserved toolkit genes: Developmental plasticity in the ant *Cardiocondyla obscurior*. *Mol. Biol. Evol.* **32**(6): 1474–1486.
- Schrader L., Helanterä H. & Oettler J. (unpublished). Evolutionary causes and consequences of developmental gene expression bias in an eusocial insect.
- Schrempf A., Heinze J. & Cremer S. (2005). Sexual cooperation: mating increases longevity in ant queens. *Curr. Biol.* **15**(3): 267–270.
- Schrempf A. & Heinze J. (2006). Proximate mechanisms of male morph determination in the ant *Cardiocondyla obscurior*. *Evol. Dev.* **8**(3): 266–272.
- Schrempf A. & Heinze J. (2007). Back to one: consequences of derived monogyny in an ant with polygynous ancestors. *J. Evol. Biol.* **20**(2): 792–799.
- Schrempf A., Aron S. & Heinze J. (2006). Sex determination and inbreeding depression in an ant with regular sib-mating. *Heredity* **97**(1): 75–80.
- Schrempf A. & Heinze J. (2008). Mating with stressed males increases the fitness of ant queens. *PLoS ONE* **3**(7): e2592.
- Schrempf A., Cremer S. & Heinze J. (2011). Social influence on age and reproduction: reduced lifespan and fecundity in multi-queen ant colonies. *J. Evol. Biol.* **24**(7): 1455–1461.
- Schrempf A., Wyschetzki K., Klein A., Schrader L., Oettler J. & Heinze J. (2015). Mating with an allopatric male triggers immune response and decreases longevity of ant queens. *Mol. Ecol.* **24**(14): 3618–3627.
- Schuler H., Bertheau C., Egan S. P., Feder J. L., Riegler M., Schlick-Steiner B. C., Steiner F., Johannesen J., Kern P., Tuba K., Lakatos F., Köppler K., Arthofer W. & Stauffer C. (2013). Evidence for a recent horizontal transmission and spatial spread of *Wolbachia* from endemic *Rhagoletis cerasi* (Diptera: Tephritidae) to invasive *Rhagoletis cingulata* in Europe. *Mol. Ecol.* **22**(15): 4101–4111.
- Schwander T. & Keller L. (2008). Genetic compatibility affects queen and worker caste determination. *Science* **322**(5901): 552.

- Schwander T., Libbrecht R. & Keller L. (2014). Supergenes and complex phenotypes. *Curr. Biol.* **24(7)**: R288–94.
- Seifert B. (2003). The ant genus *Cardiocondyla* (Insecta: Hymenoptera: Formicidae) - a taxonomic revision of the *C. elegans*, *C. bulgarica*, *C. batesii*, *C. nuda*, *C. shuckardi*, *C. stambuloffi*, *C. wroughtonii*, *C. emeryi* and *C. minutior* species groups. *Annalen Des Naturhistorischen Museums Wien* (**104B**): 203–338.
- Shukla J. N. & Palli S. R. (2012a). *Doublesex* target genes in the red flour beetle, *Tribolium castaneum*. *Sci. Rep.* **2**: 948.
- Shukla J. N. & Palli S. R. (2012b). Sex determination in beetles: production of all male progeny by parental RNAi knockdown of *transformer*. *Sci. Rep.* **2**: 602.
- Simola D. F., Ye C., Mutti N. S., Dolezal K., Bonasio R., Liebig J., Reinberg D. & Berger S. L. (2013). A chromatin link to caste identity in the carpenter ant *Camponotus floridanus*. *Genome Res.* **23(3)**: 486–496.
- Sirviö A., Gadau J., Rueppell O., Lamatsch D., Boomsma J. J., Pamilo P. & Page R. E. (2006). High recombination frequency creates genotypic diversity in colonies of the leaf-cutting ant *Acromyrmex echinator*. *J. Evol. Biol.* **19(5)**: 1475–1485.
- Sirviö A., Pamilo P., Johnson R. A., Page R. E. & Gadau J. (2010). Origin and evolution of the dependent lineages in the genetic caste determination system of *Pogonomyrmex* ants. *Evolution* **65(3)**: 869–884.
- Sirviö A., Johnston J. S., Wenseleers T. & Pamilo P. (2011). A high recombination rate in eusocial Hymenoptera: evidence from the common wasp *Vespula vulgaris*. *BMC Genet.* **12(1)**: 95.
- Smith C. R., Toth A. L., Suarez A. V. & Robinson G. E. (2008). Genetic and genomic analyses of the division of labour in insect societies. *Nat. Rev. Genet.* **9(10)**: 735–748.
- Smith C. R., Helms Cahan S., Kemena C., Brady S. G., Yang W., Bornberg-Bauer E., Eriksson T., Gadau J., Helmkampf M., Gotzek D., Miyakawa M. O., Suarez A. V. & Mikheyev A. (2015). How do genomes create novel phenotypes? Insights from the loss of the worker caste in ant social parasites. *Mol. Biol. Evol.* [Epub ahead of print]
- Solignac M., Mougél F., Vautrin D., Monnerot M. & Cornuet J. M. (2007). A third-generation microsatellite-based linkage map of the honey bee, *Apis mellifera*, and its comparison with the sequence-based physical map. *Genome Biol.* **8(4)**: R66.
- Song M. & Boissinot S. (2007). Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. *Gene* **390(1-2)**: 206–213.
- Staden R., Beal K. F. & Bonfield J. K. (2000). The Staden package, 1998. *Methods in Molecular Biology* **132**: 115–130.
- Stapley J., Santure A. W. & Dennis S. R. (2015). Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Mol. Ecol.* **24(9)**: 2241–2252.

- Stansbury M. S. & Moczek A. P. (2014). The function of Hox and appendage-patterning genes in the development of an evolutionary novelty, the *Photuris* firefly lantern. *Proc. R. Soc. B* **281**(1782): 20133333.
- Stoll S., Feldhaar H., Fraunholz M. J. & Gross R. (2010). Bacteriocyte dynamics during development of a holometabolous insect, the carpenter ant *Camponotus floridanus*. *BMC Microbiol.* **10**(1): 308.
- Stolle E., Wilfert L., Schmid-Hempel R., Schmid-Hempel P., Kube M., Reinhardt R. & Moritz R. F. (2011). A second generation genetic map of the bumblebee *Bombus terrestris* (Linnaeus, 1758) reveals slow genome and chromosome evolution in the Apidae. *BMC Genomics* **12**(1): 48.
- Suzek B. E., Ermolaeva M. D., Schreiber M. & Salzberg S. L. (2001). A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics* **17**(12): 1123–1130.
- Suzuki R. & Shimodaira, H. (2015) pvclust: Hierarchical clustering with p-values via multiscale bootstrap resampling. Retrieved from <https://cran.r-project.org/web/packages/pvclust/index.html>
- Snyder A. K. & Rio R. V. (2013). The interwoven biology of the tsetse holobiont. *J. Bacteriol.* **195**: 4322–4330.
- Talavera G. & Castresana J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biol.* **56**: 564–577.
- Tamames J., Gil R., Latorre A., Peretó J., Silva F. J. & Moya A. (2007). The frontier between cell and organelle: genome analysis of *Candidatus Carsonella ruddii*. *BMC Evol. Biol.* **7**(1): 181.
- Tanaka K., Barmina O., Sanders L. E., Arbeitman M. N. & Kopp A. (2011). Evolution of sex-specific traits through changes in HOX-dependent *doublesex* expression. *PLoS Biol.* **9**(8): e1001131.
- Tåquist H., Cui Y. & Ardell D. H. (2007). TFAM 1.0: an online tRNA function classifier. *Nucleic Acids Res.* **35**: W350–W353.
- Tomchaney M., Mysore K., Sun L., Li P., Emrich S. J., Severson D. W. & Duman-Scheel M. (2014). Examination of the genetic basis for sexual dimorphism in the *Aedes aegypti* (dengue vector mosquito) pupal brain. *Biology of Sex Differences* **5**(1): 10.
- Torday J. (2015). Pleiotropy as the mechanism for evolving novelty: Same signal, different result. *Biology* **4**(2): 443–459.
- Ugelvig L. V., Kronauer D. J. C., Schrenpf A., Heinze J. & Cremer S. (2010). Rapid anti-pathogen response in ant societies relies on high genetic diversity. *Proc. R. Soc. B* **277**(1695): 2821–2828.

- Ugelvig L. V. & Cremer S. (2012). Effects of social immunity and uniclonality on host-parasite interactions in invasive insect societies. *Functional Ecology* **26**(6): 1300–1312.
- Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A. & Speleman F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**(7): research0034.
- Verhulst E. C., Beukeboom L. W. & van de Zande L. (2010). Maternal control of haplodiploid sex determination in the wasp *Nasonia*. *Science* **328**: 620–623.
- Verhulst E. C., Lynch J. A., Bopp D., Beukeboom L. W. & van de Zande L. (2013). A new component of the *Nasonia* sex determining cascade is maternally silenced and regulates *transformer* expression. *PLoS ONE* **8**(5): e63618.
- Vigneron A., Charif D., Vincent-Monégat C., Vallier A., Gavory F., Wincker P. & Heddi A. (2012). Host gene response to endosymbiont and pathogen in the cereal weevil *Sitophilus oryzae*. *BMC Microbiol.* **12**(Suppl 1): S14.
- Vigneron A., Masson F., Vallier A., Balmand S., Rey M., Vincent-Monégat C., Aksoy E., Aubailly-Giraud E., Zaidman-Rémy A. & Heddi A. (2014). Insects recycle endosymbionts when the benefit is over. *Curr. Biol.* **24**(19): 2267–2273.
- Viljakainen L., Reuter M. & Pamilo P. (2008). *Wolbachia* transmission dynamics in *Formica* wood ants. *BMC Evol. Biol.* **8**: 55.
- Vooght L. De, Caljon G., Hees J. Van & Abbeele J. Van Den (2015). Paternal transmission of a secondary symbiont during mating in the viviparous tsetse fly. *Mol. Biol. Evol.* **32**(8): 1977–1980.
- Wagner G. P. & Lynch V. J. (2010). Evolutionary novelties. *Curr. Biol.* **20**(2): R48–52.
- Wagner G. P., Amemiya C. & Ruddle F. (2003). *Hox* cluster duplications and the opportunity for evolutionary novelties. *Proc. Natl. Acad. Sci. USA* **100**(25): 14603–14606.
- Wallin I. E. (1927) Symbiogenesis and the origin of species. Williams & Wilkins, Baltimore, MD.
- Wallberg A., Glémin S. & Webster M. T. (2015). Extreme recombination frequencies shape genome variation and evolution in the honeybee, *Apis mellifera*. *PLoS Genet.* **11**(4): e1005189.
- Wang J., Weiss B. L. & Aksoy S. (2013). Tsetse fly microbiota: form and function. *Front. Cell. Infect. Microbiol.* **3**: 69.
- Wang J., Wurm Y., Nipitwattanaphon M., Riba-Grognuz O., Huang Y.-C., Shoemaker D. & Keller L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**: 664–668.

- Ward P. S., Brady S. G., Fisher B. L. & Schultz T. R. (2014). The evolution of myrmicine ants: phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae). *Systematic Entomology* **40**(1): 61–81.
- Warnes G. R., Bolker B., Bonebakker L., Gentleman R., Huber W., Liaw A., Lumley T., Maechler M., Magnusson A., Moeller S., Schwartz M. & Venables B. (2015). Gplots: Various R programming tools for plotting data. Retrieved from <http://cran.r-project.org/web/packages/gplots/index.html>
- Watanabe K., Yukuhiro F., Matsuura Y., Fukatsu T. & Noda H. (2014). Intrasperm vertical symbiont transmission. *Proc. Natl. Acad. Sci. USA* **111**(20): 7433–7437.
- Weeks A. R., Turelli M., Harcombe W. R., Reynolds K. T. & Hoffmann A. A. (2007). From parasite to mutualist: rapid evolution of *Wolbachia* in natural populations of *Drosophila*. *PLoS Biol.* **5**(5): e114.
- Welburn S. C., Arnold K., Maudlin I. & Gooday G. W. (1993). Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies. *Parasitology* **107**: 141–145.
- Wenseleers T., Ito F., Van Borm S., Huybrechts R., Volckaert F. & Billen J. (1998). Widespread occurrence of the micro-organism *Wolbachia* in ants. *Proc. Biol. Sci.* **265**(1404): 1447–1452.
- Wenseleers T., Sundstrom L. & Billen J. (2002). Deleterious *Wolbachia* in the ant *Formica truncorum*. *Proc. Biol. Sci.* **269**(1491): 623–629.
- Werren J. H. & Beukeboom L. W. (1998) Sex determination, sex ratios, and genetic conflict. *Annu. Rev. Ecol. Syst.* **29**: 233–261.
- Werren J. H., Baldo L. & Clark M. E. (2008). *Wolbachia*: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* **6**(10): 741–751.
- West-Eberhard M. J. (2003). Developmental plasticity and evolution. Oxford University Press, USA.
- West-Eberhard M. J. (2005). Developmental plasticity and the origin of species differences. *Proc. Natl. Acad. Sci. USA* **102**(Suppl. 1): 6543–6549.
- West-Eberhard M. J. (2009). Toward a modern revival of Darwin’s theory of evolutionary novelty. *Philosophy of Science* **75**(5): 899–908.
- Wheeler D. E. (1986). Developmental and physiological determinants of caste in social hymenoptera: evolutionary implications. *The American Naturalist* **128**(1): 13–34.
- Wheeler D. E. (1990). The developmental basis of worker polymorphism in fire ants. *J. Insect Physiol.* **36**(5): 315–322.
- Wheeler D., Redding A. J. & Werren J. H. (2013). Characterization of an ancient lepidopteran lateral gene transfer. *PLoS ONE* **8**(3): e59262.

- Whitehorn P. R., Tinsley M. C., Brown M. J. F., Darvill B. & Goulson D. (2009). Impacts of inbreeding on bumblebee colony fitness under field conditions. *BMC Evol. Biol.* **9**: 152.
- Whiting P. W. (1943). Multiple alleles in complementary sex determination of *Habrobracon*. *Genetics* **24**: 110–111.
- Wiernasz D. C., Hines J., Parker D. G. & Cole B. J. (2008). Mating for variety increases foraging activity in the harvester ant, *Pogonomyrmex occidentalis*. *Mol. Ecol.* **17**(4): 1137–1144.
- Wilfert L., Gadau J. & Schmid-Hempel P. (2007). Variation in genomic recombination rates among animal taxa and the case of social insects. *Heredity* **98**(4): 189–197.
- Wilgenburg E. van, Driessen G. & Beukeboom L. W. (2006). Single locus complementary sex determination in Hymenoptera: an “unintelligent” design? *Front. Zool.* **3**: 1.
- Wilkins A. S. (1995) Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *Bioessays* **17**: 71–77.
- Willhoeft U. & Franz G. (1996). Identification of the sex-determining region of the *Ceratitis capitata* Y chromosome by deletion mapping. *Genetics* **144**(2): 737–745.
- Williams T. M., Selegue J. E., Werner T., Gompel N., Kopp A. & Carroll S. B. (2008) The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* **134**(4): 610–623.
- Wolschin F., Hölldobler B., Gross R. & Zientz E. (2004). Replication of the endosymbiotic bacterium *Blochmannia floridanus* is correlated with the developmental and reproductive stages of its ant host. *Appl. Environ. Microbiol.* **70**(7): 4096–4102.
- Worning P., Jensen L. J., Hallin P. F., Staerfeldt H.-H. & Ussery D. W. (2006). Origin of replication in circular prokaryotic chromosomes. *Environ. Microbiol.* **8**(2): 353–361.
- Wright S. I., Agrawal N. & Bureau T. E. (2003). Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Res.* **13**(8): 1897–1903.
- Wu Z., Hopper K. R., Ode P. J., Fuester R. W., Tuda M. & Heimpel G. E. (2005). Single-locus complementary sex determination absent in *Heterospilus prosopidis* (Hymenoptera: Braconidae). *Heredity* **95**(3): 228–234.
- Wu D., Daugherty S. C., Van Aken S. E., Pai G. H., Watkins K. L., Khouri H., Tallon L. J., Zaborsky J. M., Dunbar H. E., Tran P. L., Moran N. A. & Eisen J. A. (2006). Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.* **4**(6): e188–14.
- Yamamoto K., Narukawa J., Kadono-Okuda K., Nohata J., Sasanuma M., Suetsugu Y., Banno Y., Fujii H., Goldsmith M. R. & Mita K. (2006). Construction of a single nucleotide polymorphism linkage map for the silkworm, *Bombyx mori*, based on bacterial artificial chromosome end sequences. *Genetics* **173**(1): 151–161.

- Yang A. S. & Abouheif E. (2011). Gynandromorphs as indicators of modularity and evolvability in ants. *J. Exp. Zool. B Mol. Dev. Evol.* **316**(5): 313–318.
- Ye K., Schulz M. H., Long Q., Apweiler R. & Ning Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**(21): 2865–2871.
- Ye Y. H., Woolfit M., Rancès E., O'Neill S. L. & McGraw E. A. (2013). *Wolbachia*-associated bacterial protection in the mosquito *Aedes aegypti*. *PLoS Neglected Tropical Diseases* **7**(8): e2362.
- Yi W. & Zarkower D. (1999). Similarity of DNA binding and transcriptional regulation by *Caenorhabditis elegans* MAB-3 and *Drosophila melanogaster* DSX suggests conservation of sex determining mechanisms. *Development* **126**(5): 873–881.
- Yoshizawa J., Mimori K., Yamauchi K. & Tsuchida K. (2009). Sex mosaics in a male dimorphic ant *Cardiocondyla kagutsuchi*. *Naturwissenschaften* **96**(1): 49–55.
- Zarkower D. (2001) Establishing sexual dimorphism: conservation amidst diversity? *Nat. Rev. Genet.* **2**: 175–185.
- Zhang Z., Klein J. & Nei M. (2014). Evolution of the *sex-lethal* gene in insects and origin of the sex-determination system in *Drosophila*. *J. Mol. Evol.* **78**(1): 50–65.
- Zientz E., Beyaert I., Gross R. & Feldhaar H. (2006). Relevance of the endosymbiosis of *Blochmannia floridanus* and carpenter ants at different stages of the life cycle of the host. *Appl. Environ. Microbiol.* **72**(9): 6027–6033.
- Zilber-Rosenberg I. & Rosenberg E. (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* **32**: 723–735
- Zug R. & Hammerstein P. (2015). Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. *Biol. Rev. Camb. Philos. Soc.* **90**(1): 89–111.

# Acknowledgements

At this point I want to thank all people that supported me during my PhD thesis.

First of all, I want to thank Prof. Dr. Jürgen Heinze for the opportunity to do a PhD, the supervision and the support during the last 5 years since I started working in his lab. I appreciated it very much to get this opportunity, to get a deep knowledge about ants and biology and to visit a lot of conferences all over the world.

I thank Prof. Dr. Judith Korb and Prof. Dr. Stephan Schneuwly very much for taking the time to evaluate my thesis. Moreover I want to thank Prof. Dr. Heike Feldhaar for mentoring me over the last three years.

I'd like to thank all people from the "Heinze Lehrstuhl" for sharing a great time during the last years! Lukas, thank you for help with bioinformatics and your friendship! Thank you Nicky, Claudi and Susanne for the great times we had at our coffees after lunch! Thanks to all technical assistants, Tina, Helena, Doris, Birgit, Maria, Andi and Stefan, and to Sonja for help and advice! Thanks to Alex and Eva for collaborations, encouraging words and support! I had not only colleagues but found a lot of friends!

Especially, I want to thank Dr. Jan Oettler for his supervision, his support, his encouragement and his excitement about my work. Thank you Jan for getting me in the CORE team and for sharing your dedication to ants, research and biology with me. Thank you for your help and friendship over the last years and especially during the last weeks!

Mathias, thank you for your support during the last weeks and days, but - more important - during the last years. I appreciate your presence in my life so much and I am looking forward to sharing my life with you!

Regensburg, 11. September 2015