Phenotypic Plasticity in *Cardiocondyla obscurior*

DISSERTATION ZUR ERLANGUNG DES

DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)

DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN

DER UNIVERSITÄT REGENSBURG

vorgelegt von

Lukas Schrader

aus

Bochum

im Jahr

2015

Promotionsgesuch eingereicht am: 15.04.2015

Die Arbeit wurde angeleitet von Prof. Dr. J. Heinze

Unterschrift:

This thesis is composed of the following manuscripts:

Chapter 2: Lukas Schrader, Jay W Kim, Daniel Ence, Aleksey Zimin, Antonia Klein, Katharina Wyschetzki, Tobias Weichselgartner, Carsten Kemena, Johannes Stökl, Eva Schultner, Yannick Wurm, Christopher D Smith, Mark Yandell, Jürgen Heinze, Jürgen Gadau, and Jan Oettler. (2014). ***Transposable element islands facilitate adaptation to novel environments in an invasive species.*** Nature Communications, 5:5495.

Author contributions: J.O. and L.S. designed the study; J.O., L.S., J.G., and J.H. wrote the manuscript; L.S. and J.O. analysed the data; A.Z. was responsible for genome assembly; J.W.K. and C.D.S. were responsible for repeat annotation; D.E., M.Y. and L.S. were responsible for gene prediction; C.K. was responsible for CpG o/e calculation; Y.W. was responsible for data logistics; J.O., L.S., T.W., K.W. and A.K. were responsible for analysing phenotypic differences; J.S., E.S. and J.O. performed the chemical analyses; all authors read and commented on the manuscript.

Chapter 3: Lukas Schrader, Daniel F Simola, Jürgen Heinze, and Jan Oettler. (2015). ***Sphingolipids, transcription factors, and conserved toolkit genes: Developmental plasticity in the ant*** **Cardiocondyla obscurior**. Molecular Biology and Evolution, in press.

Author contributions: L.S. and J.O. designed the study; L.S., J.O., J.H., and D.F.S. wrote the manuscript; L.S. analysed the data; D.F.S. was responsible for the transcription factor binding site annotation; L.S. conducted the experiments.

Chapter 4: Lukas Schrader, Heikki Helanterä, Jan Oettler. (Submitted to eLife). ***Rates of molecular evolution correlate with gene expression bias during larval development in the ant*** **Cardiocondyla obscurior**.

Author contributions: L.S. and J.O. designed the study; L.S., J.O. and H.H. wrote the manuscript; L.S. analysed the data.

Chapter 5: Antonia Klein, **Lukas Schrader**, Rosario Gil, Alejandro Manzano-Marín, Laura Flórez, David Wheeler, John H. Werren, Amparo Latorre, Jürgen Heinze, Martin Kaltenpoth, Andrés Moya, Jan Oettler. (Under review at ISME Journal). ***A novel intracellular mutualistic bacterium in the invasive ant*** **Cardiocondyla obscurior**.

Author contributions: A.K., L.S., and J.O. designed the study; A.K., L.S., J.H. and J.O. wrote the manuscript; A.K. and J.P. performed the experiments; A.K. and L.S. analysed the data; D.W. and J.W. were responsible for prokaryote sequence annotation and HGT detection; A.L., A.M., A.M-M. and R.G. were responsible for Westeberhardia genome assembly, functional annotation and phylogenomic analysis; L.F. and M.K. were responsible for *Westeberhardia* localization (FISH); all authors read, reviewed and accepted the final version of the manuscript.

Ärgere die Ameise nicht

Sie hat ihr kleines Leben und ist fröhlich

- Firdausī -

Contents

Chapter 1 - General Introduction 1

1.1 The Phenotype 1
1.2 The Genome 2
1.2.1 Phenotypic Change through Genetic Mutations 3
1.2.2 An Emphasis on Transposable Elements 5
1.3 The Environment 8
1.4 Endosymbionts 10
1.5 *Cardiocondyla obscurior* as a Laboratory Model 11
1.6 *Cardiocondyla obscurior* as a Model to Study Phenotypic Evolution and Plasticity 13
1.7 Aims of this Thesis 14

Chapter 2 - Transposable Element Islands Facilitate Adaptation to Novel Environments in an Invasive Species 15

2.1 Introduction 17
2.2 Results 19
2.3 Discussion 29
2.4 Material and Methods 32
2.5 Acknowledgements 32

Chapter 3 - Sphingolipids, Transcription Factors, and Conserved Toolkit Genes: Developmental Plasticity in the Ant *Cardiocondyla obscurior* 33

3.1 Introduction 35
3.2 Results 40
3.3 Discussion 48
3.4 Material and Methods 52
3.5 Acknowledgements 54

Chapter 4 - Rates of Molecular Evolution Correlate with Gene Expression Bias during Larval Development in the Ant *Cardiocondyla obscurior*..... 55

4.1 Introduction....57
4.2 Material and Methods....59
4.3 Results....62
4.4 Discussion....68
4.5 Acknowledgements....71

Chapter 5 - A Novel Intracellular Mutualistic Bacterium in the Invasive Ant *Cardiocondyla obscurior* 72

5.1 Introduction....74
5.2 Material and Methods....75
5.3 Results....79
5.4 Discussion....87
5.5 Acknowledgements....91

Chapter 6 - General Discussion..... 92

6.1 *Cardiocondyla obscurior* and other Ant Species as Laboratory Models in the Genomics Era....92
6.2 Phenotypic Plasticity and Transposable Elements as Potential Sources for Rapid Adaptation in *Cardiocondyla obscurior*....94
6.3 Future Prospects....97

Chapter 7 - Summary 100

8. References 102

9. Acknowledgements..... 122

Supplementary Material ……...S1

Chapter 1

1. General Introduction

"Genetic variation and developmental plasticity are fundamental properties of all living things: all individual organisms [...] have distinctive genomes, and all of them have phenotypes that respond to genomic and environmental inputs."

– Mary Jane West-Eberhard (2005a)

1.1 The Phenotype

The diversity of organismal life on earth is stunning. Ranging from the simplest, single-celled bacteria to complex multi-cellular organisms, life is extremely versatile and it is almost impossible to conceive that it all evolved from the same ancestor some four billion years ago. Conceptually, an organism can be considered as a network of interconnected morphological, developmental, biochemical, physiological and behavioural phenotypic traits. Combined, these traits constitute a unique combination: the individual. Phenotypic traits differ not only between species but also between and within individuals (West-Eberhard 1989), and virtually every phenotypic trait is subject to change over the course of an individual's life (DeWitt and Scheiner 2004). For example, experience and learning can affect behavioural traits, and developmental programming and exercise determine morphology. In spite of recent controversy (Nowak et al. 2010; Boomsma et al. 2011; Abbot et al. 2011), the individual is generally considered to be the target of selection in evolutionary processes. However, it is the phenotype that constitutes the interface between individuals and evolution (West-Eberhard 2003).

Selection is non-random, directional change of trait frequencies in a natural population by differential survival and reproductive success of phenotypically different individuals. Evolution through adaptation and speciation is hence fuelled by continuously emerging phenotypic differences between individuals. To study evolution, it is thus necessary to assess what factors contribute and shape an organism's phenotype, and how novel phenotypes can originate from these sources.

This thesis, covering various aspects of phenotypic evolution in the ant *Cardiocondyla obscurior,* focuses on three major factors that control an organism's phenotype: The genome (Chapter 2), organismal susceptibility to the environment (phenotypic plasticity, Chapters 3 and 4) and, on the borderline between genetics and environment, endosymbionts (Chapter 5).

Based on the analysis of the draft genome sequence of *C. obscurior* and the genomic comparison of two independent populations from Brazil and Japan, Chapter 2 centres on genetic and genomic mechanisms that allow the evolution of adaptive phenotypes in spite of low genetic diversity in founder populations of *C. obscurior*. In Chapter 3, divergent gene expression patterns underlying the polyphenic/polymorphic development of *C. obscurior* are used to illuminate the genetic basis of developmental plasticity in eusocial Hymenoptera, including a discussion on the potential origins of novel phenotypes through plasticity. Based on the results obtained in Chapter 3, Chapter 4 focuses on the molecular evolutionary consequences of plastic gene expression, with a discussion on underlying selection regimes. Chapter 5 contains the first description of the intracellular endosymbionts *Candidatus* Westeberhardia cardiocondylae, which is present in most analysed populations of *C. obscurior* and conveys novel phenotypic traits to its host by contributing its genetic repertoire to the symbiosis.

The following paragraphs of this general introduction are intended to provide a synopsis of the most important aspects affecting phenotypes and their evolution, comprising a brief overview of genomes and genetic diversification, an introduction to phenotypic plasticity and its evolutionary significance, and a primer on the role of endosymbionts in insects. The last sections of this chapter introduce *C. obscurior* and aim to promote this species as a model for the study of phenotypic evolution, outlining its advantages regarding maintenance under laboratory conditions, and its high level of adaptability and plasticity.

1.2 The Genome

Containing an organism's blueprint, the genome is fundamental in shaping an individual's phenotype. The largest share of this blueprint is coded in the nuclear genome, complemented by the much smaller mitochondrial and plastidial genomes, present mostly but not exclusively in autotrophic organisms (SB Gould et al. 2008). Biochemically, genomes and any other genetic element are DNA molecules composed of 2-deoxyribose, phosphate groups, and the four nucleotides adenine (A), guanine (G), thymine (T) and cytosine (C). Beyond the mere biochemical makeup, genomes are structured semantically into discrete genetic elements (e.g. genes) and every element can itself again be highly sub-structured (e.g. genes can be sub-structured into exons and introns). In addition, groups of genetic elements can form a functional unit (e.g. operons) and cellular processes furthermore impose certain structural requirements (e.g. telomeres, centromeres) on genomes.

In spite of its rather simple biochemistry with only four different nucleotides, the coding potential of DNA is virtually infinite. In a recent effort, the ENCODE project revealed pervasive transcriptional activity across the human genome, with a remarkable 75 % of the human genome being transcribed to some extent (Djebali et al. 2012). With increasing effort to decipher the functionality of genomes, the list of different genetic elements is growing constantly, changing our understanding of genomic complexity and the diversity of transcripts.

To date, the best-characterized genetic elements are protein-coding genes. In eukaryotes, a typical protein-coding gene contains regulatory elements, exons, and introns. While regulatory elements and in part introns mainly enable transcriptional control, exons contain the genetic code for a protein's amino acid sequence. In contrast, so-called non-coding genes do not serve as templates for proteins but for functional, non-coding RNA (ncRNA) molecules (Eddy 2001). Well-known examples are RNA genes coding for enzymatically active ribosomal RNAs (rRNA) and transfer RNAs (tRNA) that play a crucial role in protein synthesis. However, following the discovery of rRNAs and their role in translation in the mid-fifties (Palade 1955), many other ncRNAs and their associated genes have been identified (Mattick 2006). While most ncRNAs have been implicated in regulating replication, transcription, splicing, and other nucleic acid metabolic processes for their high capability to interact with DNA and RNA molecules through base-pairing (Eddy 2001; Fatica and Bozzoni 2013), there are notable exceptions such as signal-recognition particle RNAs or vault RNAs involved in protein and xenobiotic translocation, respectively (Walter and Blobel 1982; van Zon et al. 2003; Gopinath et al. 2005).

Genetic elements can either serve as templates for RNA molecules and proteins, or they can themselves be functional, for example as binding sites for regulatory proteins (Wasserman and Sandelin 2004). Such regulatory genetic elements are composed of specific nucleotide sequences that allow for precise transcriptional regulation of gene expression through binding of regulatory molecules (e.g. transcription factors (TF), DNA methyltransferases (DNMT), histones). Some regulatory genetic elements are targeted for large-scale regulatory modifications through histones, changing the accessibility of an entire genomic region. Other regulatory elements provide target sites for long-term transcriptional regulation through methylation by DNMTs or for dynamic transcriptional regulation through binding of TFs (Latchman 2010; Jones 2012).

1.2.1 Phenotypic Change through Genetic Mutations

The heritability of genetic material is the mechanistic basis for evolution. Evolution can however only advance when mutations generate genetic differences between individuals. In an evolutionary context, mutations with phenotypic effects are assessed based on their impact on the mutant's fitness (Eyre-Walker and Keightley 2007). Beneficial mutations increase a carrier's fitness and deleterious

mutations decrease it, resulting in directional change in the respective mutation's frequency in a population. Whether and how a mutation affects the phenotype depends on its magnitude and genomic location, with possible effects ranging from virtually no changes in the phenotype (silent or nearly neutral mutations) to substantial phenotypic aberrations. For example, most single nucleotide substitutions ("point mutations") have very little effect on the phenotype, as they do not significantly affect the biochemical properties of the genetic element they occur in (Barrick and Lenski 2013). In contrast, mutational loss or duplication of single genes can have more severe phenotypic consequences (Conrad and Antonarakis 2007) and extreme mutational events such as whole genome duplications (WGD) provide grounds for adaptive radiations and evolutionary innovation (Edger and Pires 2009).

Point mutations are in most cases a consequence of random mistakes during DNA replication and proofreading that get incorporated into the daughter strand. In spite of this randomness, their genomic distribution follows a strict pattern produced by purifying selection against deleterious and positive selection promoting beneficial mutations (Loewe and WG Hill 2010). Depending on their position, point mutations within coding genes (and similarly for point mutations in RNA genes) can affect the gene's product. For protein coding genes, point mutations at exonic non-synonymous sites often have the highest potential to alter the gene's function by changing the amino acid sequence of the coded protein or by introducing premature stop-codons. In contrast, point mutations at synonymous sites, introns, or other non-coding and regulatory sites in a genome are less likely to have strong phenotypic effects, as they do not alter a protein or RNA product but potentially affect expression and regulation (Wray 2007).

The recurrent emergence of point mutations generates mild differences in traits between individuals of a population, forming the basis for gradual evolutionary progress through selection and adaptation. However, the paradigm of "evolution through gradual change", a core concept of classic evolutionary theory, fails to explain episodes of rapid adaptation and organismal diversification (SJ Gould 1980). Among other mechanisms, large-scale mutations that affect entire genes, chromosomes or genomes hold the potential for such rapid evolutionary change (Singh et al. 2012). Similar to single nucleotide substitutions, large-scale mutations can result from aberrations in cellular processes. For example, genes or other genomic sequences can be duplicated or deleted by unequal crossing over and WGD can occur during incomplete meiosis (Brown 2002). In addition to direct and potentially severe phenotypic consequences, gene or whole genome duplications also increase a mutant's long-term adaptive potential by introducing genetic redundancy (Flagel and Wendel 2009; Van de Peer et al. 2009). Such genetic redundancy is expected to release constraints of purifying selection and pleiotropy, allowing duplicated loci to "escape from the ruthless pressure of natural selection" (Ohno 1970) and thus evolve new functions (Conant and Wolfe 2008).

1.2.2 An Emphasis on Transposable Elements

While coding genes, RNA genes and regulatory elements are widely accepted as functionally integral parts of an organism's genome, the role of another group of genetic elements discovered less than seventy years ago is still much more enigmatic: transposable elements (TEs). In 1947, Barbara McClintock, at the time a geneticist at Cold Spring Harbor, first mentioned that the *Ds* gene she was studying at the time "may change its position in the chromosome" (McClintock 1948). Until that point, genes were thought to be neatly aligned along the arms of the chromosomes (Ravindran 2012). With increasing research efforts over the last decades, TEs are today widely recognized as ubiquitous and influential genetic elements populating the genome of virtually every organism (Fedoroff 2012). Nevertheless, due to their still largely unresolved and likely underappreciated role in evolution they remain enigmatic, and we only begin to appreciate their potential as adaptive agents.

1.2.2.1 TE structure and frequency in genomes

With the emergence of whole genome studies, the diversity and commonness of TEs became even more apparent (Hurst and Werren 2001). Several thousand copies of different TEs typically populate a genome, making up approximately 85 % of the maize and 45 % of the human genome. TEs vary in length ranging from a few hundred to several thousand base pairs (Feschotte and Pritham 2007), depending on their mode of action to achieve replication. With less than 600 bp, MITEs (miniature inverted-repeat transposable elements) are among the shortest TEs discovered so far, consisting only of terminally inverted repetitive sequence at both ends (Feschotte et al. 2002). Longer TEs can however be much more complex, containing several protein-coding genes, regulatory sequence, and recognition sites. In general, TEs are flanked by repetitive sequence motifs that are recognized by transposases or polymerases, depending on the TEs mode of transposition. A unifying classification system was developed in 2007, dividing mobile genetic elements into two classes with nine orders and 29 superfamilies based on mechanistic and enzymatic criteria (Wicker et al. 2007, Figure 1.1). Class I comprises retrotransposons, which rely on the activity of polymerases and reverse transcriptases for transposition via RNA intermediates. The classification system furthermore divides Class I elements into five orders in two subclasses (LTR (long terminal repeat) and non-LTR retrotransposons), based on mechanistic, structural and phylogenetic features (RK Slotkin and Martienssen 2007). DNA transposons belong to Class II, which is further divided into two subclasses based on the number of DNA strings that are cut during transposition. Most elements in Subclass 1 move by a "cut-and-paste" mechanism and contain terminal inverted repeats (TIR), which are recognized by transposase enzymes during transposition through cutting of both DNA strands (Fedoroff 2013). Subclass 2, which comprises two orders, Helitron-like elements and Maverick-like elements, holds DNA transposons that replicate by a "copy-and-paste" mechanism where only a single DNA strand is cut (Wicker et al. 2007).

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
Class I (retrotransposons)					
LTR	*Copia*	GAG AP INT RT RH	4–6	RLC	P, M, F, O
	Gypsy	GAG AP RT RH INT	4–6	RLG	P, M, F, O
	Bel–Pao	GAG AP RT RH INT	4–6	RLB	M
	Retrovirus	GAG AP RT RH INT ENV	4–6	RLR	M
	ERV	GAG AP RT RH INT ENV	4–6	RLE	M
DIRS	*DIRS*	GAG AP RT RH YR	0	RYD	P, M, F, O
	Ngaro	GAG AP RT RH YR	0	RYN	M, F
	VIPER	GAG AP RT RH YR	0	RYV	O
PLE	*Penelope*	RT EN	Variable	RPP	P, M, F, O
LINE	*R2*	RT EN	Variable	RIR	M
	RTE	APE RT	Variable	RIT	M
	Jockey	ORF1 APE RT	Variable	RIJ	M
	L1	ORF1 APE RT	Variable	RIL	P, M, F, O
	I	ORF1 APE RT RH	Variable	RII	P, M, F
SINE	*tRNA*		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O
Class II (DNA transposons) - Subclass 1					
TIR	*Tc1–Mariner*	Tase*	TA	DTT	P, M, F, O
	hAT	Tase*	8	DTA	P, M, F, O
	Mutator	Tase*	9–11	DTM	P, M, F, O
	Merlin	Tase*	8–9	DTE	M, O
	Transib	Tase*	5	DTR	M, F
	P	Tase	8	DTP	P, M
	PiggyBac	Tase	TTAA	DTB	M, O
	PIF– Harbinger	Tase* ORF2	3	DTH	P, M, F, O
	CACTA	Tase ORF2	2–3	DTC	P, M, F
Crypton	*Crypton*	YR	0	DYC	F
Class II (DNA transposons) - Subclass 2					
Helitron	*Helitron*	RPA Y2 HEL	0	DHH	P, M, F
Maverick	*Maverick*	C-INT ATP CYP POL B	6	DMM	M, F, O

Structural features

Long terminal repeats; Terminal inverted repeats; Coding region; Non-coding region; Diagnostic feature in non-coding region; Region that can contain one or more additional ORFs

Protein coding domains

AP, Aspartic proteinase; APE, Apurinic endonuclease; ATP, Packaging ATPase; C-INT, C-integrase; CYP, Cysteine protease; EN, Endonuclease; ENV, Envelope protein; GAG, Capsid protein; HEL, Helicase; INT, Integrase; ORF, Open reading frame of unknown function; POL B, DNA polymerase B; RH, RNase H; RPA, Replication protein A (found only in plants); RT, Reverse transcriptase; Tase, Transposase (* with DDE motif); YR, Tyrosine recombinase; Y2, YR with YY motif

Species groups

P, Plants; M, Metazoans; F, Fungi; O, Others

Figure 1.1: The classification system proposed for transposable elements by Wicker et al. 2007. The diverse transposable elements are grouped into classes, subclasses, orders, and superfamilies based on mechanistic and enzymatic criteria. DIRS, *Dictyostelium* intermediate repeat sequence; LINE, long interspersed nuclear element; LTR, long terminal repeat; PLE, Penelope-like elements; SINE, short interspersed nuclear element; TIR, terminal inverted repeat.

TEs depend on the activity of proteins that recognize, mobilize and reintegrate the element, and TEs often contain the necessary genes to be autonomous. For example, Class I retrotransposons of the order LTR contain an ORF coding for a reverse transcriptase, a proteinase, an RNase and an integrase

(Wicker et al. 2007). Mutational events can however impair the ability of TEs to independently produce their transpositional machinery, giving rise to non-autonomous elements. Yet, by exploiting the machinery produced by other TEs, non-autonomous elements remain mobile. MITEs for example are dramatically reduced non-autonomous TEs, consisting only of two TIRs that successfully recruit transposases from other, autonomous Class II DNA transposons.

1.2.2.2 The effect of TEs on the genome and the phenotype

Initially, TEs were considered as selfish parasitic elements that inflate genomes with "junk" sequence (Ohno 1972), without any benefit to the host. This poor reputation was bolstered by the ability of TEs to generate substantial mutations (e.g. Anxolabéhère et al. 1988) – either as a consequence of their mobility or of their high frequency in the genome. TE transposition can disrupt, modify or duplicate genes and regulatory elements, affect alternative splicing and expression patterns, or interfere with epigenetic regulation. In addition, with rising copy-numbers in a genome, the likelihood of aberrant transposition and ectopic recombination events increases, potentially causing large-scale chromosomal rearrangements (Hua-Van et al. 2010). It is thus not surprising that mechanisms evolved to constrain the activity and the disruptive potential of TEs in genomes. For example, TEs are often particularly abundant in genomic regions of low recombination frequency and gene density, and they are assumed to have driven the evolution of epigenetic silencing mechanisms (RK Slotkin and Martienssen 2007; Shabalina and Koonin 2008; Levin and JV Moran 2011). In spite of this apparent menace to genome integrity, the reputation of TEs changed substantially when evidence for TE-induced beneficial genetic innovation gathered (reviewed in Volff 2006; KR Oliver and Greene 2009; Fedoroff 2013). For example, TEs are suspected to be crucial to the evolution of the placenta of Eutheria (Mi et al. 2000) or the evolution of RNAi (RK Slotkin and Martienssen 2007) and long ncRNA (lncRNA, Kapusta et al. 2013). In *Drosophila*, telomeres are maintained not by telomerases but by two non-LTR retrotransposons, HeT-A and TART (Pardue and DeBaryshe 2003), and DNA transposons have been implicated in the radiation of the primate lineage (Pace and Feschotte 2007) and the success of invasive species (Stapley et al. 2015).

These and countless other examples of TE-driven evolutionary innovations highlight that adaptive phenotypes can emerge from TE-induced genetic change. Like for other mutations, most TE-induced changes will be deleterious or nearly neutral and the ratio of beneficial to deleterious mutations is likely to be similar between single-nucleotide and TE-induced mutations (Akagi et al. 2013). However, the potential for major evolutionary innovations appears to be particularly high in TE-induced mutations due to stronger phenotypic effects (Feschotte and Pritham 2007). As transposon-induced genetic change often involves shuffling and rewiring of entire genetic elements, genotypic and thus also phenotypic consequences are often much more severe compared to the gradual changes induced by single nucleotide mutations.

This feature of TE-induced mutations constitutes the basis for the compelling "epi-transposon hypothesis" that advocates a key role for TEs in adaptive evolution and speciation through leaps of rapid phenotypic diversification (Zeh et al. 2009). The hypothesis is furthermore founded on the fact that TE activity fluctuates over evolutionary time and appears to peak during periods of stress (Capy et al. 2000). Either by directly activating transposition or by inhibiting genomic silencing mechanisms (RK Slotkin and Martienssen 2007), environmentally induced physiological or genomic stress liberates TEs, enabling them to restructure the genome. In response to changes in the environment (e.g. climatic), the physiological and genomic stress response thus triggers frequent genetically induced, random phenotypic variation that would provide the necessary diversity in a population for subsequent adaptation through natural selection (Zeh et al. 2009).

1.3 The Environment

Without external, environmental influence, genomic information alone does not suffice to produce fit phenotypes (SF Gilbert 2012b). For environmental cues to have an effect on the phenotype, organisms need to be susceptible and responsive to these stimuli, a phenomenon described as phenotypic plasticity.

As a universal feature of living organism, phenotypic plasticity is a basic principle in biology and many attempts have been made to define its parameters and dimensions (e.g. Smith-Gill 1983; Schlichting and Pigliucci 1995; Via et al. 1995; Schlichting and Pigliucci 1998; West-Eberhard 2003; SF Gilbert and Epel 2009; Fusco and Minelli 2010; Forsman 2014). Even though these definitions often are oversimplifications of a complex and in most cases gradual phenomenon, they provide a useful vocabulary to conceptualize phenotypic plasticity.

The phenomenon of different phenotypes arising from a single genotype can be divided into polyphenism (or phenotypic polymorphism) with discrete, discontinuous morphs and graded plasticity (in which a continuum of morphs is expressed) (Fusco and Minelli 2010). A prominent example of polyphenism is found in eusocial insects, where queens and workers are produced by the same genotype (usually) without intermediate phenotypes. In contrast, graded plasticity describes plastic responses to the environment where a phenotypic trait shows continuous variability, for example body size differences between workers in the same eusocial insect colony. Closely related to the distinction between polyphenisms and graded plasticity are the concepts of developmental conversions and phenotypic modulations (Smith-Gill 1983). Developmental conversion is defined as a phenotypically plastic response that is based on the activation of alternative genetic programs controlling the expression of certain traits (e.g. development) by the perception of specific environmental cues. Hence, developmental conversions can also be described as active, specific, often anticipating plasticities that are based on established and fine-tuned mechanisms to perceive, process and transduce

a particular cue, and subsequently alter an organism's developmental (or physiological, or behavioural) state. In contrast, phenotypic modulation by definition describes nonspecific phenotypic variation in response to environmental stimuli or influences that affect trait expression without changing the underlying genetic program (Smith-Gill 1983). Phenotypic modulation is usually considered a passive, general and responsive plasticity in that the plastic response is a direct consequence of environmental conditions.

While these distinctions provide a useful conceptual framework, they often fail to describe actual forms of phenotypic plasticity, for most cases usually settle between both extremes. For example, the discrete queen-worker dimorphism of social insects can be undermined by the occurrence of intercastes (Heinze 1998) that are thought to result from aberrant developmental processes producing graded phenotypes ranging between queens and workers.

As any organismal trait, phenotypic plasticity is subject to selection and evolutionary change. A key aspect in describing phenotypic plasticity is thus to assess to what extent the plasticity of a trait has been shaped by selection and whether the plasticity itself is adaptive. An evolutionary perspective on phenotypic plasticity is important, because it can help to resolve conceptual inaccuracies resulting from such discriminations as active and passive or specific and unspecific.

For example, if the plasticity of a trait is highly adaptive and has likely evolved under strong selection in a sufficiently predictable environment, it is likely to appear as an active, anticipatory and specific form of plasticity, with highly integrated underlying physiological or developmental networks involved. In contrast, passive phenotypic modulations that are unspecific physical responses to environmental conditions can often be considered neutral and non-adaptive (Schlichting and Pigliucci 1995). However, any form of plasticity likely contains both active and passive components and, likewise, its effect on the phenotype can be adaptive or neutral (Via et al. 1995). Furthermore, phenotypic plasticity can also be maladaptive, in particular under aberrant environmental conditions (Langerhans and DeWitt 2002). Depending on the fitness effects of a phenotypically plastic response, selection will act to either buffer or canalize plasticity (Nijhout 2003). If the plastic response is detrimental, selection acts to decrease plasticity, thus stabilizing the phenotype (homeostasis). However, if the plasticity is beneficial, selection will increase the sensitivity to the inducing environmental stimulus, enabling the evolution of more elaborate responses to the cue – a process described as genetic accommodation (West-Eberhard 2005a). Similarly, exposure to stressful conditions can enable the evolution of adaptive phenotypic plasticity through directional selection, if the resulting phenotype is closer to a new phenotypic optimum than to the one originally favoured (Badyaev 2005; Ghalambor et al. 2007).

1.4 Endosymbionts

It is difficult to apply the distinction between environmental and genetic contributors to the phenotype to endosymbionts. On the one hand, endosymbionts are external factors acquired from the environment, and on the other hand, by contributing their genome to the symbiosis, they can enrich their host's available genetic repertoire.

Endosymbioses have been key in major evolutionary transitions such as the evolution of eukaryotic and auto-phototrophic cells (Dyall 2004). Examples of lesser evolutionary significance are abundant, in particular in insects, where endosymbionts continue to push diversification, speciation and evolutionary innovation (Moya et al. 2008).

The transition from incipient to obligate endosymbiosis is a long-lasting evolutionary process beginning with a loose association between independent partners and ending in complete inter-dependency of host and symbiont. In general, this consolidation involves the emergence of novel, adaptive traits increasing host fitness (and thus also symbiont fitness), and the evolution of vertical transmission of endosymbionts from one host generation to the next (Toft and Andersson 2010; Gil et al. 2010). In addition, genomes of obligate endosymbionts tend to become highly reduced (Moya et al. 2008) and hosts develop specialized, often morphological adaptations to control and foster their endosymbiont population (Toft and Andersson 2010).

The diversity of endosymbionts is particularly well studied in insects, and many different levels of specialization and inter-dependency are recognized in extant symbioses (Kikuchi 2009). In insects, endosymbionts occur extracellularly in the gut, the body cavity or the hemolymph and intracellularly in various tissues or in specialized cells (bacteriocytes) and organs (bacteriome) (Kikuchi 2009; Gil et al. 2010). Bacteriocytes and bacteriomes usually harbour primary symbionts that are beneficial and obligate to the host. Secondary endosymbionts vary in their localization within the host and are considered facultative and non-essential, providing only conditional benefits to the host (Kikuchi 2009).

In general, costs and benefits apply to both partners in a mutualistic endosymbiosis (Herre et al. 1999). For endosymbionts, costs usually involve their loss of independence and benefits include nutrition and a protected microhabitat. Benefits to the host can be diverse. In most cases, endosymbionts provision rare nutrients to complement a host's unbalanced diet, but other beneficial effects have been documented as well, including pathogen resistance (Kaltenpoth and Engl 2014) and increased stress-resistance (JA Russell and NA Moran 2006). In general, these acquired phenotypic traits are expected to offer great fitness advantages to the host, so that benefits of the symbiosis outweigh the imposed costs (Feldhaar 2011). In addition to maintenance costs of providing nutrition and microhabitat, insect hosts often face costs resulting from a conflict of interest over reproductive output between endosymbionts and host. Conflicts arise as endosymbionts strive to bias the host's sex-ratio towards

females through reproductive manipulation, because endosymbionts are in most cases transmitted only by female hosts (Werren et al. 2008; Cordaux et al. 2011).

1.5 *Cardiocondyla obscurior* as a Laboratory Model

Most ant species are challenging to work with in the laboratory, because their colonies require intense care if maintained in an artificial environment. In addition, colonies often comprise several thousand individuals and propagation is in many cases nearly impossible under laboratory conditions. However, among the many different ant species that are studied in laboratories across the world, there are some few exceptional species whose biology and life history allow consistent maintenance and controlled experimental manipulation under laboratory conditions. One of these exceptional species is *C. obscurior* (Figure 1.2). Originally from Southeast Asia, human commerce led to the spread of *C. obscurior* to different habitats in the tropics and subtropics (Heinze et al. 2006), but also in green houses and university offices. Colonies of *C. obscurior* are usually found in plant structures, occupying naturally formed cavities (Seifert 2003) and ephemeral and diverse nest sites, such as furled leaves, aborted fruits or loose bark, require frequent relocation and adaptation to alternative nesting conditions. In the laboratory, colonies of *C. obscurior* readily adapt to artificial nesting conditions in Petri dishes with plaster grounding in a damp climate at 23° C to 30° C.

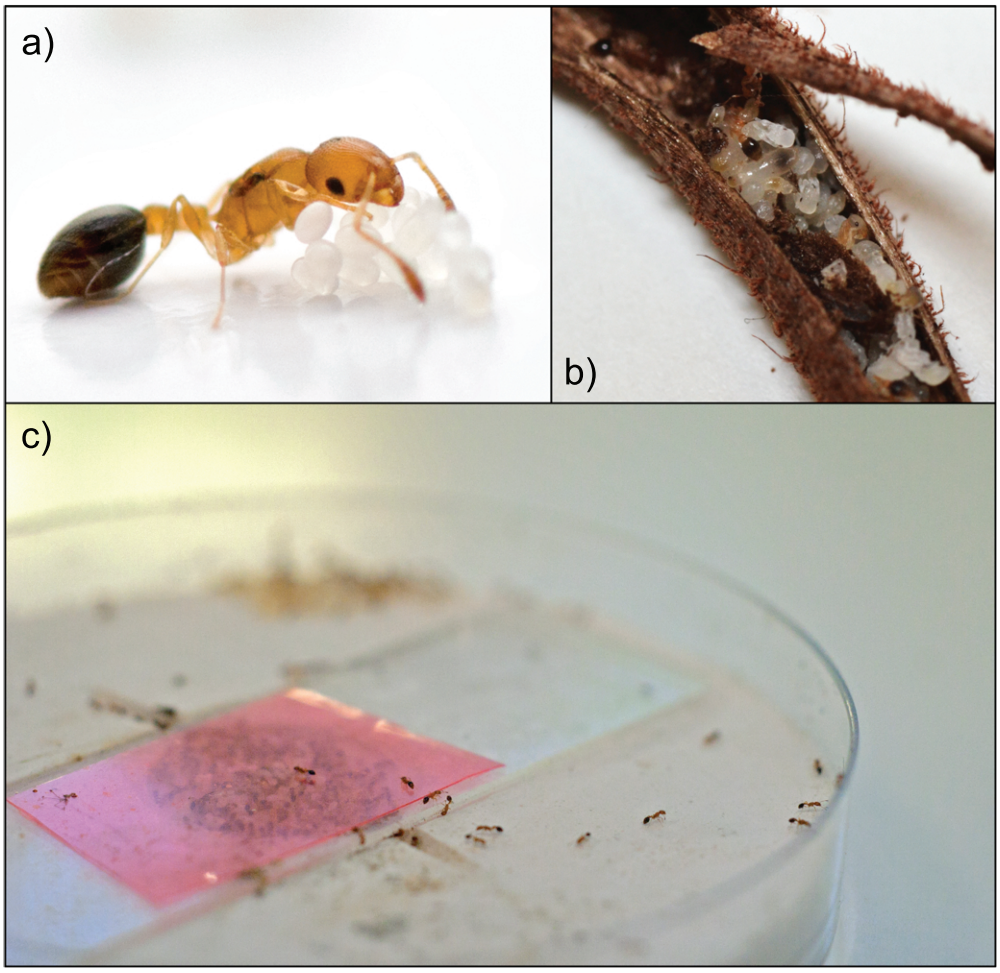


Figure 1.2: The model system *Cardiocondyla obscurior*. (a) Queen of *C. obscurior* tending to a pile of eggs. (b) Colony of *C. obscurior* nesting inside a dried *Nepenthes* leaf. (c) Laboratory colony of *C. obscurior* in a Petri dish.

Consisting of only few workers and a single queen, incipient colonies can be easily split from larger stock colonies (Heinze et al. 2006). Their minute body size (~2 mm) and frugality concerning nesting space and conditions allow for the maintenance and manipulation of several hundred independent colonies from different source populations in the laboratory. Propagation of colonies of *C. obscurior* is remarkably simple, because in contrast to most other ant species, mating in *C. obscurior* takes place regularly within the colony and in most cases between closely related individuals. Fully developed colonies contain multiple queens (polygyny) and, a rarity among ants, non-dispersing males (Kugler 1983). Most species of ants produce winged males that leave the natal colony and disperse soon after reaching maturity (Hölldobler and Wilson 1990). In *Cardiocondyla* however, males remain in the colony where they mate with virgin queens (Kinomura and Yamauchi 1987). This mode of reproduction is linked to substantial adaptations in the male sex, culminating in a remarkable male diphenism with docile winged males and ergatoid fighter males in several species (Oettler et al. 2010). In *C. obscurior*, the vast majority of males produced by a colony develop into the ergatoid phenotype.

While the winged male phenotype is rare, its expression can be triggered by altered rearing conditions (Schrempf and Heinze 2006).

Confrontations between ergatoid males in a colony usually end fatally for one of the combatants, allowing a single victorious male to monopolise reproduction in its colony (Stuart et al. 1987). Such a life history imposes strong selection pressures on males, resulting in striking morphological, behavioural and developmental adaptations in the fighter (Heinze and Hölldobler 1993) but also in the winged phenotype (Cremer, Sledge, et al. 2002b). For example, ergatoid males of *C. obscurior* are long-lived and have life-long spermatogenesis and enlarged mandibles, while winged males mimic queen odour to evade aggression by ergatoid males.

1.6 *Cardiocondyla obscurior* as a Model to Study Phenotypic Evolution and Plasticity

In general, ants are rich models for studies on phenotypic evolution. For example, the discrete queen-worker diphenism and worker polyethism provide sufficient substance for countless studies on the basis of developmental and phenotypic plasticity (e.g. DE Wheeler 1991; Chittka et al. 2012; Yan et al. 2014). In addition, ants are tremendously diverse and often show extreme levels of adaptations to ecological niches (Hölldobler and Wilson 1990; Guénard 2013). However due to the difficulties in cultivating ants under laboratory conditions, it is often difficult to study the mechanisms underlying phenotypic evolution. Among the ant species that are rather easily maintained in the laboratory, *C. obscurior* is particularly well suited to study both developmental plasticity and adaptation in ants.

The species distribution of *C. obscurior* is wide, including habitats in Bahia (Brazil), Okinawa (Japan), Mississippi (USA), Holguin (Cuba), Chiapas (Mexico), Alajuela (Costa Rica), Al Bahah (Saudi Arabia), Fiji, and Tenerife (Spain) (*AntWeb*, Macgown 2012, A. Schrempf pers. comm.). Even though environmental differences between habitats might be small, each colonization event will require some level of local adaptation (Reznick and Ghalambor 2001). In incipient, introduced populations of *C. obscurior*, genetic diversity is expected to be extremely low due to genetic bottlenecks, high levels of inbreeding, and low effective population size (N_e) (Nei et al. 1975; D Charlesworth and Wright 2001; Romiguier et al. 2014). In spite of these constraints, *C. obscurior* appears to be able to establish stable populations following the introduction to novel habitats, indicating a high level of adaptability. The rapid production of locally adapted phenotypes is particularly important in invasive species that face different environmental conditions following introduction (Prentis et al. 2008).

The possibility to maintain colonies from different source populations of *C. obscurior* offers the opportunity to study rapidly emerging phenotypic adaptations under controlled laboratory conditions.

In addition, the relatively short generation time of approximately 4 weeks, and the species' disposition for inbreeding, allow for genetically isolated inbred lines to be maintained over several generations, enabling for example experimental evolution or studies on the genetic and environmental determinants of phenotypic traits. Similarly, the ability to extensively manipulate colonies and alter rearing conditions provides powerful tools to affect development in studies on the basis of (environmentally induced) polyphenism (Schrempf and Heinze 2006; Du et al. 2007).

1.7 Aims of this Thesis

The fields of ecology, development and evolution are growing closer, for an increasing appreciation of environmental effects on phenotype and evolution. The emerging field of eco-evo-devo ("ecological, evolutionary developmental biology") aims at building an evolutionary concept that integrates development, environment, and genes to explain how novel phenotypes can evolve (Abouheif et al. 2013). In essence, the concept revolves around the environment's potential to induce novel phenotypic and genetic variants that are shaped by developmental processes and constraints. Novel phenotypes and genotypes are thus exposed to natural selection, allowing evolution to proceed.

The aim of this thesis is to assess principles and mechanisms of phenotypic evolution in *C. obscurior* in the context of eco-evo-devo. Focusing on the genomic basis of rapid adaptation (Chapter 2), the genetics of developmental plasticity (Chapters 3 and 4), and the role of mutualism in adaptation (Chapter 5), this thesis provides insight into the interactions of environment, genes and development and their role in evolutionary progress. In addition, the genomic and transcriptomic resources developed in this thesis serve as valuable groundwork for further studies on the evolutionary biology of this astonishing species.

Chapter 2

2. Transposable Element Islands Facilitate Adaptation to Novel Environments in an Invasive Species

Lukas Schrader[1,11], Jay W. Kim[2], Daniel Ence[3], Aleksey Zimin[4], Antonia Klein[1], Katharina Wyschetzki[1], Tobias Weichselgartner[1], Carsten Kemena[5], Johannes Stökl[1], Eva Schultner[6], Yannick Wurm[7], Christopher D. Smith[8], Mark Yandell[3,9], Jürgen Heinze[1], Jürgen Gadau[10], Jan Oettler[1,11,12]

[1] Institut für Zoologie, Universität Regensburg, 93053 Regensburg, Germany

[2] Department of Biomolecular Engineering, University of California at Santa Cruz, Santa Cruz, CA 95064, USA

[3] Eccles Institute of Human Genetics, University of Utah, Salt Lake City, Utah 84112, USA

[4] Institute for Physical Sciences and Technology, University of Maryland, College Park, MD 20742, USA

[5] Institute for Evolution and Biodiversity, Westfälische Wilhelms-Universität, 48149 Münster, Germany

[6] Department of Biosciences, University of Helsinki, 00014 Helsinki, Finland

[7] School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK

[8] Department of Biology, San Francisco State University, San Francisco, CA 94132, USA

[9] Utah Center for Genetic Discovery, University of Utah, Salt Lake City 84112, USA. 10

[10]School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA

[11] Contributed equally

[12] Corresponding author:
Jan Oettler
Institut für Zoologie,
Universität Regensburg,
Universitätsstr. 31,
93053 Regensburg, Germany Tel +49 9419432996, Fax +499419433304
joettler@gmail.com

Running title: Mutation hotspots in TE islands

Keywords: *Cardiocondyla obscurior*, Genetic drift, Adaptation, Transposable Elements

Abstract

Adaptation requires genetic variation, but founder populations are generally genetically depleted. Here we sequenced two populations of an inbred ant that diverge in phenotype to determine how variability is generated. *Cardiocondyla obscurior* has the smallest of the sequenced ant genomes and its structure suggests a fundamental role of transposable elements (TEs) in adaptive evolution. Accumulations of TEs (TE islands) comprising 7.18 % of the genome evolve faster than other regions with regard to single nucleotide variants, gene/exon duplications and deletions and gene homology. A non-random distribution of gene families, larvae/adult specific gene expression, and signs of differential methylation in TE islands indicate intragenomic differences in regulation, evolutionary rates, and coalescent effective population size. Our study reveals a tripartite interplay between TEs, life history, and adaptation in an invasive species.

2.1 Introduction

Depletion of genetic variation is detrimental to species evolution and adaptation (D Charlesworth and B Charlesworth 1987). Low genetic and phenotypic variation is common in founder populations, where only one or a few genotypes are isolated from a source population. Under such conditions, reduced effective population size (N_e) should decrease selection efficiency and increase genetic drift, resulting in only weak selection against mildly deleterious alleles, which can thus accumulate (Lynch 2007). These effects should be even stronger in inbreeding species (D Charlesworth and Wright 2001) and taxa with generally low Ne such as social insects (Romiguier et al. 2014). Despite these constraints on adaptive evolution, many inbred or selfing species thrive and are able to invade novel habitats. This raises the question of how genetic variation as the raw material for adaptation is generated in such systems.

Single nucleotide substitutions are an important factor in adaptation (McDonald and Kreitman 1991) and species diversification (Lanfear et al. 2010; Lynch 2010). However, other structural and regulatory units, such as transposable elements (TEs) and epigenetic modifications, may act as drivers in adaptation and evolution (Fontdevila 2011). TEs play a particularly vital role in genome evolution (Fedoroff 2013), and recurringly generate adaptive phenotypes (Madlung and Comai 2004; González et al. 2010; Rostant et al. 2012; Casacuberta and González 2013) primarily through (retro-)transposition (Kazazian 2004), and secondarily through ectopic recombination and aberrant transposition (Hua-Van et al. 2010).

The invasive, inbreeding ant *Cardiocondyla obscurior* (Figure 2.1) provides a suitable model to study how species adapt to novel habitats in spite of constraints imposed by invasion history, life history, or both. Originally from Southeast Asia, *C. obscurior* has established populations in warm climates around the globe from founder populations that presumably consisted of only one or a few inbred colonies, each with a few reproductive queens and several dozen sterile workers. In this species, related wingless males and females (queens) mate within the colony, after which queens leave the colony with a group of workers to found a new nest nearby. While greatly reducing the extent of gene flow between colonies, this behaviour enables sexual reproduction within the same colony and allows single founder colonies to rapidly colonize novel habitats. At the same time, the combination of prolonged inbreeding with severe genetic bottlenecks strongly reduces N_e in this species. Under such conditions, genetic drift is predicted to drastically deplete genetic variation, thus leaving little for selection to act on.

Figure 2.1: Two workers of *Cardiocondyla obscurior* and the remains of a fly. Hidden in small cavities of plants, the inconspicuous colonies of this species are frequently introduced to new habitats by global commerce. In spite of strong genetic bottlenecks, even single colonies with few reproductive individuals suffice to establish stable populations.

Here we explore the genomes of *C. obscurior* from two invasive populations (Brazil BR and Japan JP) to identify signatures of divergence on a genomic level and to determine how the species can rapidly adapt to different habitats. We find clear phenotypic differences between the populations and strong correlation between accumulations of TEs ("TE islands") and genetic variation. Our results suggest that TE islands might function as spring-wells for genetic diversification in founder populations of this invasive species. The distinct organization of TE islands, their gene composition, and their regulation by the genome adds compelling evidence for the role of TEs as players in differentiation, adaptation, and speciation.

2.2 Results

Phenotypic differences between BR and JP lineages

Colonies from the two populations contained similar numbers of workers (Mann-Whitney-U= 778.5, Z= -0.634, p= 0.526; BR: median= 28, quartiles 21.75 and 51.25, n= 27 colonies; JP: median= 29, quartiles 16 and 47, n= 64), but queen number was higher in Japan (Mann-Whitney-U= 501, Z= -3.084, p< 0.003; BR: 5 queens, quartiles 3, 8; JP: median= 10, quartiles 4 and 19). Body sizes of queens and workers from BR were significantly smaller than in JP individuals, yet wingless males did not differ in any of the measured characters (see Supplement).

In ants, cuticular chemical compounds play a particular prominent role in kin recognition, which is crucial for species integrity but on a deeper level also a requirement for the maintenance of altruism (van Zweden et al. 2010). Analysis of cuticular compound extracts from BR and JP workers showed that compound composition differed significantly between the two lineages (MANOVA: df= 2, F= 10.33, R2= 0.39, p < 0.001) and samples were classified correctly according to population of origin in 83.3 % of cases (Supplementary Table S1.1; Supplementary Figure S1.1).

The lineages also differed in behaviour, with BR colonies being significantly more aggressive towards both workers and queens from their own lineage, while JP colonies more readily accepted JP workers and queens ($p_{Workers}$ JPxJP vs. BRxBR= 0.000296, p_{Queens} JPxJP vs. BRxBR= 7.98e-07, Supplementary Figure S1.2). Confronted with individuals from the other lineage, BR colonies were as aggressive as in within-population encounters ($p_{Workers}$ BRxJP vs. BRxBR= 0.39, p_{Queens} BRxJP vs. BRxBR= 0.94), while JP colonies were again significantly less aggressive ($p_{Workers}$ JPxBR vs. BRxBR= 0.000131, p_{Queens} BRxJP vs. BRxBR= 1.23e-07). Testing discrimination against workers of another ant species, *Wasmannia auropunctata,* evoked similarly high aggressive responses in both lineages, suggesting that the BR and JP populations do not generally differ in their aggressive potential.

The *C. obscurior* genome is compact and rich in class I TEs

Using MSR-CA version 1.4, we produced a 187.5-Mb draft reference genome based on paired-end sequencing of several hundred diploid females (454 Titanium FLX sequencing) and a 200-bp library made from five haploid males (Illumina HiSeq2000) (Supplementary Table S1.2), all coming from a single Brazilian colony. Automatic gene annotation using MAKER version 2.20 (Holt and Yandell 2010) was supported by 454 RNAseq data of a normalized library made from a pool of all castes and developmental stages. We filtered the assembly for prokaryotic scaffolds and reduced the initial 11,084 scaffolds to 1,854 scaffolds, containing all gene models and a total of 94.8 % (177.9 Mb) of the assembled sequence. The genome can be accessed under antgenomes.org and hymenopteragenome.org.

The final gene set contains 17,552 genes, of which 9,552 genes have a known protein domain as detected by IPRScan (ebi.ac.uk/interpro/), and falls within the range of recent estimates for eight other sequenced ant species (Weinstock et al. 2006; Werren et al. 2010; Bonasio et al. 2010; CR Smith et al. 2011b; Suen et al. 2011; CD Smith et al. 2011a; Wurm et al. 2011; Nygaard et al. 2011; Oxley et al. 2014). Of all genes, 72.5 % have an annotation edit distance of less than 0.5, which is consistent with a well-annotated genome (Yandell and Ence 2012) (Supplementary Table S1.3).

The *C. obscurior* genome is the smallest so far sequenced ant genome (Weinstock et al. 2006; Werren et al. 2010; Bonasio et al. 2010; CR Smith et al. 2011b; Suen et al. 2011; CD Smith et al. 2011a; Wurm et al. 2011; Nygaard et al. 2011; Oxley et al. 2014). Although there is no physical genome size estimate for *C. obscurior*, assembled sequences and physical estimates are tightly correlated in seven ant genomes (LM in R: R^2= 0.73, $F_{1,5}$= 13.7, p= 0.014, from Gadau et al. 2011), suggesting that *C. obscurior* has the smallest genome reported so far for an ant species (Tsutsui et al. 2008). Overall, draft genome size of the analysed sequenced ants is negatively correlated to relative exon content (GLM in R: df= 6, F= 150.55, p < 0.001) but not to relative intron content (df= 5, F= 0.65, p= 0.460; Figure 2.2), indicative of stabilizing selection on coding sequence. In contrast, intron size distribution is diverse between ant genomes and is not correlated with genome size (Supplementary Figure S1.3; Supplementary Table S1.4).

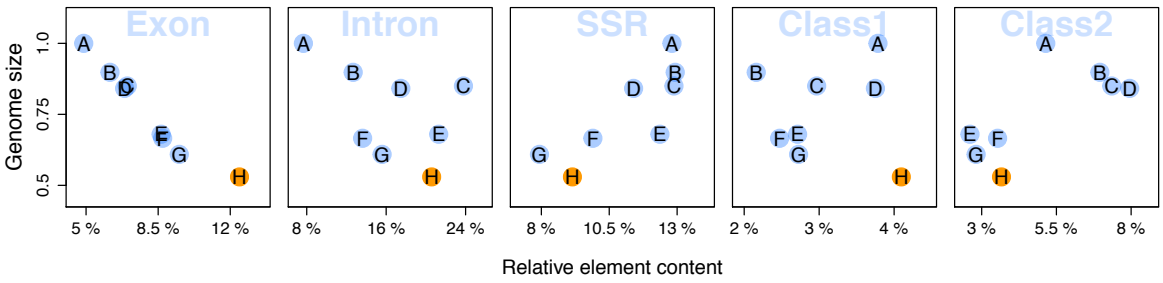


Figure 2.2: Assembly size in Mbp plotted against the relative proportion of exons, introns, and different repetitive elements. The analysed genomes show a negative correlation between relative exon but not intron content. Genome size is positively correlated with relative short simple repeat (SSR) but not Class I and II TE content. A= *S. invicta*, B= *A. cephalotes*, C= *A. echinatior*, D= *H. saltator*, E= *C. floridanus*, F= *P. barbatus*, G= *L. humile*, H= *C. obscurior*.

We used a custom pipeline (see Supplementary Information) to identify simple repeats, Class I retrotransposons, and Class II DNA transposons in *C. obscurior*, seven ant genomes (*Acromyrmex echinatior* (*Aech*), *Atta cephalotes* (*Acep*), *Solenopsis invicta* (*Sinv*), *Linepithema humile* (*Lhum*), *Pogonomyrmex barbatus* (*Pbar*), *Harpegnathos saltator* (*Hsal*), *Camponotus floridanus* (*Cflo*)), the parasitic wasp *Nasonia vitripennis* (*Nvit*), and the honeybee *Apis mellifera* (*Amel*). Across the analysed ants, genome size is significantly correlated with relative simple repeat content (lm, R^2= 0.66, F= 11.83, p= 0.014; Figure 2.2) but not with Class I and Class II TE content. However,

it appears that the larger genomes contain more relative Class II sequence. Relative Class I retrotransposon content was highest in *C. obscurior* (7.6 Mb, 4.31 %, Supplementary Figure S1.4) and in particular, many Class I non-LTR retrotransposons (e.g. 14 types of LINEs) and several types of LTR transposons (Ngaro, Gypsy, DIRS, and ERV2), TIR elements (e.g. hAT, MuDR, P), and Helitrons are more abundant in *C. obscurior* (Supplementary Table S1.5).

Genomic signatures of an inbred lifestyle

Based on TE content calculations for 1- and 200-kb sliding windows, we identified 18 isolated "TE islands" located in "LDR" (low density regions) in the *C. obscurior* genome (Figure 2.3). These TE islands were defined containing TE accumulations in the 95-100 % quantile within scaffolds over 200 kb (87 scaffolds, representing 96.02 % or 170.8 MB of the assembly). In total TE islands cover 12.78 Mb of sequence (7.18 % of total sequence) and range between 0.19 and 1.46 Mb in size. The TE islands contain 27.54 % (4.92 Mb) of the assembly-wide TE sequence (17.87 Mb), 6.6 % of all genes (1,160), and have reduced exon content (TE islands 87.0 exon bp kb^{-1}, LDRs 124.5 exon bp kb^{-1}). Note that some larger scaffolds contain more than one TE island.

Retroelements of the superfamilies BEL/Pao, DIRS, LOA/Loa, Ngaro, R1/R2, and RTE as well as DNA transposons of the superfamilies Academ, Kolobok-Hydra, Maverick, Merlin, on, and TcMar-Mariner/-Tc1 populate TE islands with significantly higher copy numbers than other elements (Fisher's exact test, FDR<0.05, Figure 2.4, Supplementary Table S1.6). Furthermore, both Class I and Class II elements show a length polymorphism, with elements in TE islands being significantly longer compared to elements in LDRs (U-tests, W= 109,089,018, p< 2e-16 for Class I and W= 152,340,067, p< 2e-16 for Class II, Bonferroni corrected, Figure 2.6a, Supplementary Figure S1.5).

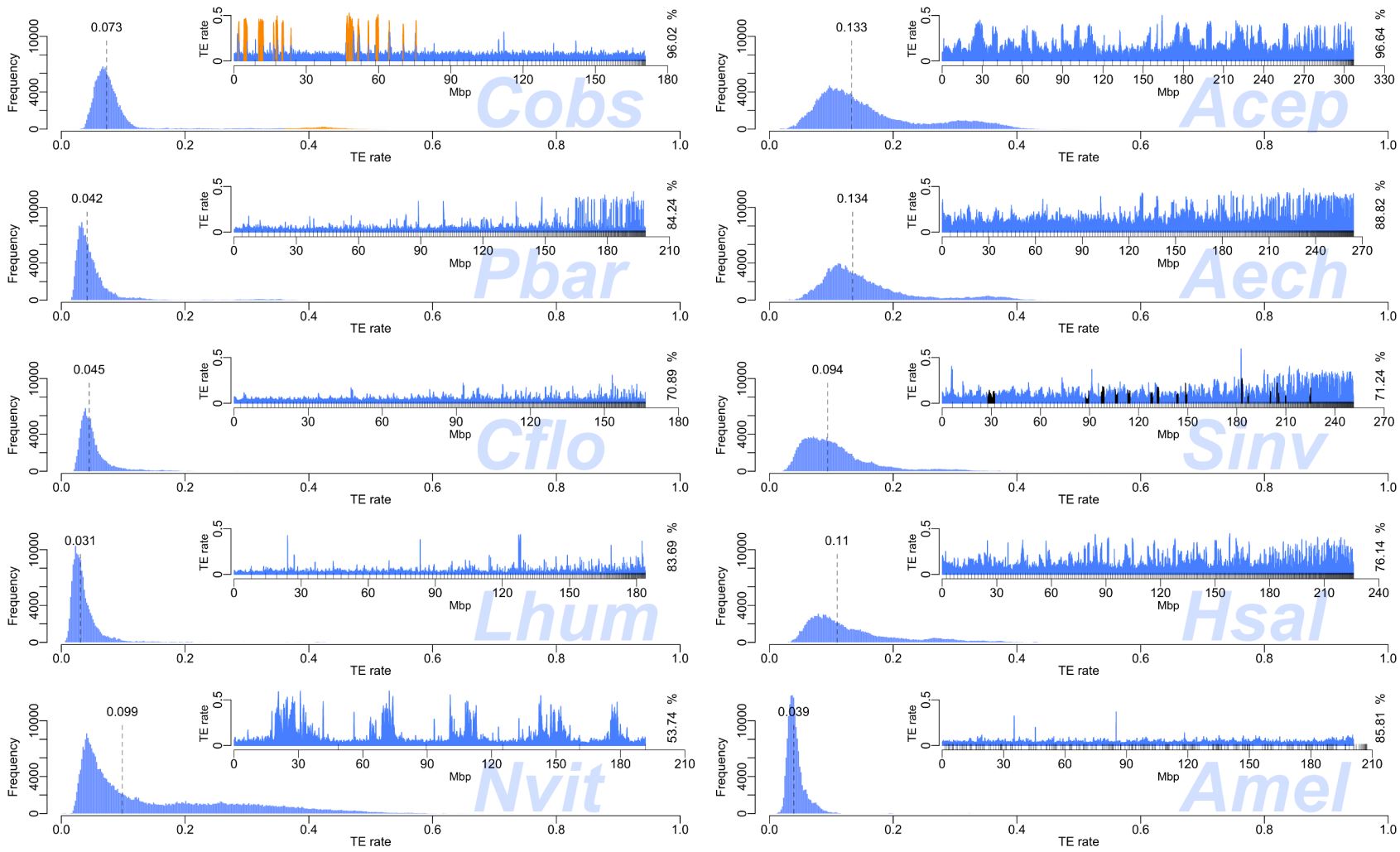


Figure 2.3: Frequency and distribution (insert plots) of TE content in 200-kb windows. Frequency plots: Dashed lines denote median TE content. Distribution plots: Different proportions of total draft genome sequence were analysed (in %), depending on assembly quality. Scaffolds are sorted by size; small upward tick marks indicate scaffold boundaries. For *C. obscurior*, regions defined as TE islands are coloured in orange. For *S. invicta*, scaffolds mapping to a non-recombining chromosomal inversion (John Wang et al. 2013b) are shown in black. For *A. mellifera* scaffolds were sorted according to linkage group.

We also assessed genome-wide TE distributions for seven published ant genomes, *Amel* v4.5, and *Nvit* v2.0 (Figure 2.3). The smaller ant genomes (*Pbar*, *Lhum,* and *Cflo*) and *Amel* are similar in TE sequence distribution. In contrast, the larger genomes (*Aech*, *Acep*, *Sinv,* and *Hsal*) are more variable, have higher median TE content, and a much broader and tailed TE frequency distribution with longer stretches of high or low TE content. The genome of *C. obscurior* is distinct from the other ant genomes, with low TE content in LDRs but exceptional clustering with high TE densities in TE islands. The genome of the inbred wasp *N. vitripennis* contains regions with up to 60 % TE content that are surrounded by LDRs containing much less TE sequence (~10 %), resembling the pattern observed in *C. obscurior*.

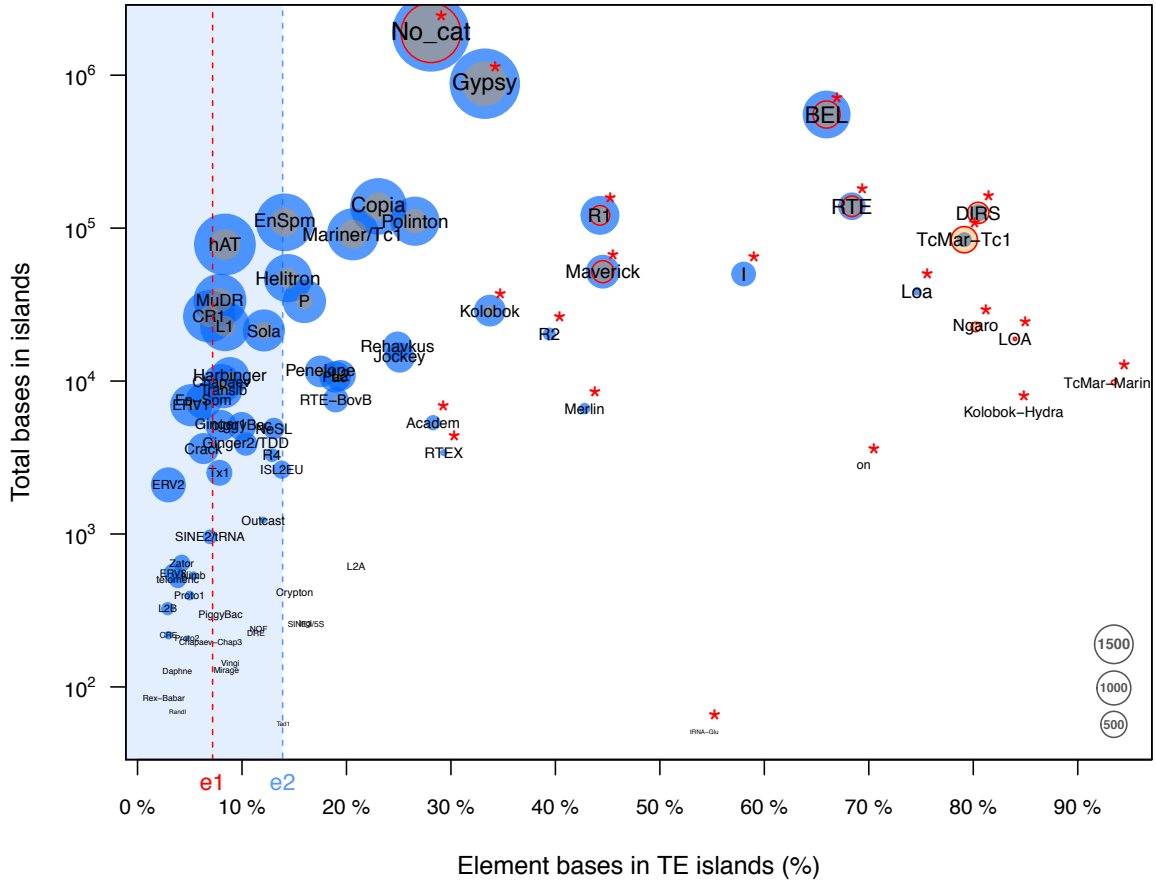


Figure 2.4: The proportion of bases annotated in TE islands in *C. obscurior* against the log-scaled total base count in TE islands for each TE superfamily. Point size is relative to the copy number of the respective element found in TE islands (orange) and in LDRs (blue). Red circles indicate superfamilies with significantly higher frequency in TE islands than other superfamilies. Superfamilies with a significantly higher base count in TE islands are denoted by a red asterisk.

TE islands diverge faster than LDRs in the two populations

We mapped ~140 Gb of genomic DNA Illumina reads (~60 x coverage for each population) from pools of 30 (BR) and 26 (JP) male pupae, respectively, against the reference genome (BWA;

bio-bwa.sourceforge.net) and analysed local coverage ratios to detect genetic divergence. Deviations from the mean coverage ratio (Figure 2.5) are in part caused by sequence deletions, insertions and duplications (Medvedev et al. 2009). Such variations are particularly frequent in TE islands (Figure 2.5, Figure 2.6b), suggesting accelerated divergence within islands (median deviation from mean coverage ratio: 0.288 in TE Islands, 0.163 in LDRs; U-test, W= 640,300,902; p< 2e-16, Bonferroni corrected).

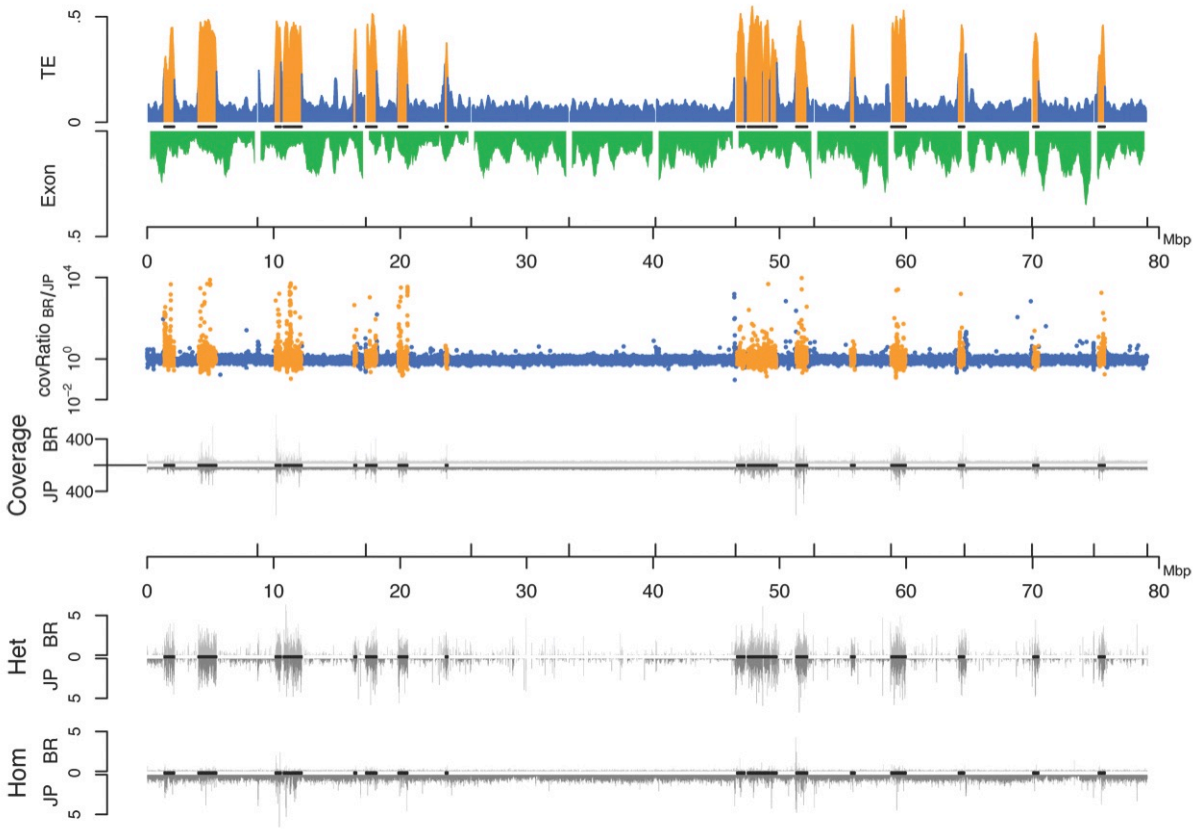


Figure 2.5: Genomic divergence and subgenomic structure of the 12 largest *C. obscurior* genome scaffolds (including all 18 TE islands). High TE content in TE islands correlates with deviations from the average coverage ratio, very high absolute coverage in both lineages, and high numbers of SNV calls. First track: Relative TE (blue and orange within TE islands) and exon content (green) per 200 kb. Second track: Coverage ratio BR/JP (blue and orange within TE islands). Third track: Absolute coverage for BR (top) and JP (bottom). Fourth track: Heterozygous SNV calls per kb in BR (top) and JP (bottom) relative to the reference genome. Fifth track: Homozygous SNV calls per kb in BR (top) and JP (bottom) relative to the reference genome. Black lines on x-axes indicate localization of TE islands.

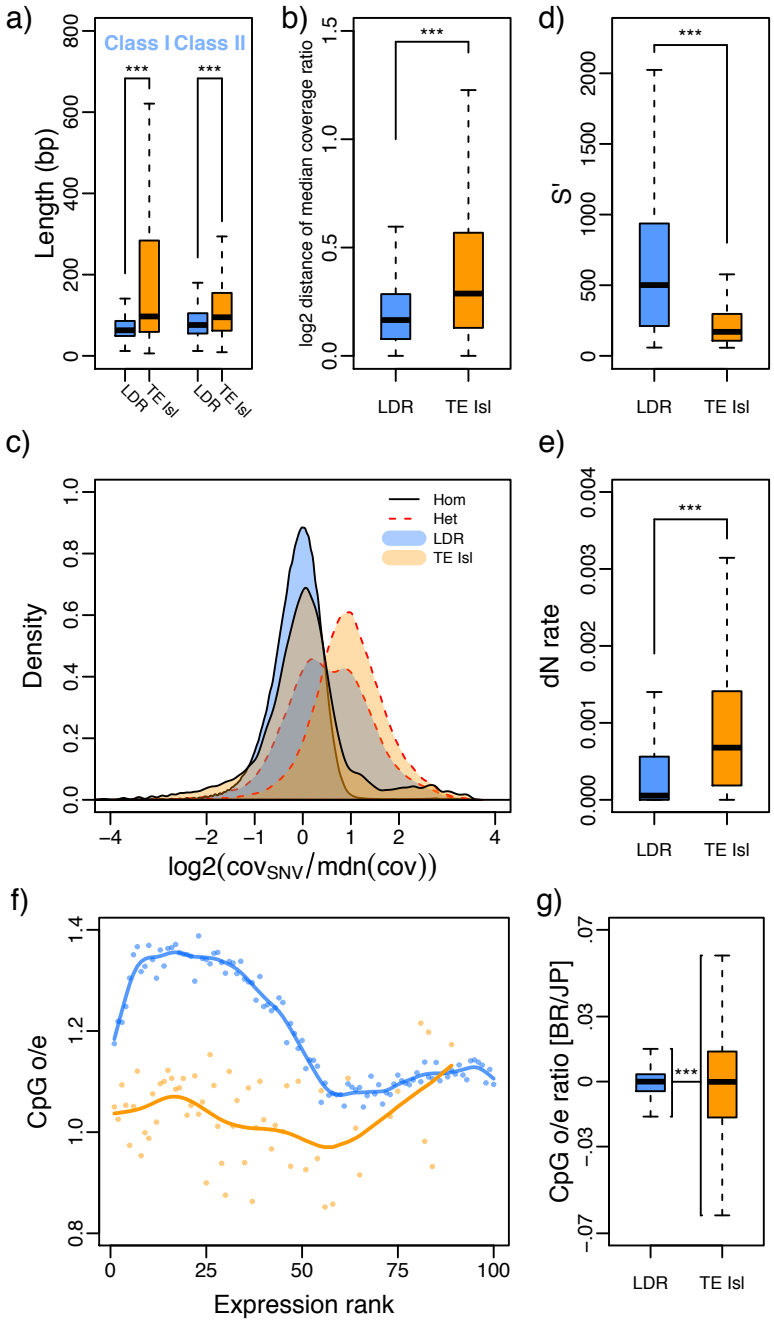


Figure 2.6: Quantitative measures on the divergence of TE islands and LDRs. (a) Length polymorphism for Class I and Class II transposable elements in LDRs (blue) and TE islands (orange). U-tests, $n_{LDR=}$ 54,950, $n_{TE=}$ 6,466 for Class I and $n_{LDR=}$ 59,054, $n_{TE=}$ 6,813 for Class II. (b) Deviations from the median coverage ratio calculated for 1 kb windows in LDRs (blue) and TE islands (orange). U-test, $n_{LDR=}$ 157,296, $n_{TE=}$ 12,165. (c) Log2-scaled density plots of the coverage for all homozygous (solid black lines) and heterozygous SNV (dotted red lines) calls divided by the median coverage (orange= calls within TE islands; blue= calls in LDRs). Coverage at homozygous calls is not different from the median overall coverage, neither in TE islands nor in LDRs. The shift for heterozygous SNV calls within TE islands shows that most calls result from diverging duplicated loci. The bimodal distribution for heterozygous calls in other genomic regions suggests two distinct populations of SNV calls, i.e. true heterozygous loci (first peak) and diverging sequence in duplicated loci (second peak). (d) Bitscores for genes in LDRs (blue) and TE islands (orange) retrieved by BLASTx against annotated proteins from seven ant genomes. U-test, $n_{LDR=}$ 12,065, $n_{TE=}$ 902. (Continued on next page)

Figure 2.6e-g: (e) Rates of non-synonymous substitutions (calculated as dN/(dN+dS)) in LDR (blue) and TE island genes (orange). U-test, $n_{LDR=}$ 6,806, $n_{TE=}$ 423. (f) Exon-wide CpG o/e values were plotted against the expression rank from 0 (least expressed) to 100 (most expressed) genes for LDRs (blue) and TE islands (orange). (g) Calculated ratios (BR/JP) for exon CpG o/e values in LDRs (blue) and TE islands (orange). F-test, $n_{LDR=}$ 16,379, $n_{TE=}$ 1,159. (***: p< 0.0001, boxplots show the median, interquartile ranges (IQR) and 1.5 IQR.)

We retrieved SNV (single nucleotide variant) calls using consensus calls from samtools (samtools.sourceforge.net) and the GATK (broadinstitute.org/gatk/). Although TE islands only comprise 7.18 % of the genome, they combine 15.59 % (86,236 of 553,052) of all SNV calls. Given that we sequenced haploid males from highly inbred lineages, heterozygous SNVs should be rare. A large fraction of heterozygous SNVs in both lineages are within TE islands (62.95 % of 62,879 in BR, 50.52 % of 98,353 in JP), while rates of homozygous calls (Figure 2.5) are not increased (11.88 % of 16,277 in BR, 6.91 % of 445,316 in JP). High numbers of false positive heterozygous SNV calls can arise in duplicated regions that collapsed into a single locus due to misassemblies (Treangen and Salzberg 2012). Accordingly, such SNVs can be identified by a two-fold increase in coverage and in fact mark diverging duplicated loci within the same lineage (Figure 2.6c).

Genes in TE islands should also show signatures of accelerated divergence from orthologs if overall sequence evolution is increased in these regions. Indeed, BLASTp searches against seven ant proteomes produced significantly lower bit scores for genes within TE islands than for genes in LDRs (Figure 2.6d, U-test, W= 120,460,260, p< 2e-16). In accordance, SNV annotation revealed higher rates of non-synonymous substitutions between the BR and JP lineage in TE island genes (Figure 2.6e, U-test, W= 923,754, p< 2e-16). Surprisingly however, on average TE island genes contained less synonymous SNVs than LDR genes (LDR 0.67 kb^{-1}, TE island 0.42 kb^{-1}, U-test, W= 10,743,397, p< 2e-16).

Copy number variation within and between TE islands

We inspected 512 candidate loci (155 in TE islands) of 1 kb length, by plotting the coverage of each lineage relative to SNVs, genes and TEs at the respective position, to find genes potentially affected by deletion or copy number variation events and compiled a list of 89 candidate genes (Supplementary Table S1.7). Experimental proof-of-principle was conducted by PCR and Sanger sequencing for two deletion candidates (*Cobs_13563* and *Cobs_01070*) and by real-time quantitative PCR for four duplication candidates (*Cobs_13806*, *Cobs_17872*, *Cobs_13486*, and *Cobs_16853*) (Supplementary Figure S1.7).

A majority of these genes are located in TE islands (61.8%) and 34 genes show at least weak expression in BR individuals in RNAseq data (see below). The affected genes play roles in processes that may be crucial during invasion of novel habitats, such as chemical perception, learning, and

insecticide resistance. In particular, four different odorant/gustatory receptor genes show signs of either multiple exon (*Cobs_05921*, *Cobs_13418*, *Cobs_14265*) or whole gene duplication (*Cobs_17892*). A gene likely involved in olfactory learning, *Cobs_13711* a homolog to *pst* (Dubnau et al. 2003), also shows signs of duplication. Three genes homologous to fatty acid synthase (FAS) genes, a key step in cuticular odour production, contain partial deletions (*Cobs_16510*, *Cobs_14262*) or duplications (*Cobs_15866*). Furthermore, we found differences in genes associated with insecticide response (*Cobs_00487*, a homologue of *nAChR*α6 (FBgn0032151) (Millar and Denholm 2007) and *Cobs_17834*, coding for a homologue to Cyp4c1 (EFN70878.1) (Hemingway and Ranson 2000). Other key genes affected are associated with circadian rhythm (*Cobs_17789*, homologue to *per* (FBgn0003068)), caste determination (*Cobs_01070*, with homology to *Mrjp1* (gi406090) (Drapeau et al. 2006), development (*Cobs_17755*, coding for a homolog of VgR (Q6X0I2.1) (M-E Chen et al. 2004), and aging (*Cobs_14758*, with homology to *Mth2* (FBgn0045637) (Duvernell et al. 2003).

De novo assembly of ~23 M Illumina paired-end reads from the JP lineage that could not be mapped to the BR reference genome resulted in 17 contigs after filtering with highly significant BLASTx hits against proteins of other ants, suggesting that these conserved sequences were lost in the BR lineage instead of being gained in the JP lineage. According to functional annotation, among others these contigs code for homologs involved in development (Vitellogenin-like (XP_003689693)) (LI Gilbert 2012a), cellular trafficking (Sorting nexin-25 (EGI65030)) (Worby and Dixon 2002), immune response (Protein Toll (EGI66069)) (LI Gilbert 2012a), and neuronal organization (Peripheral-type benzodiazepine receptor-associated protein 1 (EFN68490)) (Galiegue et al. 1999) (Supplementary Table S1.8).

Gene composition and regulation of TE islands

Increased TE activity may incur costs to fitness by disrupting gene function. A two-tailed Gene Ontology (GO) enrichment analysis revealed that 59 GO terms associated with conserved processes (e.g. cytoskeleton organization, ATP binding, organ morphogenesis) are underrepresented in TE islands, while 18 GO terms are enriched (Supplementary Tables S1.9 and S1.10). Four of the overrepresented terms relate to olfactory receptors (ORs) (GO:0004984, GO:0005549, GO:0050911, GO:0007187) and two terms relate to FAS genes (GO:0005835, GO:0016297). The remaining twelve terms most likely relate to TE derived genes.

Gene body CpG depletion as a result of increased CpG to TpG conversion due to cytosine methylation is a measure for germline methylation (i.e. epigenetic regulation) in past generations. In TE island genes, the exon-wide median observed/expected (o/e) CpG ratio is significantly lower than in other genes (t-test, TE island genes: 1.05, LDR genes: 1.20, $p < 1e-16$). However, both sets of genes show

strikingly different correlations of expression and o/e CpG values (Figure 2.6f). For LDR genes, o/e CpG values are high in moderately expressed genes and low in highly expressed genes. In contrast, in TE islands, weakly to moderately expressed genes contain less CpG dinucleotides, while highly expressed genes have higher o/e CpG values. To furthermore identify traces of differential regulation of TE islands, we compared exon o/e CpG values between the lineages by calculating BR/JP ratios for each exon's o/e CpG values and found higher variance in BR/JP ratios in TE islands than in LDRs (Figure 2.6g, F-test, F= 0.136, p< 2e-16, ratio of variances= 0.136, Bonferroni corrected).

Finally, to assess whether gene expression levels differed between LDRs and TE islands we generated ~14 Gb and ~17 Gb transcriptomic RNAseq data of seven queens and seven queen-destined larvae (3rd larval stage), respectively, from the BR lineage. We estimated mean normalized expression values for each gene using DESeq2 (bioconductor.org/packages/release/bioc/html/DESeq2.html), revealing that expression in TE islands was much lower than in LDRs (median expression of all LDR genes= 25.45; in TE islands: 0.49; U-test, W= 14,461,310, p< 2e-16). While larvae and adult queens did not differ in the expression of LDR genes (median expression in queens= 21.16; in larvae= 23.72; U-test, W= 133,301,709, p= 0.221), TE island genes were more expressed in adult queens (median expression in queens= 0.84; in larvae= 0; W= 1,031,038, p< 2e-16) (Figure 2.7, see Supplementary Figure S1.6 for details on differential expression between queen and larvae).

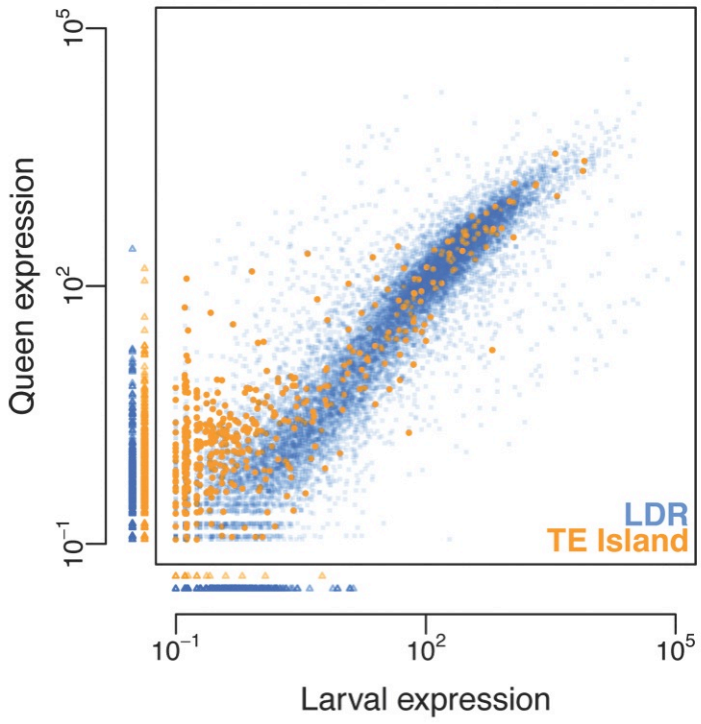


Figure 2.7: Mean normalized expression in 3rd instar queen larvae and mated adult queens for all Cobs1.4 genes. Small triangles indicate genes with no expression in queens (plotted below the x axis) or larvae (plotted left to the y axis). Ninety-five TE island genes and 1,382 LDR genes were not expressed at all (orange= TE island genes; blue= LDR genes).

2.3 Discussion

Cardiocondyla obscurior is a textbook example for successful biological invasion. Its small size allows for interspecific avoidance, it can rapidly establish colonies in disturbed habitats, and multiple generations per year allow for fast adaptation. While variation in CHCs and body size between the populations point to adaptations to different environments, higher queen number in the JP lineage is likely correlated with reduced intra-specific aggression.

The small genome of *C. obscurior* differs markedly from the other analysed ant genomes in TE distribution and overabundance of several Class I subclasses. Importantly, the genome contains low frequencies of TEs in LDRs but well-defined islands with high densities of TEs. In these islands, TEs are on average longer than in LDRs, suggesting overall higher TE activity (Kaminker et al. 2002). Differences in mutation rates and sequence divergence between LDRs and TE islands reveal distinct evolutionary dynamics acting within the *C. obscurior* genome. Moreover, in TE islands, key genes are removed and the majority of genes are less expressed in larvae than adult queens. The non-random distribution of TEs suggests that intra-genomic differences in selection efficiency against TEs may have further supported the formation of such locally confined TE accumulations.

Inbreeding can facilitate the accumulation of TEs (D Charlesworth and Wright 2001) and repeated exposure to stress induced by novel environmental conditions can further amplify TE proliferation (Capy et al. 2000). Small N_e is expected to increase the effects of genetic drift and in turn reduce selection efficiency against mildly deleterious mutations (Lynch 2007). Under such conditions, local accumulations of TEs might have formed in genomic regions under relaxed selection. Similarly, a reduction in N_e in inbred *Drosophila* leads to a shift in the equilibrium between TE proliferation and purifying selection against TEs, thus allowing TEs to accumulate (Nuzhdin et al. 1997).

How can we explain extensive proliferation and diversification of TEs within islands, but purifying selection against TEs in LDRs? Coalescent effective population size of a genomic region is positively correlated with its recombination frequency and thus the local efficiency of selection and mutation rate (Casacuberta and González 2013). The initial foundation of TE islands could hence be facilitated in genomic regions with low recombination frequency, providing a refugium of relaxed selection for TE insertions. Indeed, elevated rates of non-synonymous substitutions suggest relaxed selection on TE island genes. Increased frequency of DNA repair processes as a consequence of higher DNA transposition frequencies in TE islands should lead to more errors in DNA replication and double strand break repair (Shee et al. 2012) in comparison to LDRs. Large-scale mutations on the other hand, such as exon or gene duplications/deletions or gene shuffling, can directly be introduced during TE transposition (Fedoroff 2012). TE islands may frequently produce genetic novelty and eventually, by chance but despite high stochastic drift, adaptive phenotypes, corroborating the view of TEs as genetic innovators.

The list of genes affected by duplications or deletions contains a number of candidates that might be key to the divergence of the lineages. For example, differences in homologs to genes involved in larval development (e.g. *Mrjp1*) might explain body-size differences. Two other candidates, *Cobs_00487* and *Cobs_17834,* show homology to genes that are involved in pesticide resistance against Chlorpyrifos and Imidacloprid (*nAChRα6*) and Deltamethrin (*Cyp4c*) in different invertebrate species (Bergé et al. 1998; T Slotkin and Seidler 2009; Casida and Durkin 2013; Xu et al. 2013). Imidacloprid treatment of gall wasp infested *Erythrina variegate* coral trees of the Japan habitat occurred at least once the year prior to collection of the colonies in 2010 (pers. comm. S. Mikheyev). In the Brazil habitat Chlorpyrifos, Deltamethrin, and the organophosphate Monocrotophos have routinely been used over the last 10 years (pers. comm. J.H.C. Delabie).

Furthermore, several within-island genes involved in the production (FAS, Blomquist et al. 1987) and perception (ORs) of chemical cues contained deletions or duplications in one of the lineages. These results suggest that variation in FAS genes may be responsible for diverging CHC profiles in *C. obscurior* (Foley et al. 2006) while variation in OR genes affects olfactory perception. Chemosensory neurons express highly sensitive ORs (Vosshall et al. 2000), which are particularly diverse (Zhou et al. 2012) and under strong selection in ants (Kulmuni et al. 2013). Gene loss and duplication in the OR gene family has been significantly frequent (Guo and J Kim 2007) and differences are assumed to be shaped by adaptive processes in response to a species' ecological niche (CA Hill et al. 2002; Bohbot et al. 2007). Intriguingly, the diversification of OR genes is thought to be largely caused by gene duplications and interchromosomal transposition (Conceição and Aguade 2008), two mechanisms known to be by-products of TE activity. While the distinct patterns of kin recognition and aggressive behaviour in the two lineages of *C. obscurior* may in part be explained by TE-mediated variation in these genes, they also suggest lineage-specific dynamics of the interaction of phenotype and genome evolution. Reduced aggression between colonies in the JP lineage should promote gene flow by exchange of reproductives and thus increase N_e, heterozygosity, and the efficiency of sexual recombination, facilitating the spread of novel arising genotypes. Our findings contrast the view of reduced aggression between colonies of invasive ants (Tsutsui et al. 2003), but so far it is unclear whether lineage specific differences are caused by variation in perception or downstream neuronal processes.

Mechanisms controlling TEs are as old as prokaryotes (Fedoroff 2013) and in fact most TEs are epigenetically silenced (Feschotte 2008; Fedoroff 2012), through either methylation, histone modifications (Rebollo et al. 2012), or RNAi (Buchon and Vaury 2006). Even though many genes in TE islands are expressed, the overall expression is significantly lower than in LDRs. In line with previous correlations on methylation and expression in eusocial insects (Bonasio et al. 2012; BG Hunt, Glastad, et al. 2013a), o/e CpG ratios in *C. obscurior* LDR genes are negatively correlated with expression. However, TE island genes do not follow this trend, in that they are weakly expressed

while having low o/e CpG rates. Proximity to TEs can positively affect gene body methylation (Veluchamy et al. 2013), which could explain stronger methylation of TE island genes and thus CpG depletion. Also, relaxed selection in island genes should in general increase fixation frequency of base mutations, including CpG to TpG conversions thus depleting CpG content. Gene expression differences in TE island genes between larvae and adult queens suggest stronger regulation of these potentially disruptive genes during the sensitive developmental phase. Finally, key regulatory genes are underrepresented in TE islands. These gene set differences between TE islands and LDRs can either be explained by selection processes, removing vital genes from linkage to TE islands or, by selective restriction of TE accumulations to genomic regions devoid of such genes.

The current understanding of TE activity dynamics in genomes is that periods of relative dormancy are followed by bursts of activity, often induced by biotic and abiotic stress, such as exposure to novel habitats. Frequent TE transposition during bursts leads to genomic rearrangements, thus producing new genetic variants and eventually even promoting speciation (Hurst and Werren 2001; Bailey et al. 2003; Ungerer et al. 2006; de Boer et al. 2007). TE dynamics can also be strongly affected by mating system (D Charlesworth and B Charlesworth 1995; D Charlesworth and Wright 2001; Wright et al. 2008; Boutin et al. 2011), and the life history of *C. obscurior* likely challenges the genomic integrity resulting in genomic regions with over 50 % TE content. In conclusion, TE dynamics in *C. obscurior* seem to have shifted from a serial to a parallel mode, where a fraction of the genome is reshaped repeatedly in a continuous burst of TE activity. Strikingly, the inbred parasitoid wasp *N. vitripennis* has similar TE frequency patterns suggesting that similar life history strategies and their consequences on N_e and drift can lead to convergent genomic organization. TEs represent a major force in evolution, contributing to the generation of genetic variation especially in species confronted with hurdles like inbreeding or repeated bottlenecks. They furthermore seem to play an important role in the rapid adaption of invasive species to novel environments, making it particularly crucial to understand their origin, function, and regulation.

2.4 Material and Methods

Methods and any associated references are available in the Supplementary Material. All animal treatment guidelines applicable to ants under international and German law have been followed. Collecting the colonies that form the basis of the laboratory population used in this study was permitted by the Brazilian Ministry of Science and Technology (RMX 004/02). No other permits were required for this study.

Data access

Sequencing data from this study have been deposited in a NCBI bioproject (ncbi.nlm.nih.gov/bioproject/) under accession number 237579.

2.5 Acknowledgements

We like to thank John Wang for helpful discussions during the initial phase of the project and Andrés Moya for encouraging comments at an earlier stage of writing. We are most grateful to Masaki Suefuji and Jacques HC Delabie for collecting the colonies. We thank Christine Elsik and Chris Childers, who do not appear in the author list, and host Cobs1.4 on hymenopteragenome.org. We thank Axel Strittmatter (MWG) and Thomas Stempfl (KFB) for advising on sequencing and Nicole Lehmann for advise on RNA amplification. We thank Helena Lowack for PCR validation of small-scale structural variants, Erich Bornberg-Bauer who assisted with CpG o/e calculation and Robert Waterhouse for comparing the gene sets with OrthoDB. This study was supported by a DFG grant He1623/31 to JO, JG, and JH.

Chapter 3

3. Sphingolipids, Transcription Factors, and Conserved Toolkit Genes: Developmental Plasticity in the Ant *Cardiocondyla obscurior*

Lukas Schrader[1,*], Daniel F. Simola[2], Jürgen Heinze[1], Jan Oettler[1]

[1] Institut für Zoologie, Universität Regensburg, 93053 Regensburg, Germany

[2] Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, PA 19119, USA

* Corresponding author:
Lukas Schrader
Institut für Zoologie,
Universität Regensburg,
Universitätsstr. 31,
93053 Regensburg, Germany Tel +49 9419432996, Fax +499419433304
lukas.schrader@ur.com

Running title: Transcription factors and sphingolipid metabolism underlie developmental plasticity in *Cardiocondyla obscurior*

Keywords: Developmental plasticity, *Cardiocondyla obscurior*, RNAseq, caste determination, sphingolipid, transcription factor binding site

Abstract

Developmental plasticity allows for the remarkable morphological specialization of individuals into castes in eusocial species of Hymenoptera. Developmental trajectories that lead to alternative caste fates are typically determined by specific environmental stimuli that induce larvae to express and maintain distinct gene expression patterns. While most eusocial species express two castes, queens and workers, the ant *Cardiocondyla obscurior* expresses diphenic females and males; this provides a unique system with four discrete phenotypes to study the genomic basis of developmental plasticity in ants. We sequenced and analysed the transcriptomes of 28 individual *C. obscurior* larvae of known developmental trajectory, providing the first in-depth analysis of gene expression in eusocial insect larvae. Clustering and transcription factor binding site analyses revealed that different transcription factors and functionally distinct sets of genes are recruited during larval development to induce the four alternative trajectories. In particular, we found complex patterns of gene regulation pertaining to sphingolipid metabolism, a conserved molecular pathway involved in development, obesity and aging.

3.1 Introduction

Developmental plasticity is a core biological phenomenon through which a genotype can produce a variety of phenotypes in response to different environmental cues (West-Eberhard 2003; SF Gilbert and Epel 2009). This responsiveness requires a sensitivity to (external) environmental input and affects most, if not all, complex developmental processes (SF Gilbert 2005). Furthermore, in a fluctuating environment, developmental plasticity can induce phenotypic variation of different magnitude that can be either non-adaptive (e.g. teratogenesis, Hamdoun and Epel 2007) or adaptive (e.g. predator-induced defences in *Daphnia* (Boersma et al. 1998)) (West-Eberhard 1989; DeWitt et al. 1998; SF Gilbert 2001; Ghalambor et al. 2007; Beldade et al. 2011). One specific form of developmental plasticity called polyphenism (or environmentally induced polymorphism) allows for the generation of two or more distinct phenotypes and underlies the evolutionary success of eusocial insects (ants, bees, wasps, termites). Eusocial insect colonies are founded on the principle of division of labour (Hölldobler and Wilson 1990), where morphologically and behaviourally specialized individuals are produced from the same genotype due to gene by environment interactions. In this way, the ontogeny of a superorganism (i.e. a eusocial insect colony) parallels the differentiation of cells or organs in a multicellular organism (Bourke 2011). Such polyphenism is a prime example of the flexibility of an organism's genetic makeup (Schlichting and Pigliucci 1998; West-Eberhard 2003).

The stable differentiation of individual eusocial insect larvae into distinct adult castes is known to be affected by external stimuli (e.g. temperature, pheromones, nutrition, and tactile interactions (Hölldobler and Wilson 2009; Penick and Liebig 2012)), but we are only beginning to understand how such stimuli are transduced into neuronal, endocrine and metabolic signals that ultimately regulate gene expression to control caste fate. This molecular cascade is best resolved for the honeybee *Apis mellifera*, where development of the queen caste is induced by feeding larvae with "royal jelly", a special glandular secretion produced by worker bees. Components of this secretion, including Royalactin, Major Royal Jelly proteins, and a fatty acid Histone Deacetylase inhibitor, likely activate epidermal growth factor receptor (EGFR) signalling in the larval fat body (Kamakura 2011), which in turn activates downstream pathways that directly or indirectly affect developmental trajectories (e.g. Insulin-like signalling, PI3K/TOR/S6K, Ras/Raf/MAPK (Patel et al. 2007; Wolschin et al. 2011; Kamakura 2011; Mutti et al. 2011; Badisco et al. 2013)). Experimental gene knock-downs, as well as comparative transcriptome and methylome studies, performed mostly in honeybees, have revealed extensive caste-dependent changes in gene expression and epigenetic regulation (Kucharski et al. 2008; Elango et al. 2009; Bonasio et al. 2012; Shi et al. 2012; Simola, Ye, et al. 2013b) for a wide range of genes, e.g. hexamerins (Hoffman and Goodisman 2007; JH Hunt et al. 2007; Cameron et al. 2013), cytochrome P450s, spliceosomal genes (Cameron et al. 2013), the hypoxia pathway (Azevedo

et al. 2011), and several other metabolic and developmental genes (Evans and D E Wheeler 2001; Barchuk et al. 2007; Hoffman and Goodisman 2007; J Li et al. 2010; Begna et al. 2011).

The mechanisms regulating caste polyphenism in the evolutionarily independent lineage of ants are expected to be more sophisticated, owing to generally greater diphenism between queens and workers and the presence of extensive worker subcaste differentiation or queen polyphenism in some species (Hölldobler and Wilson 1990; Ruppell and Heinze 1999; Johnson and Linksvayer 2010). Furthermore, caste diphenism may be accompanied by striking morphological variation among males (Kugler 1983; Yamauchi et al. 1991; Heinze and Hölldobler 1993; Oettler et al. 2010).

Though transcriptomes (Feldmeyer et al. 2013), methylomes (Bonasio et al. 2010; 2012), and chromatin structure (Simola, Ye, et al. 2013b) have been compared among different adult castes of ants, detailed studies about the molecular regulation of gene expression during larval ontogeny have not yet been conducted. This is partly due to an inherent difficulty in most ant species, as an individual's caste-fate remains elusive until a late larval instar or the pupal stage. Yet, it is hypothesized that many of the key differences between adult castes emerge during larval development. Therefore, characterizing genes that are involved in this initial phase of caste divergence will help to unravel the ontogenetic and evolutionary underpinnings of caste determination and differentiation in eusocial insects.

Here we used *Cardiocondyla obscurior*, a minute myrmicine ant, as a model to study genome-wide gene expression variation among larvae of known caste fate within a replicated experimental design based on individual whole larval samples. Colonies of *C. obscurior* can be reared easily in the laboratory, and in this controlled setting we are able to induce larval development along the desired caste-fate trajectory in a reproducible manner.

Furthermore, *C. obscurior* offers a particularly promising system to unravel the mechanisms underlying developmental plasticity in both female and male sexes. In addition to female queens (QU) and workers (WO), *C. obscurior* colonies regularly express wingless males (ergatoid (literally "workerlike") males, EM) and occasionally also winged disperser males under laboratory conditions (WM) (Figure 3.1, Table 3.1) (Heinze and Hölldobler 1993; Cremer and Heinze 2003; Schrempf and Heinze 2006; Cremer et al. 2010). The phenotypes of these two male forms differ substantially in many aspects (Table 3.1). Winged males resemble typical males of other ant species in morphology, physiology and behaviour, whereas ergatoid males have smaller eyes, a slender thorax, smaller brains, an unusually long life span and – unique for males of social Hymenoptera – a life-long spermatogenesis (Kugler 1983; Heinze and Hölldobler 1993). EMs also feature enlarged, sickle-shaped mandibles that are deployed in lethal fights with rival males over reproductive rights (Stuart et al. 1987; Kinomura and Yamauchi 1987).

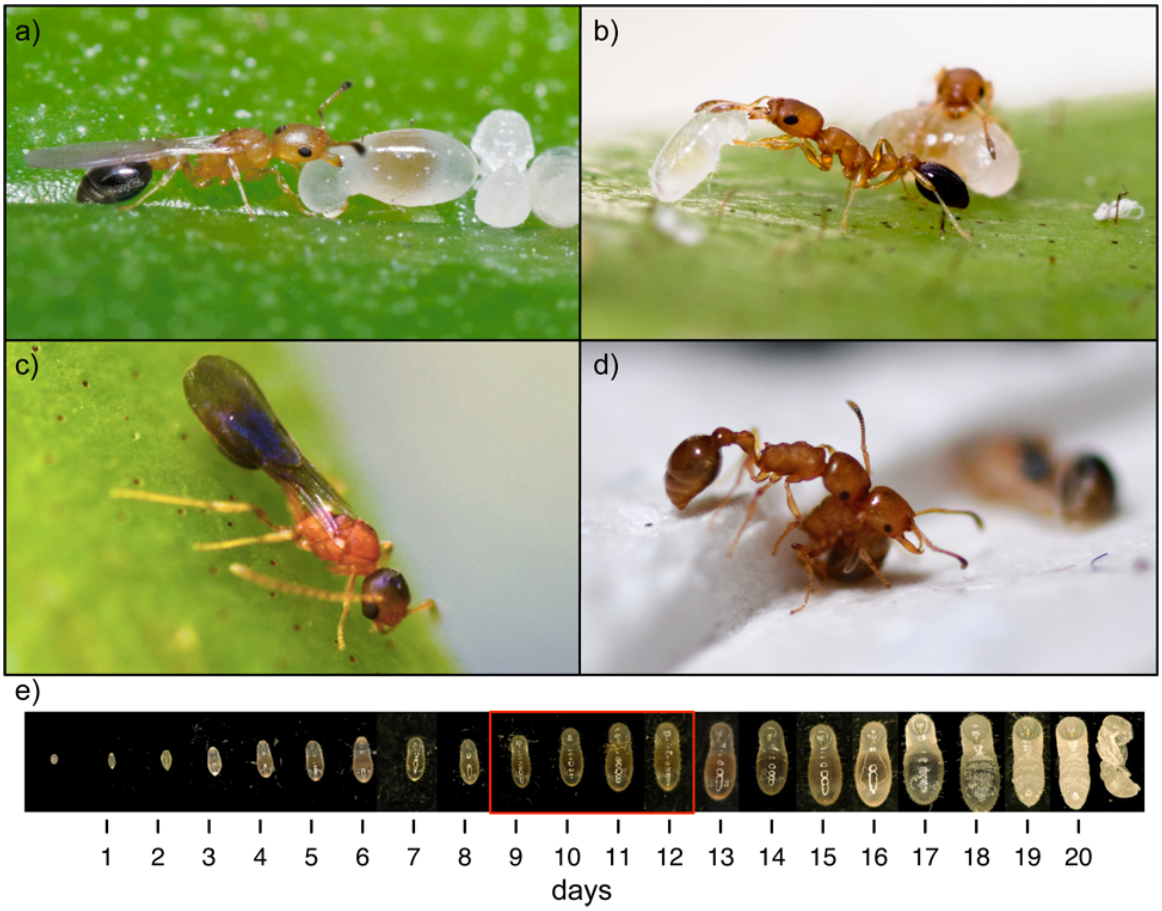


Figure 3.1: Phenotypes of *Cardiocondyla obscurior*. Fertilized, diploid eggs develop either into a) queens (QU) or b) workers (WO), while unfertilized eggs remain haploid and develop into males of either the c) winged (WM) or d) wingless, ergatoid phenotype (EM). e) Development of a *C. obscurior* worker larva from egg to pupa. The red box indicates the approximate sampling period for RNAseq experiments.

Table 3.1: General features of the four different adult phenotypes expressed in *Cardiocondyla obscurior*. QU= Queen, WO= Workers WM= Winged male, EM= Ergatoid male.

	QU	WO	WM	EM
Sex	female	female	male	male
Winged	yes	no	yes	no
Average life span (weeks)	24	8	2	6
Average size (mm)	~2	~1.7	~2.1	~1.7
Fertility	high	sterile	low	high[1]
Individuals per colony	5	28	0[2]	1
Number of ocelli	3	0	3	0
Antennae	elbowed, long scape	elbowed, long scape	short scape	elbowed, long scape
Gaster pigmentation	dark	dark	dark	light
Eye size	medium	medium	large	small
Mandible type	normal	normal	reduced	enlarged

1: Life long spermatogenesis
2: Winged males are only produced under lab conditions

To deepen our understanding of the mechanisms underlying polyphenism in ants, we characterized the developmental transcriptomes of *C. obscurior* larvae by high throughput mRNA sequencing (RNAseq) using seven early third instar larvae from each of four known developmental trajectories (caste fate), rendering this study the first analysis of genome-wide gene expression differences of individual ant larvae. By analysing gene expression patterns based on caste fate, we identified the genes that are putatively functional during the early stages of caste divergence.

Transcriptional regulation of gene expression is achieved through many factors, including transcription factors (TFs), DNA methylation, histone post-translational modifications and histone variants, and noncoding RNAs. Among these factors, TFs are perhaps the most well studied and are established regulators of development and phenotypic plasticity in eukaryotes (Carroll 2000; SF Gilbert and Epel 2009; Latchman 2010). By binding to canonical and often evolutionarily conserved binding sites (TFBSs) typically located in intergenic sequence proximal to genes (promoters) and in tandem groups of TFBSs distal to genes (enhancers), TFs can coordinately regulate transcription for groups of target genes, thereby driving specific developmental pathways. It was previously shown that there are significant evolutionary changes in the genome-wide TFBS landscape between solitary and eusocial insect genomes, suggesting that the regulation of gene transcription by TFs is a major mechanism facilitating caste differentiation in eusocial insects (Simola, Wissler, et al. 2013a). In support of this, we found enrichment of different TFBSs in promotor regions of genes exhibiting differential expression by caste fate, thereby relating specific TFs to early developmental divergence of ant castes.

Sphingolipids are versatile structural compounds of cell membranes and have gained increasing interest in cancer research and developmental biology for their activity as second messengers affecting highly conserved molecular pathways, thus regulating cell growth, proliferation, differentiation, and apoptosis (Spiegel and Merrill 1996; Adachi-Yamada et al. 1999; Lebman and Spiegel 2008; Gault et al. 2010; Hamel et al. 2010; Kraut 2011; Zhu et al. 2013; Pepperl et al. 2013; Sasamura et al. 2013). Intriguingly, the sphingolipid metabolism is tightly interwoven with several pathways implicated in caste differentiation (TOR (Zhu et al. 2013), Wnt (Pepperl et al. 2013), Notch/EGF (Baker and Thummel 2007) and is also involved in the regulation of CytP450 expression (Merrill et al. 1999) and JH metabolism (Q Yang et al. 2010). These interactions led us to analyse expression patterns for genes involved in sphingolipid metabolism. We found differential regulation in all expressed key genes (Walls et al. 2013), indicating that developmental plasticity in ants is in part regulated by changes in sphingolipid homeostasis.

Together, our findings confirm that significant gene expression differences emerge early in larval development, priming larvae for the vast morphological reorganization during metamorphosis as pupae. Furthermore, such changes likely are controlled by a specific set of TFs and prominent in

developmental toolkit genes (e.g. Wnt, EGFR, Notch), metabolic pathways (e.g. sphingolipids, oxidative phosphorylation), and cell-cycle processes (e.g. ribosomal and proteasomal proteins).

3.2 Results

To assess gene expression differences at an early stage of caste fate divergence, we performed high throughput sequencing of mRNA complements (RNAseq) from 28 individual and morphologically indistinguishable early third instar larvae, sampled shortly after their commitment to each of the four mature phenotypes (Figure 3.1e) (Schrempf and Heinze 2006). By sequencing seven independent biological replicates per treatment, our experimental design explicitly incorporates inter-individual biological variation (Auer and Doerge 2010) (Figure 3.2a-c), which increases the sensitivity and specificity of our analyses (Y Liu et al. 2014b).

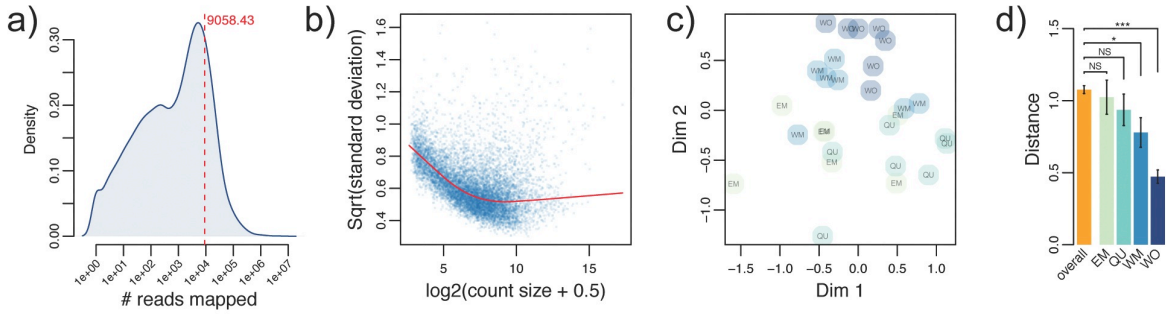


Figure 3.2: a) Distribution of raw read coverage per gene. The red dashed line indicates mean read count per gene. b) Gene-wise mean-variance relationship suggesting high levels of biological variation between samples (Law et al. 2014). c) Multi dimensional scaling (MDS) plot for the top 500 most variable genes. d) Mean Euclidean distances (±SE) of MDS-scaled samples between all samples (overall) and within each caste fate.

Differential gene expression and Gene set enrichment analyses

The current official gene set of *C. obscurior* includes annotations for 17,552 protein-coding genes (Schrader et al. 2014). In our larval RNAseq data 10,012 of these genes (57 %) were expressed with a mean raw coverage of 9058.43 reads per gene (Figure 3.2a) and high-level biological variation as suggested by the mean-variance relationship (Figure 3.2b, Law et al. 2014). To assess whether transcriptome samples cluster according to their developmental trajectory, we performed a multi-dimensional scaling (MDS) analysis with limma (Law et al. 2014) of the 500 most variable genes across all samples (Figure 3.2c) and calculated mean Euclidean distances between MDS-scaled data points (Figure 3.2d). This calculation of distance is conservative, as it relies on a lower dimensional projection of the data that does not account for the total variance. Compared to mean distances between pairs of all 28 samples, distances between EM and QU samples, respectively, did not differ significantly (t-tests, EM: df= 22.2, p= 0.67, QU: df= 22.58, p= 0.23). However, among WM and WO

samples, respectively, distances were significantly smaller (t-test, WM: df= 22.95, p= 0.01, WO: df= 36.18, p< 0.0001). In addition, we found strong overlap between groups in the MDS analysis, in particular for QU and EM.

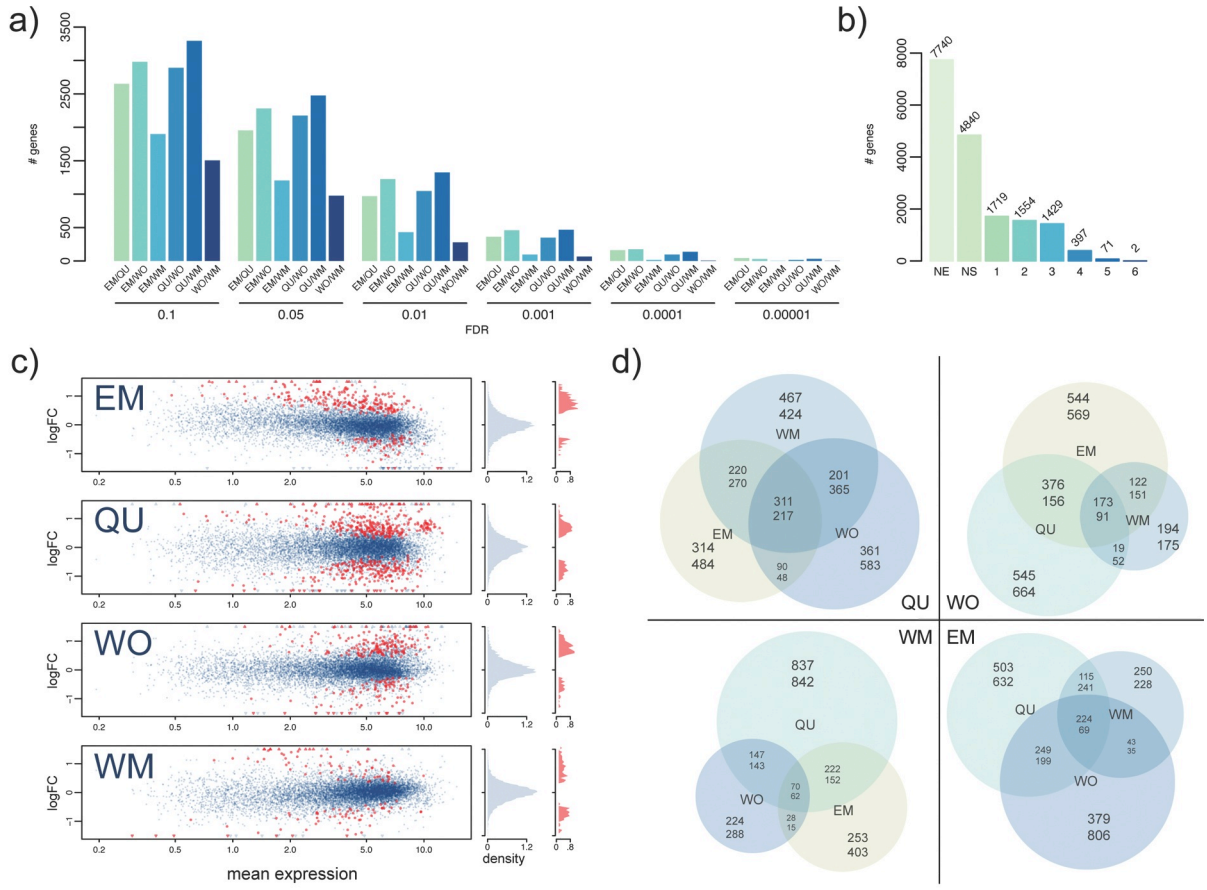


Figure 3.3: a) Number of differentially expressed genes for the six possible pairwise comparisons at different FDR cut-offs. b) Summary of differential gene expression calls with six pairwise comparisons at an FDR of 0.05, NE= not expressed, NS= no significant difference in expression, 1-6: Number of genes with significant expression differences in 1 to 6 pairwise comparisons. c) MA plots showing the mean expression across all 28 samples on the x-axis and the log-fold change for one over three other groups. QU= queen, WO= worker, EM= ergatoid male, WM= winged male. Each dot represents one expressed gene, red dots show genes included in the compiled gene sets. Density plots show frequency distribution of all genes (blue) and of genes in the set genes (red) d) Venn diagrams comparing phenotype-specific expression differences in winged males (WM), ergatoid males (EM), queens (QU), and workers (WO). Each Venn diagram shows the number of genes with significantly higher (upper numbers) or lower expression (lower numbers) at an FDR < 0.05 in one, two or three pairwise comparisons per caste fate. For each caste fate, sets of genes differentially expressed in all three pairwise comparisons (i.e. the central intersection) were used for enrichment analyses.

Using a false discovery rate (FDR) of 0.05 (Figure 3.3a), 5,172 genes showed differential expression in at least one of the six pairwise comparisons (EM/QU, EM/WO, EM/WM, QU/WO, QU/WM, WO/WM, Figure 3.3b). Since we were interested in phenotype-specific gene expression profiles, we retrieved the subsets of differentially expressed genes (at FDR < 0.05) that showed consistent changes in one group compared to the three others (Figure 3.3c, Supplementary Tables S2.1-S2.4), e.g. 311

genes showed significantly higher and 217 genes showed significantly lower expression in QU in each of the three pairwise comparisons with WO, EM and WM (Figure 3.3d).

In the different gene sets, we assigned *Drosophila* orthologs to 69.70 % (QU), 74.24 % (WO), 62.46 % (EM) and 53.79 % (WM) of the genes using a reciprocal best alignment approach. When orthology was not resolved clearly, we assigned the closest homolog (BLASTp e-value < 1e-10). We then performed functional enrichment analyses using Gene Ontology (GO) terms (see Supplementary Material) and DAVID's Functional Annotation Clustering (v6.7), which combines enrichment scores from different resources (e.g. GO terms, protein-protein interactions, homologies) based on common biological themes. DAVID analysis returned 25 distinct Annotation Clusters on the 152 EM-specific genes and fewer clusters for the other gene sets (QU 12 (199 genes), WO 5 (53), WM 0 (0)) (Table 3.2). In QU, most clusters are associated with cell-cycle processes (e.g. mitosis, DNA replication, nucleosome organization, Table 3.2a). Furthermore, we found clusters associated with transcript splicing and diacylglycerol kinase activity. The functionally clustered gene set in QU contains twelve stronger expressed RpS, 14 RpL (13 overexpressed), and four mitochondrial ribosomal protein genes, which are significantly downregulated in QU. Furthermore, seven different histone genes (four H2A, one H1/H5, one H2b, and one H3) and the proteasome subunits α3, -6, -7 as well as β2, -4, -6 and -7 are expressed to a significantly lower level in the QU gene set (Supplementary Table S2.1). As histone occupancy generally inhibits transcription factor binding to DNA (Boyle et al. 2008; J Wang et al. 2012; Neph et al. 2012), these results are consistent with a model of caste differentiation in which queen destined larvae increase chromatin accessibility to increase gene expression and cell proliferation.

Table 3.2a: Functional enrichment of genes significantly differentially expressed in queens (QU) (annotations by DAVID, AC= Annotation Cluster, ES= $-\log(p_{enrichment})$).

Cluster	ES	Description
AC 1	14.85	Mitosis 1
AC 2	10.29	Mitosis 2
AC 3	6.20	DNA replication
AC 4	3.95	Protein-DNA interaction
AC 5	3.76	Nucleotide binding
AC 6	3.65	Proteasome
AC 7	3.38	Mitosis 3
AC 8	3.15	Nucleosome organization
AC 9	2.83	Pre-replicative complex
AC 10	2.71	Spindle organization
AC 11	2.71	Splicing
AC 12	2.56	Kinase activity

Table 3.2b: Functional enrichment of genes significantly differentially expressed in workers (WO) (annotations by DAVID, AC= Annotation Cluster, ES= -log($p_{enrichment}$)).

Cluster	ES	Description
AC 1	11.34	Oxidative phosphorylation
AC 2	6.69	Energy generation
AC 3	4.01	ATP metabolism
AC 4	2.86	Oxidoreductase activity
AC 5	2.46	Valine, leucine and isoleucine degradation

Table 3.2c: Functional enrichment of genes significantly differentially expressed in ergatoid males (EM) (annotations by DAVID, AC= Annotation Cluster, ES= -log($p_{enrichment}$)).

Cluster	ES	Description
AC 1	11.70	Neurogenesis
AC 2	10.12	Organogenesis
AC 3	8.51	Cellular morphogenesis
AC 4	7.47	Respiratory system development
AC 5	6.61	Transcription factor activity
AC 6	5.83	Imaginal disc differentiation
AC 7	5.54	Exocrine System development
AC 8	5.46	Limb & Wing morphogenesis
AC 9	5.06	Organismal development & organogenesis
AC 10	4.87	Axon guidance
AC 11	4.68	Wnt pathway
AC 12	4.67	Eye morphogenesis
AC 13	4.64	Cell adhesion involved in morphogenesis
AC 14	4.47	Homeobox
AC 15	4.00	Cell-cell junction
AC 16	3.80	EGF signalling
AC 17	3.60	Gliogenesis
AC 18	3.28	Compound eye pigmentation
AC 19	3.02	Genitalia development
AC 20	2.92	Vesicle formation
AC 21	2.84	Immunoglobulin-like proteins
AC 22	2.64	Frizzled
AC 23	2.52	Metal ion binding
AC 24	2.42	Gene expression
AC 25	2.19	Apoptosis

In WO, each of the five Annotation Clusters is associated with energy generation in mitochondria (e.g. oxidative phosphorylation, ATP metabolism, or valine, leucine and isoleucine degradation, Table 3.2b). We found stronger expression of five mRp genes (L20, L21, L32, S14, S28) and three proteasome genes (α6, β3, β7). In addition, four different Cytochrome oxidase C subunits (IV, Vb, VIb, VIIc) were expressed significantly higher in WO (Supplementary Table S2.2).

Many of the 20 functional clusters in EM are associated with basic development (e.g. neurogenesis, imaginal disc development, exocrine system development, Table 3.2c). In addition, we found clusters indicating substantial changes in core molecular pathways (e.g. Wnt, EGFR) (Hurlbut et al. 2007). For example, six of eight annotated *Wnt* homologs (*wg*, *Wnt10*, 3 homologs of *Wnt2*, *Wnt5*, and a homolog of *Wnt6*) in the *C. obscurior* genome are significantly overexpressed in EM. In addition, orthologs of the Wnt receptors *fz2* and *fz4*, the co-receptor *arr*, and associated genes like *stan*, *arm*, *nkd*, and four different Cadherins (*ft*, *Cad87A*, *Cad74A*, and a homolog to *Cad96Ca*) are upregulated. Apart from Wnt pathway genes, we found differential regulation in various genes involved in Notch (*N*, *Dl*, *H*, *Ser*, *shg*, *heph*) or EGFR signalling (*Egfr*, *Hh*, *Ptc*) and also in three Toll-receptor homologs (*Toll-6* and two *Tollo* homologs). In addition, differential expression in EM was abundant in transcription factors, e.g. *abd-b*, *castor*, *twist*, *Sp1*, *vielfältig*, *jumeau*, *jim*, and homeobox genes (e.g. *dll*, *en*, *smo*) (see Supplementary Table S2.3).

There were no significant Annotation Clusters for WM. However, some individual candidate genes might be of particular interest in the WM specific gene set. For instance a homolog to *Edem1*, a gene coding for an alpha-mannosidase affecting life span in *D. melanogaster* and *Caenorhabditis elegans* (Y-L Liu et al. 2009), is strongly downregulated in WM. In addition, we found differential expression in well-characterized developmental genes such as *headcase*, *visgun*, *eyegone*, as well as homologs to *engrailed*, *prospero*, *AGO3*, and *capricious* (Supplementary Table S2.4).

Together, the enrichment analyses revealed strong functional differences between the gene sets, suggesting that development of the discrete phenotypes is regulated by distinct developmental programs in cell division (QU), energy generation (WO), and developmental toolkit genes (EM).

Transcription factor binding site enrichment

As TFs function as important regulators of developmental plasticity, we tested whether gene expression may be regulated by distinct sets of TFs for each caste fate (Figure 3.4). Significant binding sites (TFBSs) for 59 factors were predicted by scanning 2 kb promoter sequences using available position weight matrix models. Caste-biased gene sets were subsequently analysed for enrichment of each TFBS model individually using two measures (enrichment probability pE and presence probability pP). In the QU gene set, 13 TFBSs were either significantly enriched ($pP < 0.01$, Bonferroni corrected) or overrepresented ($pP < 0.01$, Bonferroni corrected), with five TFBSs being significant for both measures. Of eight significant TFBSs in the WO gene set, six were significant in both tests. The highest number of significantly enriched TFBSs was found in EM, with 19 TFBSs being significantly enriched ($pE < 0.01$, Bonferroni corrected) or overrepresented ($pP < 0.01$, Bonferroni corrected). Of these, ten TFBSs were significant in both tests. In contrast, we did not find

any TFBSs to be significantly overrepresented in the WM gene set and only one TFBS (OVO) was significantly enriched (pE < 0.01, Bonferroni corrected).

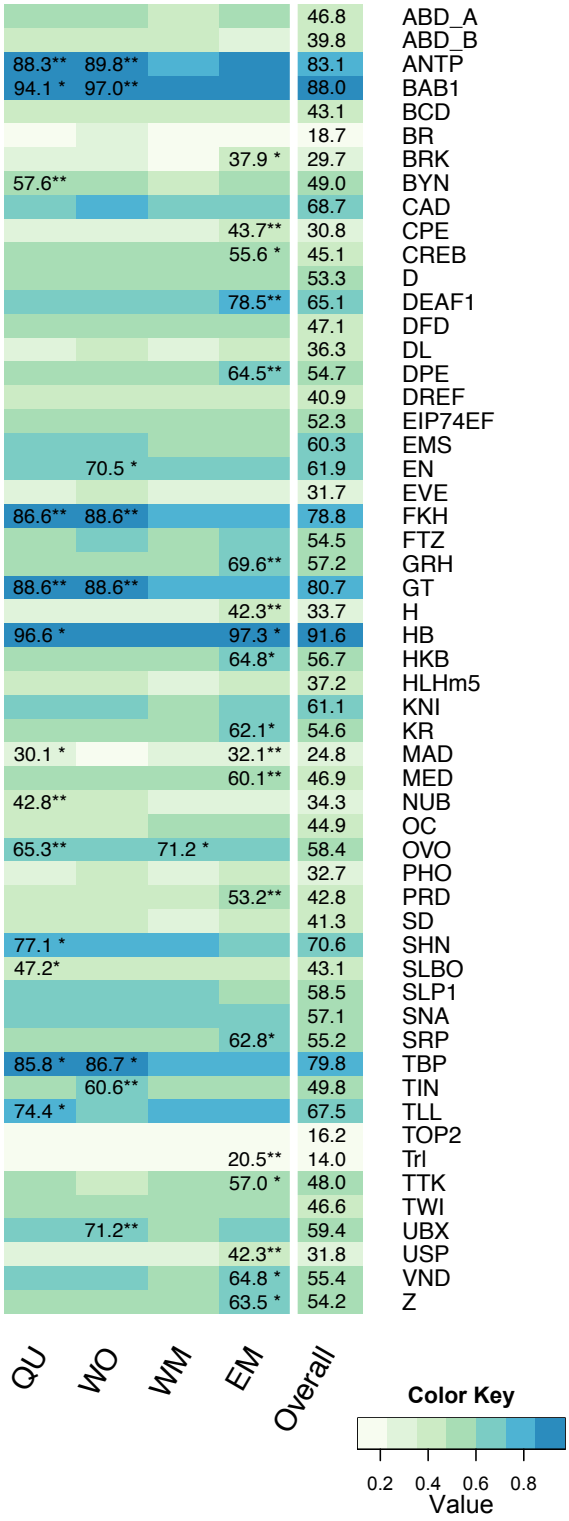


Figure 3.4: Transcription factor binding site enrichment for caste specific gene sets. The heat map shows the proportion of genes in the different gene sets that have at least one of the respective TFBSs in their promoter sequence. Numbers for significantly enriched/overrepresented TFBSs give the percentage of genes having the respective TFBS in their promotor. Asterisks indicate whether we found significant enrichment (pE<0.01, asterisk at first position), overrepresentation (pP<0.01, asterisk at second position), or both (two asterisks).

Sphingolipids

Based on previous evidence suggesting that sphingolipid metabolism regulates pathways involved in caste differentiation, we analysed expression of sphingomyelin cycle genes in our dataset. Interestingly, we found significant expression differences in all eight key genes (Walls et al. 2013) involved in the metabolic cycle (Figure 3.5, Supplementary Table S2.5). *Lace*, which codes for a serine-palmitoyltransferase responsible for *de novo* synthesis of sphingolipids, was weaker expressed in WO. Following the formation of dihydrosphingosine, the dihydroceramide synthase *schlank* (significant differences in expression: WM>WO, WM>QU, WM>EM, QU>WO) and the sphingosine desaturase *infertile crescent* (*ifc*) (QU>EM) catalyse the production of the bioactive lipid ceramide. The conversion of ceramide to sphingosine involves different ceramidases coded by *brain washing* (*bwa*, with two homologs in *C. obscurior*. *Bwa1*: not expressed, *bwa2*: QU>WM), and *ceramidase* (*CDase*), which has two homologs in *C. obscurior* of which both show stronger expression in QU (*CDase1*: QU>EM, QU>WM and *CDase2*: QU>EM, QU>WM, QU>WO). Sphingosine is then converted to sphingosine-1-phosphate (S-1P) through the activity of a kinase coded by *Sk2* (QU>EM, QU>WM, QU>WO). The membrane-bound S1P-lyase *Sply,* which is down regulated in EM, converts S-1P to an aldehyde and phosphoethanolamine, thus removing S-1P irreversibly from the cycle.

In addition to the eight key genes of the sphingomyelin cycle, we also analysed expression of downstream genes affected by cellular sphingolipid levels. For example, we found differences in *slipper* (QU>WO), *hemipterous* (QU>WO, QU>WM) and *basket* (EM>WO, WM>WO), three genes that form a ceramide-sensitive MAP k^3 → JNK pathway involved in stress response and apoptosis in *Drosophila* (Kraut 2011).

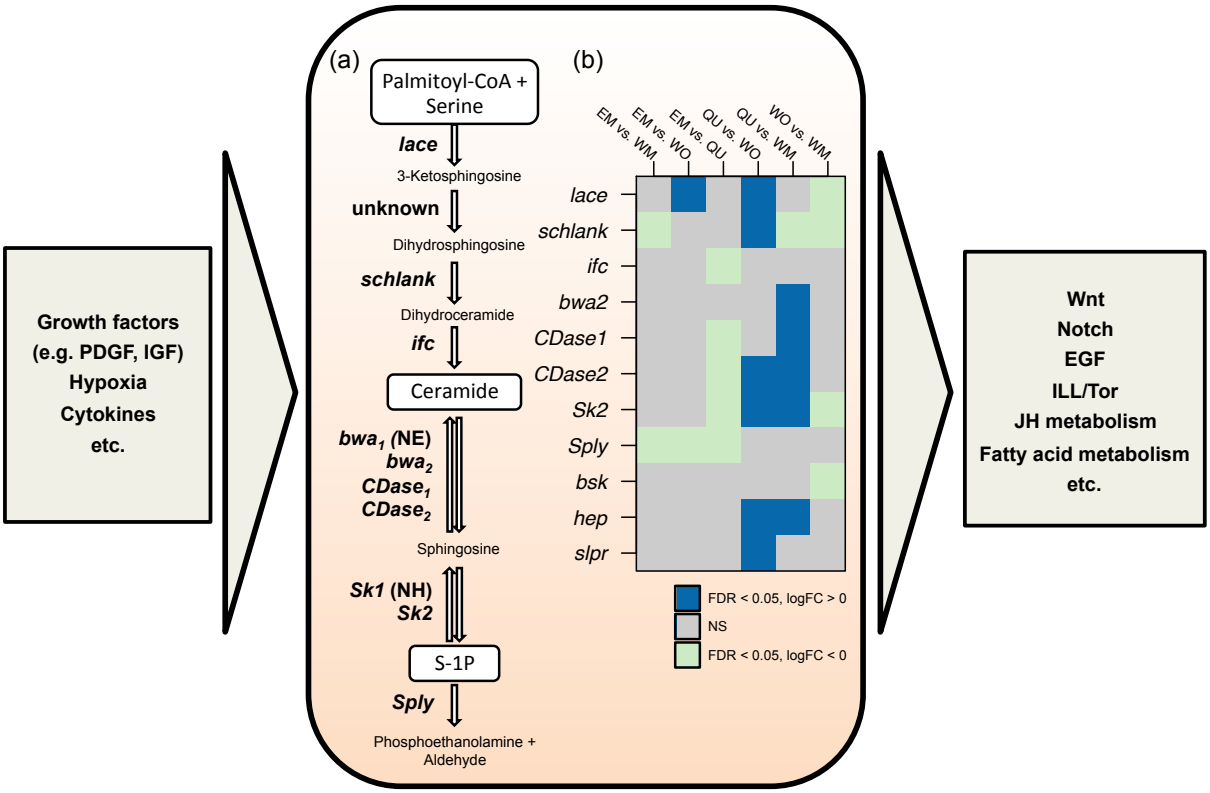


Figure 3.5: Schematic representation of the sphingomyelin cycle (central box) and its involvement with other developmental pathways and mechanisms (left and right boxes). (a) The central metabolism in sphingolipid biosynthesis (Walls et al. 2013) with the involved genes (left of arrows) and metabolic products (between arrows) (b) Differential expression of key genes of the sphingomyelin cycle. The heat map shows significant (FDR < 0.05) positive (dark blue) or negative (light blue) log fold changes for each of the six pairwise comparisons in sphingomyelin cycle and downstream genes (*basket* (*bsk*), *hemipterous* (*hep*), *slipper* (*slpr*)). *CDase2* and *bwa2* are homologs to the *Drosophila CDase* and *bwa* genes, respectively. NH: no homolog identified in *C. obscurior*. NE: not expressed. NS: no significant difference.

3.3 Discussion

The pronounced differences between queen and worker castes make ants and other social insects compelling models to study the regulation of developmental plasticity. However, genetic and molecular studies focusing on the divergence during larval and pupal development are still rare (Abouheif and Wray 2002; Ometto et al. 2011; X Chen et al. 2012; Berens et al. 2015). Studies on queen determination in different ant species have revealed differences in developmental switch points in caste differentiation during larval development, suggesting that the extent of queen-worker-diphenism correlates with the timing of developmental divergence (Abouheif and Wray 2002; Penick et al. 2012). Furthermore, as shown in *A. mellifera*, caste determination might be a progressive process not controlled by a single fixed time point in development (Cameron et al. 2013). In the polyphenic ant *C. obscurior*, queen/worker and winged/ergatoid male determination has been studied in cross-fostering experiments and by juvenile hormone analogue application (Schrempf and Heinze 2006; Du et al. 2007), indicating that in both males and females, caste fate is determined late in second larval instar development. In this study, we characterized gene expression during larval development shortly after the mature phenotype has been determined with high likelihood to unravel the genetic framework directing specific morphological differentiation into four distinct adult phenotypes.

Gene expression patterns

The great number of genes found to be differentially expressed in our comparison shows that distinct phenotypes adopt largely different gene expression profiles already during larval development, preceding morphological differentiation that occurs during metamorphosis. Therefore, the specific morphologies of queens, workers, ergatoid males, and winged males likely depend on gene expression patterns that are pre-established, as opposed to emerging first during pupation.

Furthermore, despite the apparent morphological similarities between alate and wingless phenotypes between sexes (e.g. female workers and ergatoid males), we did not find consistent overlap in gene expression profiles (Figure 3.3d) or TFBS enrichment (Figure 3.5) comparing these phenotypes. Assessing whether analogous morphological traits (e.g. winglessness) do in fact emerge by separate developmental routes however requires detailed studies on the vestigial imaginal discs of EM and WO.

Though an RNAseq experiment with whole body samples does not allow for a deeper analysis of tissue-specific expression profiles, it did allow for the identification of systemic gene expression changes. Furthermore, by sampling during early larval development when morphological differences have not yet emerged, we are able to minimize allometric bias in cell number or proportion of tissues.

Functional enrichment analysis revealed substantial ontogenetic differences between EMs and the other three phenotypes, driven by changes in pleiotropic and interconnected signalling networks (Notch, EGFR, Wnt, etc.). The general importance of Wnts in development and morphogenesis in arthropods is well known (e.g. in CNS, eye, genitalia, imaginal discs, mouthparts (Murat et al. 2010)) and it is also likely to be involved in wing loss in ants (Abouheif and Wray 2002; Rajakumar et al. 2012). The great number of differentially expressed Wnt, Notch, and EGFR pathway genes in EM thus indicates that changes in timing and/or orchestration of gene expression in such evolutionarily conserved developmental circuits underlie the evolution of the ergatoid male phenotype in *Cardiocondyla*.

In addition, absence of developmental toolkit pathways in DAVID analyses of the three other gene sets suggests that, during the particular period of sampling, basic developmental mechanisms experience differential regulation specifically in EMs. This is likely a consequence of strong sexual selection on EM specific traits. EMs attack and kill young rivals upon their eclosion, resulting in an evolutionary arms race over production of the first male in multi-queen colonies (Yamauchi et al. 2006; Suefuji et al. 2008). Consequently, selection is predicted to favour the early emergence of EMs; accordingly, EMs do exhibit reduced developmental time compared to the other male phenotype (Schrempf and Heinze 2006). Thus, phenotype-specific expression profiles in our analysis might emerge from alternative developmental rates that are set by upstream changes in developmental signalling pathways, e.g. through Notch and Wnt signalling. In support of this, a recent study also found DNA methylation differences in Wnt genes between queen- and worker-destined honeybee larvae, indicating a prominent role for these networks in plastic development of eusocial insects (Shi et al. 2012).

The prevalence of genes involved in mitosis and nucleosome organization in QU implies specific changes in cell cycle and chromatin structure. Many mitosis-related genes are downregulated in QU, potentially reflecting a state of arrested cell division. In fact, restrained growth in queen larvae is known from honeybees, where queen-destined larvae remain smaller during early differentiation and only later outgrow their worker-destined siblings (Stabe 1930). Conversely, the high frequency of overexpressed ATP metabolic genes in the WO-specific gene set indicates an increased energy requirement in worker larvae during early 3rd instar.

This is in contrast to studies done in *A. mellifera* that suggest an increase in energy generation in queen-destined larvae (Corona et al. 1999; J Li et al. 2010; Begna et al. 2011; Cameron et al. 2013). However, in honeybees, growth rates of queen-destined larvae predominantly increase late in development and queens develop faster than workers (Page and Peng 2001). In *C. obscurior* (Schrempf and Heinze 2006) and likely other ants (Bowsher et al. 2007), queens develop at the same rate than workers but for a longer period. Thus, caste-specific energy requirements of larvae might not be conserved across these two distant taxa.

Transcription factor binding sites

Testing for enrichment of TFBSs in *cis*-regulatory regions, we again found specific differences between QU, WO, EM, and WM. The lack of a single TFBS enriched in all four phenotypes suggests that larval development is not controlled by common TFs, but instead that each developmental phenotype recruits its own set of TFs to produce the respective gene expression signatures. Yet, we found five TFBSs (antennapedia [ANTP], bric a brac 1 [BAB1], fork head [FKH], giant [GT], and TATA-binding protein [TBP]) with significant enrichment in both QU and WO, potentially revealing TFs specifically recruited to female larval development, e.g., as shown for *bab1* in *Drosophila* (Kopp et al. 2000). We did not find a similar overlap between the two male phenotypes, as WMs showed only one TFBS (OVO/Shavenbaby, a germline expressed factor shared with queens). How the interplay of different TFs directs gene expression and development in each caste remains elusive, in particular since significant overabundance of binding sites did not correlate with stronger TF gene expression in the respective larvae.

Sphingomyelin cycle

We discovered significant differences in the expression of most sphingolipid metabolism genes comparing all four phenotypes, suggesting a role for the sphingomyelin cycle in caste differentiation. In addition, previous studies revealed that mutations in sphingolipid metabolism genes in *Drosophila* and *Caenorhabditis* produce phenotypes that strikingly resemble differences observed between castes in eusocial insects, in particular concerning fecundity and longevity. For example, loss of *bwa* decreases longevity and increases larval developmental time in *Drosophila*, and pharmacological inactivation of Bwa increased Juvenile Hormone acid methyltransferase activity (Q Yang et al. 2010). Sterility was observed in *ifc* deficient *Drosophila* (Phan et al. 2007). Deletion of *sply* leads to flight muscle degradation and apoptosis in reproductive organs (Herr et al. 2003); similar defects were found in *Sk2* mutants (Herr et al. 2004). *Schlank*$^{-}$ mutant flies show defects in fat storage and larval growth, and corresponding mutants in *C. elegans* (*hyl-1; lagr-1*, *hyl-1*, and *hyl-2*) show significant lifetime extension (Tedesco et al. 2008; Menuz et al. 2009; Mosbech et al. 2013).

Similarly, the four phenotypes of *C. obscurior* are characterized by considerable differences in longevity, fecundity, larval development, and morphology (Table 3.1). The distinct differential expression in most key genes in *C. obscurior* and their established regulatory functions and roles in *Drosophila* development strongly suggest previously unappreciated role for sphingolipids in caste fate differentiation in ants. We therefore propose to extend the current model of caste determination and differentiation in eusocial insects to include sphingolipids as important mediators of developmental plasticity.

Conclusion

Eusocial insect castes illustrate the concept of developmental conversion (Smith-Gill 1983), in that plastic responses in development are not continuous reaction norms proportional to the environmental stimulus, but threshold dependent discrete phenotypes. Previous studies on caste differentiation and developmental plasticity in eusocial insects already showed that evolutionarily conserved developmental pathways are involved in this process and that the initial environmental signal is transduced to substantially alter gene expression profiles during development. In accordance, our comparison of gene expression patterns in early larvae of *C. obscurior* revealed that a large number of genes are differentially expressed between larvae of different caste-fate and that many developmental pathways and TFs appear to be involved in caste differentiation. Caste differentiation can be considered as active, regulatory plasticity involving versatile developmental and physiological changes rather than passive plasticity, a (largely non-adaptive) direct response to environmental influences (Schlichting and Pigliucci 1995; 1998; Whitman and Agrawal 2009; Forsman 2014). Eusocial insects evolved from monophenic ancestors and the worker-queen diphenism thus likely evolved from passive plasticity expressed by the solitary ancestors (JH Hunt and Amdam 2005). This transition from putatively non- or even mal-adaptive plasticity to highly adaptive plasticity is expected to occur, if the ancestral plasticity produces phenotypes that are close enough to a new phenotypic optimum for directional selection to act upon (Ghalambor et al. 2007), allowing for the evolution of a stable queen-worker diphenism as we see it today. In *Cardiocondyla*, a similar mechanism might have also enabled the evolution of male diphenism, in that the extant male phenotypes evolved from a single, merely passively plastic winged male phenotype (Tsuji et al. 1994). Intriguingly, the extent of male phenotypic plasticity is highly variable in the genus *Cardiocondyla*, with several species producing ergatoid males exclusively, others producing both morphs (Oettler et al. 2010), and some species additionally producing intermorphs between winged and wingless males (Cremer, Lautenschläger, et al. 2002a; Yamauchi et al. 2005; Heinze et al. 2013). Hence it appears that phenotypic plasticity in males evolved under species-specific selection regimes in *Cardiocondyla*, potentially rendering this genus a showcase for the role of genetic accommodation in the evolution of novelty (West-Eberhard 1989; 2005a; Pfennig et al. 2010; Moczek et al. 2011; Schlichting and Wund 2014).

3.4 Material and Methods

Organism

Cardiocondyla obscurior is a cosmotropical tramp ant species with a presumed native range in Southeast Asia (Seifert 2003). Our experimental colonies used in this study are derived from colonies that were originally collected in an introduced population in Brazil (permitted by the Brazilian Ministry of Science and Technology (RMX 004/02)) and thereafter propagated in the lab for several years.. Experimental colonies were reared at 12 h 28° C light/12 h 23° C dark in plastered Petri dishes and fed twice a week with parts of cockroaches or *Drosophila* and honey-soaked shreds of paper. Water was provided *ad libitum*. All animal treatment guidelines applicable to ants under international and German law have been followed.

Larvae sampling

Experimental colonies of *C. obscurior* were treated to exclusively produce only one of the four different phenotypes. Production of worker larvae (WO) was achieved by establishing colonies comprised of one fertilized queen, one ergatoid male and ten workers. Larvae produced within the first two months after colony foundation will develop into workers (Suefuji et al. 2008). Queen larvae (QU) were produced by treating eggs from worker-producing colonies (see above) with the juvenile hormone analogue methoprene (Schrempf and Heinze 2006). Collected eggs were rinsed twice with 1 mg/ml methoprene dissolved in acetone and transferred to a queen-less colony of 20 workers. To produce males, we reared virgin queens that only produce haploid eggs, which develop either into either ergatoid or winged males. Under standard conditions (see above), colonies produce ergatoid males. Winged males are rare. However, stressful environmental conditions can induce their production (Cremer and Heinze 2003). We established colonies consisting of two unfertilized queens and 10 workers and reared these at a constant 23° (colder than normal; see above). Some of these colonies ceased production of ergatoid males and exclusively produced winged males, especially in the later stage of their queens' life (>30 weeks, (Heinze and Schrempf 2012)), which allowed us to sample winged male larvae (WM).

We sampled larvae that had chitinized mandibles, a characteristic feature of the 3rd instar (Schrempf and Heinze 2006), and were smaller than 800 µm (Fig. 3.1e). All colonies that were used for sampling larvae were screened twice a week from two weeks before the first larva was sampled until three weeks after the last larva had been sampled to confirm the exclusive production of workers, queens, ergatoid males, or winged males. For RNAseq studies, sampled larvae were placed individually in 1.5 ml Eppendorf tubes, snap-frozen in liquid nitrogen, and stored at -80° C.

Gene expression analysis with RNAseq

Total RNA was extracted from 28 individuals during the early 3rd larval instar with the RNeasy Plus Micro kit (Qiagen), yielding 27 to 270 ng per individual larvae.

Starting from 20 ng input RNA, double strand, unstranded, multiplexed cDNA libraries for single-end sequencing of the 28 separate samples were generated. Briefly, cDNA was generated by reverse transcription using pseudo-random and oligo-dT priming. Using single-primer isothermal linear amplification (SPIA), dsDNA was generated and amplified (Ovation RNAseq system V2, NuGEN). After clean up with QIAgen's MinElute clean-up kit, 1.5 µg cDNA were fragmented to 100-300 bp by shearing (Covaris S2 AFA). Libraries from 150 ng cDNA were prepared with the Encore Rapid DR Multiplex System (NuGEN), quantified (KAPA library quantification kit), and distributed randomly across four different lanes (Auer and Doerge 2010). Sequencing was carried out at the in-house sequencing centre (KFB, University of Regensburg, Germany) on an Illumina HiSeq1000. Raw sequencing reads have been deposited in the NCBI short read archive (Accession no. SRX879674, SRX879676, SRX879678). Sequencing reads produced from QU were published previously (Schrader et al. 2014, SRX692538). Sequencing yielded ~532 M raw reads that were filtered for adapter contamination (cutadapt, Martin 2011), parsed through quality filtration (Trimmomatic v0.27, options: LEADING:10 TRAILING:10 SLIDING:4:10 MINLEN:15), and mapped against the reference genome (Schrader et al. 2014) using the tophat2 (v2.0.8) and bowtie2 (v2.1.0) package (Phan et al. 2007; Langmead and Salzberg 2012; D Kim et al. 2013, --b2-sensitive, default settings). On average a mapping rate of ~55 % was obtained. *De novo* assembly of unmapped reads with velvet (Zerbino and Birney 2008) and subsequent BLAST analyses did not yield contigs of traceable origin, suggesting the unmapped reads are of no biological significance. Indeed, similar mapping rates have been reported by other studies using the same or very similar library preparation protocols (H Chen et al. 2011; Beane et al. 2011; Leal et al. 2013; Malboeuf et al. 2013; Sun et al. 2013; Burruel et al. 2013). Gene expression analysis was performed with limma (Smyth et al. 2002), DEseq (Anders and Huber 2010), and DEseq2 (Love et al. 2014) based on count tables produced with HTseq (Q Yang et al. 2010; Anders et al. 2015) against the Cobs1.4 official gene set (Schrader et al. 2014). Even though limma was originally designed for microarray data, it outcompetes some recently published software in RNAseq analysis (Rapaport et al. 2013). In addition, it more extensively supports modelling multi-sample comparisons in the underlying generalized linear model, so that we chose limma for the analysis of gene expression differences in this particular experiment. Read counts were converted to log2 counts per million ("library size normalization"), quantile normalized and precision weighted with voom (Law et al. 2014) for subsequent modelling in limma.

Functional clustering and gene set enrichment

We performed a reciprocal BLAST analysis against *Drosophila* annotated proteins (r5.39) to retrieve the ortholog or closest homolog (BLASTp, e-value cut-off 1e-10) for each *C. obscurior* gene. Caste specific gene sets were tested for functional clustering by parsing lists of the respective *Drosophila* closest homologs/orthologs through DAVID 6.7 (Database for Annotation, Visualization and Integrated Discovery, david.abcc.ncifcrf.gov; last accessed 22.02.2015) at an EASE cut-off of 0.01 (Huang et al. 2008; 2009). Each retrieved Annotation Cluster received a descriptive name summarizing the annotation terms.

Transcription factor enrichment analysis

Transcription factor binding site (TFBS) annotations were obtained in promotor regions (0-2 kb upstream) of all 17,552 annotated protein-coding genes for 59 different transcription factors (TFs) using pwm_scan and a P-value cut-off of 2e-04 (Levy and Hannenhalli 2002; Simola, Wissler, et al. 2013a). Caste specific gene sets were then analysed for enrichment of each TF individually using two measures (enrichment probability pE and presence probability pP) to identify TFs likely to be involved in the regulation of the query gene set (Veerla et al. 2010). The enrichment probability pE was estimated by bootstrap analysis using the total number of occurrences of a given TFBS in the promotors of a given gene set. A background distribution of occurrences was generated by randomly sampling gene lists of the same length as the query list (10^5 iterations). The probability pE of obtaining the number of occurrences of a TFBS by chance in the query gene set was then calculated as the percentile of the observed number compared to the background distribution ($1/10^5 \leq pE \leq 1$). Similarly, the presence probability pP was also estimated by bootstrap by computing the proportion of genes in a gene set (query or background) that contains at least one occurrence of a given TFBS in their promoters.

3.5 Acknowledgements

We would like to thank Mary Jane West-Eberhard for her encouraging comments on the manuscript and two anonymous reviewers for providing helpful comments. This work was funded by the Deutsche Forschungsgemeinschaft (He1623/31). DFS was supported by a Howard Hughes Medical Institute Collaborative Innovation Award (HCIA) #2009005.

Chapter 4

4. Rates of Molecular Evolution Correlate with Gene Expression Bias during Larval Development in the Ant *Cardiocondyla obscurior*

Lukas Schrader[1,2], Heikki Helanterä[3], Jan Oettler[1]

[1] Institut für Zoologie, Universität Regensburg, 93053 Regensburg, Germany

[3] Centre of Excellence in Biological Interactions, Department of Biosciences, University of Helsinki, 00014 Helsinki, Finland

[2] Corresponding author:
Lukas Schrader
Institut für Zoologie,
Universität Regensburg,
Universitätsstr. 31,
93053 Regensburg, Germany Tel +49 9419432996, Fax +499419433304
Lukas.Schrader@ur.de

Running title: Evolutionary consequences of plastic gene expression

Keywords: *Cardiocondyla obscurior* / gene evolution / caste-biased genes / developmental plasticity

Abstract

Phenotypic plasticity is based on alternative gene expression in response to environmental cues. Consequently, recruitment to the expression of a plastic trait leads to changes in the selection pressures under which a gene evolves. In most cases, plastically expressed genes have higher rates of molecular evolution and current efforts attempt to dissect underlying selection regimes. Social insects are well-suited model systems to study the evolution and expression of biased genes, because of discrete phenotypes (queens and workers) that are produced from the same genetic background. Here we study the relationship between gene evolution/regulation and expression bias in the ant *Cardiocondyla obscurior* that expresses four discrete phenotypes. In addition to queens and workers, *C. obscurior* produces winged males and fighter males, rendering this species well suited for comparative analyses. By applying a high-resolution dataset of gene expression patterns from 28 individual larvae of known developmental fate, we found that strong expression bias during larval development is negatively correlated with expression level and positively correlated with inter-individual expression variation and molecular evolution of genes. Furthermore, using observed/expected (o/e) CpG ratios, we show that CpG dinucleotides are more conserved in phenotype-biased genes, suggesting higher rates of germline gene body methylation in unbiased genes. Comparing gene expression biases in two different contexts, we found that genes biased in one context are more likely to be biased in a different context as well. Together, our study constitutes a detail-rich analysis of the correlations between gene expression, sequence evolution and gene regulation in eusocial insects.

4.1 Introduction

Phenotypic plasticity plays an important role in adaptive evolution (Schlichting and Pigliucci 1998; West-Eberhard 2003). Such plasticity is driven by changes in gene expression in response to environmental cues, directing differentiation from a single genotype into discrete phenotypes. Adaptive phenotypic plasticity thus decreases the necessity for mutation based genetic adaptation. However, phenotypic plasticity is also expected to promote sequence evolution through genetic accommodation in response to different selection regimes acting on alternative phenotypes and by exposing cryptic genetic variation to selection (West-Eberhard 2003; Price et al. 2003; West-Eberhard 2005a; Schlichting 2008; Schlichting and Wund 2014). Gene expression bias towards a particular (environmentally induced) morph or sex and the rate of sequence evolution of these genes have been shown to be positively correlated (see Helanterä and Uller 2014 for a review). However disentangling evolutionary causes and consequences of this correlation is difficult, because of opposing and frequently changing selection regimes, and confounding correlates of evolutionary rates (Helanterä and Uller 2014). Furthermore, since most studies so far have used pooled samples, the extent and importance of variation among individuals within morphs is poorly understood. As a consequence, the signal to noise ratio in genome wide correlative analyses remains low, hampering the dissection of the selection regimes driving the molecular evolution of biased genes.

The propensity for a gene to initially become involved in the expression of a plastic trait is predicted to be negatively correlated with the strength of purifying selection controlling both the genotype and the unsteadiness of expression of a given gene (Helanterä and Uller 2014). Consequently, genes under relaxed selection with accelerated evolutionary rates are more likely to be recruited to phenotypic plasticity (BG Hunt et al. 2011; Leichty et al. 2012). In addition, morph-biased genes should experience relaxed pleiotropic constraints (and *vice versa* genes with multiple functions should be less likely to evolve morph-specificity), further reducing the strength of purifying selection (Mank et al. 2008; Snell-Rood et al. 2010). In accordance, relaxed selection preceding the evolution of eusociality has been suggested to be the main factor for accelerated molecular evolution of morph-biased genes in fire ants (BG Hunt et al. 2011). Similarly, higher rates of molecular evolution of biased genes in polymorphic aphids have been suggested to result mainly from relaxed purifying selection (Purandare et al. 2014). However, genes evolving under relaxed selection might become both subject to directional (e.g. Harpur et al. 2014) and purifying selection, following their recruitment to the expression of plastic traits (Helanterä and Uller 2014).

In the ant *Cardiocondyla obscurior* reproductive queens (QU) and sterile workers (WO) occur next to winged (WM) and ergatoid (EM) males, providing an unrivalled complex system within the social hymenoptera with four distinct morphs developing from the same genetic background. Even though the foundation of phenotypic differentiation in eusocial insects lies in different developmental trajectories (D E Wheeler 1986), studies on the molecular evolutionary rates of biased genes have not

been conducted in larvae. In this study, we used gene expression data from larvae of *C. obscurior* with known developmental fate (Schrader et al. 2015) to study transcriptome-wide evolutionary dynamics and correlations between biased gene expression and intra- and interspecific evolutionary rates, methylation target sites, expression level, variation across biological replicates, and expression bias between larvae and adult queens, respectively. Overall, our study contributes to our understanding of the evolutionary underpinnings and consequences of developmental plasticity in social insect castes and polymorphic species in general. In essence, we found that strong expression bias during larval development is positively correlated with the rate of molecular evolution of genes. Furthermore, we show that CpG dinucleotides are more conserved in morph-biased genes, suggesting higher rates of germline gene body methylation in unbiased genes. Our analysis of gene expression also shows that the variation of expression across seven biological replicates is significantly stronger in biased genes than unbiased genes and that expression levels are negatively correlated with expression bias. Finally, comparing gene expression biases in two different contexts (adult vs. larva and larva vs. larva), we found that genes biased in one context are more likely to be biased in a different context as well.

4.2 Material and Methods

Gene expression data generated by RNAseq from 28 individual third instar larvae of *C. obscurior* (seven QU, seven WO, seven EM and seven WM) were used for this study (Schrader et al. 2015). The unambiguous identification of genes with true phenotype-specific expression biases is generally difficult. For one, expression biases can arise if morphological differences (e.g. in organ size) are strong, so that constitutively expressed organ-specific genes appear as differentially expressed when analysing data from whole body samples. By sampling larvae that are not yet distinguishable by morphology, we reduced this confounding effect. In addition, using independent individual biological replicates substantially increases statistical power and allows for the analysis of individual-level variation in gene expression.

We used average expression levels of genes in each of the four morphs to calculate $\log_2$ fold changes (logFC) for each of six possible pairwise comparisons between the four morphs (EM/QU, EM/WO, EM/WM, QU/WO, QU/WM, WO/WM). Each gene's expression pattern is summarized in a three-dimensional vector with the following fold change measures: QU/WO, EM/WM and EM/QU. From these, each of the remaining pairwise fold changes can be inferred. To get a single measure for a gene's overall expression bias, we calculated the Euclidean length (EL) of its vector (Formula 1, Figure 4.1). As a measure of a gene's morph-specific expression bias, we calculated the EL of a vector defined by the three respective pairwise comparisons (e.g. for EM the vector's dimensions were EM/QU, EM/WO and EM/WM, Formula 2).

$$(1)\ EL = \sqrt{logFC_{EM/WM}{}^2 + logFC_{QU/WO}{}^2 + logFC_{EM/QU}{}^2}$$

$$(2)\ EL_{EM} = \sqrt{logFC_{EM/WM}{}^2 + logFC_{EM/WO}{}^2 + logFC_{EM/QU}{}^2}$$

As a measure of intra-morph variation in gene expression, we calculated the coefficient of variance for each gene across each of the seven biological replicates. To assess divergence from homologous genes, we performed BLASTx searches of all annotated protein-coding genes from the *C. obscurior* genome (Cobs1.4, Schrader et al. 2014) against protein sequences from seven ant species (*Solenopsis invicta*, *Pogonomyrmex barbatus*, *Harpegnathos saltator*, *Camponotus floridanus*, *Acromyrmex echinatior*, *Atta cephalotes*, *Linepithema humile*) downloaded from antgenomes.org. BLAST identity scores (iS%), as a primer of divergence, indicate to which extent two sequences (i.e. the *C. obscurior*

protein and its closest ant homologue) share the same amino acid residues in the BLAST alignment (i.e. a score of 100 indicates full identity).

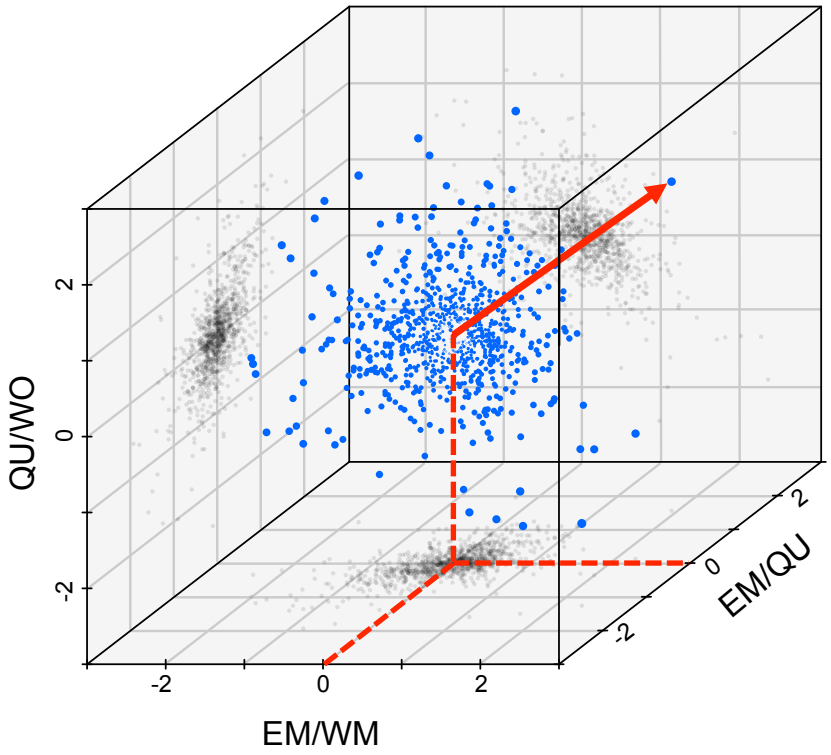


Figure 4.1: Euclidean length (EL) of a three-dimensional vector as a measure of expression bias. The three dimensional space is defined by the three independent dimensions $logFC_{EM/WM}$, $logFC_{QU/WO}$ and $logFC_{EM/QU}$. Each blue dot represents one of 1000 randomly sampled genes from the data set and the red arrow illustrates the concept of Euclidean length of three-dimensional vector. Black dots show two-dimensional projections of each three-dimensional data point.

Synonymous and non-synonymous homozygous single nucleotide variant (SNV) calls were retrieved from a genomic comparison of two populations of *C. obscurior* from Brazil and Japan (Schrader et al. 2014). From these, we calculated SNV rates per kb for each gene (dS). Intraspecific rates of non-synonymous SNVs were calculated as the number of non-synonymous SNVs divided by the sum of synonymous and non-synonymous SNVs (N/(N+S)). We retrieved CpG observed/expected (o/e) ratios (CpG_{exon} o/e) and log fold change expression values between adult queens and larvae ($|logFC_{A/L}|$) from Schrader et al. 2014. Lastly, we calculated pairwise Dn/Ds rates for orthologous gene pairs from *C. obscurior* and *P. barbatus*, which we chose because the same annotation pipeline was used. (CR Smith et al. 2011b). We identified *P. barbatus* orthologs for 9,405 *C. obscurior* genes by a reciprocal best BLAST approach. Pairwise alignments generated with clustalw2 (Larkin et al. 2007) were

trimmed to a multiple of three and up to in-frame internal stop codons in either sequence. Subsequently, we used the yn00 algorithm implemented in PAML (Z Yang 2007) to estimate pairwise Dn/Ds rates for each ortholog pair with a pairwise alignment longer than 100 bp (6,764 ortholog pairs).

To test for monotonic positive or negative correlations, genes were sorted in 500 groups based on their EL with the first group containing genes with the lowest and the last group containing genes with the strongest expression bias. We then calculated group-wise medians, inter-quartile ranges (IQR) and 1.5xIQR of average gene expression ($Expr_{Ave}$), $Expr_{Var,}$ $|logFC_{A/L}|$, dS, N/(N+S), iS%, Dn/Ds, and CpG_{exon} o/e. To assess whether evolutionary dynamics differ between genes with specific morph-biases, we compiled sets of highly biased genes for each morph. For this, we sampled all genes with a morph-specific Euclidean length above the 90 % quantile and a positive expression bias towards the respective morph. For example, the highly EM-biased gene set comprises genes with higher expression in EM compared to QU, WO and WM, respectively and a Euclidean length $EL_{EM} > 1.51$. The set of unbiased genes ("not biased", nb) is comprised of the 250 least biased genes (based on overall EL) in the dataset. We chose a cut-off of 250 to have similar sample sizes for the unbiased and the biased gene sets (see Results). For statistical analyses we calculated correlations and semi-partial correlations using Kendall's rank correlation coefficient τ and two-tailed pairwise Mann-Whitney U-tests, adjusted for multiple testing with a Bonferroni correction.

4.3 Results

Among the 17,552 genes annotated in the *C. obscurior* genome, 10,012 were expressed in larvae. The 250 least biased genes had on average an expression bias of EL= 0.11. Strong morph-bias (EL_{morph} 90 – 100 % quantile) was found in 280 (QU), 259 (WO), 247 (QU), and 217 (WM) genes, respectively. On average, genes included in the unbiased set had absolute logFC values of 0.05 across all six pairwise comparisons and genes in the biased sets on average had logFC values of 1.13 (Figure 4.2a). Euclidean length values were increased over ten-fold in the unbiased compared to the biased gene sets (EL: nb 0.17, QU 1.87, WO 2.07, WM 1.70, EM 1.79, Figure 4.2b).

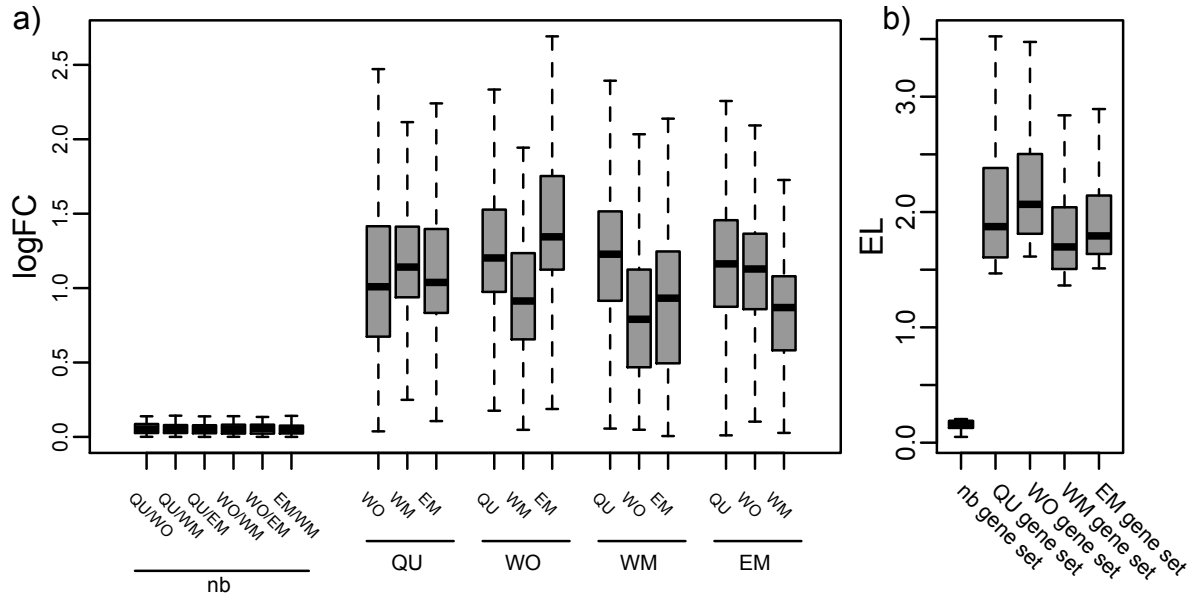


Figure 4.2: Gene expression bias in the unbiased and highly biased gene sets. (a) Boxplots show absolute logFC values for the unbiased genes in all six pairwise comparisons and logFC for each of the three pairwise comparisons used for calculating EL_{morph}. (b) Differences in EL of the unbiased and biased gene sets. Boxplots show median, inter-quartile range (IQR, box) and 1.5xIQR (whiskers).

We found a significant and strong monotonic negative correlation between gene expression bias, as measured by EL, and average gene expression levels $Expr_{Ave}$ (Figure 4.3, Table 4.1). As expression levels are known to be negatively correlated with evolutionary rates (Drummond et al. 2005), we used semi-partial correlations with Kendall's coefficients for each following correlative analysis, excluding the effect of expression levels (S-H Kim and Yi 2006). To further assess the relationship of gene expression bias and gene regulation, we tested for correlations (correcting for expression level)

between expression bias and expression variation (EL vs. $Expr_{Var}$) and found a strong significant positive correlation (Figure 4.3, Table 4.1).

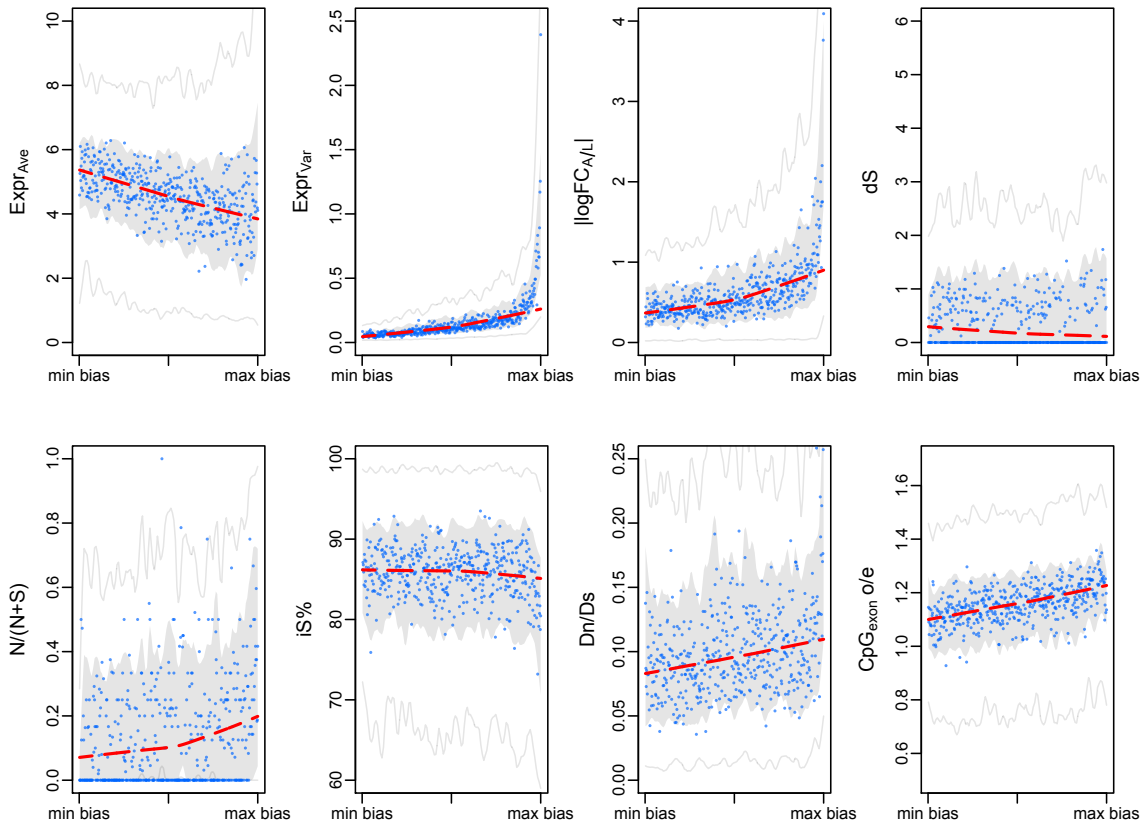


Figure 4.3: Relationship between gene expression bias during larval development and gene regulation/evolution in *C. obscurior*. Genes were sorted in 500 groups based on their overall expression bias (EL). For each group, median (blue dots), inter-quartile ranges (IQR, grey shading) and 1.5xIQR (grey lines) were calculated and plotted. Red dashed lines illustrate the general trend of the relationship and are based on Lowess-smoothed lines (f=0.9).

Table 4.1: Simple correlation between expression bias (EL) and average expression ($Expr_{Ave}$) and semi-partial correlations ($Expr_{Ave}$) with Kendall's correlation coefficient.

	Estimate (τ)	p-value	Statistic	n	Method
Exp_{Ave}	-0.391	< 2.2e-16	-12.757	10,012	Correlation
$Expr_{Var}$	0.352	< 2.2e-16	52.849	10,012	Semi-partial correlation
$\|logFC_{A/L}\|$	0.183	2.86E-166	27.482	10,012	Semi-partial correlation
dS	0.007	0.32	0.999	10,012	Semi-partial correlation
N/(N+S)	0.058	3.08E-10	6.295	5,196	Semi-partial correlation
iS%	-0.022	1.70E-03	-3.138	9,071	Semi-partial correlation
Dn/Ds	0.056	2.93E-12	6.981	6,981	Semi-partial correlation
CpG_{exon} o/e	0.085	4.03E-37	12.730	10,012	Semi-partial correlation

An in-depth analysis of inter-individual variation in gene expression revealed that weakly biased genes also exhibit little variation between individuals of the same morph. In contrast, genes strongly biased between morphs also showed much higher variation in expression between individuals of the same morph (Figure 4.4a, Table 4.2a). However, in each of the four gene sets, inter-individual variation was lowest among samples of the respective focal morph (e.g. lowest variation in EM samples in the EM gene set, Figure 4.4b, Table 4.2b). This decrease was significant in three of four gene sets (WO, WM, EM).

The semi-partial correlation between gene expression biases in two different contexts (EL vs. $\log FC_{A/L}$) was also highly significant (Figure 4.3, Table 4.1). With regard to sequence evolution, we found no significant semi-partial correlation between overall SNV rates and expression bias (Figure 4.3, Table 4.1). However, we found significant monotonic semi-partial correlations between expression bias and intraspecific non-synonymous SNV rates, BLAST identity scores, and inter-specific Dn/Ds rates (Figure 4.3, Table 4.1). In addition, there was a significant and positive monotonic correlation between expression bias and exon CpG o/e rates (Figure 4.3, Table 4.1).

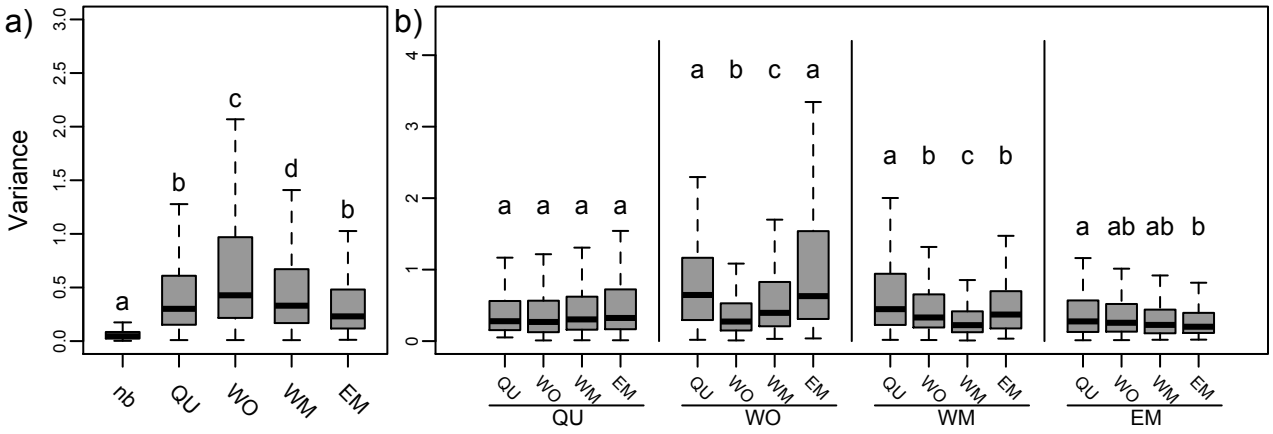


Figure 4.4: Inter-individual gene expression variation for each morph between strongly biased and unbiased genes. (a) Comparison of average within-morph expression variation between the unbiased (nb) and biased gene sets (QU, WO, WM, EM). (b) Comparison of expression variation in the biased gene sets among QU, WO, WM and EM samples. Boxplots show median, inter-quartile range (IQR, box) and 1.5xIQR (whiskers).

Table 4.2a: Bonferroni-corrected p-values from two-tailed pairwise Mann-Whitney U-tests on variance between unbiased and morph-biased gene sets

	EM	ns	QU	WM
ns	< 2e-16	-	-	-
QU	5.90E-06	< 2e-16	-	-
WM	8.30E-11	< 2e-16	0.27	-
WO	< 2e-16	< 2e-16	< 2e-16	1.30E-08

Table 4.2b: Bonferroni-corrected p-values from two-tailed pairwise Mann-Whitney U-tests on variance between morphs in the morph-biased gene sets

QU gene set	EM	QU	WM
QU	0.35	-	-
WM	1	1	-
WO	0.11	1	1

WO gene set	EM	QU	WM
QU	0.904	-	-
WM	8.50E-06	0.001	-
WO	<2e-16	1.90E-14	5.10E-05

WM gene set	EM	QU	WM
QU	0.2	-	-
WM	4.30E-06	5.70E-11	-
WO	1	0.02	3.40E-05

EM gene set	EM	QU	WM
QU	0.039	-	-
WM	1	0.219	-
WO	0.089	1	0.479

In accordance with the correlation analyses, strongly morph biased genes had higher intra- (median N/(N+S): nb 0, QU 0.25, WO 0.25, WM 0.33, EM 0.13, Figure 4.5, Table 4.3) and interspecific rates of non-synonymous substitutions (median Dn/Ds: nb 0.11, QU 0.14, WO 0.17, WM 0.14, EM 0.14, Figure 4.5, Table 4.3), and higher relative CpG content (median CpG_{exon} o/e: nb 1.09, QU 1.23, WO 1.17, WM 1.21, EM 1.25, Figure 4.5, Table 4.3) than unbiased genes. In addition, strongly WO-biased genes had higher overall SNV rates compared to unbiased genes (dS: WO 0.48 kb^{-1}, others 0 kb^{-1}, Figure 4.5, Table 4.3). Between morphs, EM-biased genes had significantly lower rates of intra-specific non-synonymous substitutions than WM-biased genes. There were no significant

differences between morphs in inter-specific Dn/Ds rates in strongly biased genes. CpG_{exon} o/e ratios were significantly increased in QU- and EM-biased genes compared to WO-biased genes.

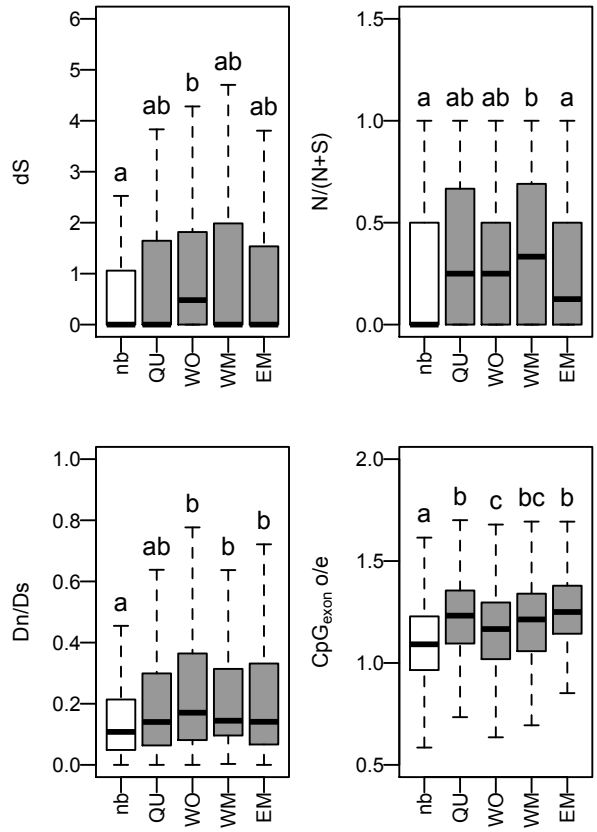


Figure 4.5: Comparison of unbiased and strongly morph-biased genes. Comparisons of overall SNV rates (dS), intra-specific non-synonymous SNV rates (N/(N+S)), inter-specific non-synonymous SNV rates (Dn/Ds) and observed/expected CpG dinucleotide rates (CpG_{exon} o/e) between unbiased (nb) and biased gene sets (QU, WO, WM, EM). Boxplots show median, inter-quartile range (IQR, box) and 1.5xIQR (whiskers).

Table 4.3: Bonferroni-corrected p-values from two-tailed pairwise Mann-Whitney U-tests testing for significant differences in dS, N/(N+S), Dn/Ds and CpG_{exon} o/e rates between the unbiased and morph-biased gene sets.

dS	**EM**	**nb**	**QU**	**WM**
nb	1	-	-	-
QU	1	0.891	-	-
WM	1	0.178	1	-
WO	0.544	0.011	1	1

N/(N+S)	**EM**	**nb**	**QU**	**WM**
Nb	1	-	-	-
QU	1	0.160	-	-
WM	0.006	2.60E-04	0.6114	-
WO	0.910	0.102	1	0.4139

Dn/Ds	**EM**	**nb**	**QU**	**WM**
nb	0.04363	-	-	-
QU	1	0.169	-	-
WM	1	6.50E-04	1	-
WO	1	1.20E-04	0.50499	1

CpG_{exon} o/e	**EM**	**nb**	**QU**	**WM**
nb	2.00E-16	-	-	-
QU	0.732	5.10E-11	-	-
WM	0.051	1.00E-05	1	-
WO	3.40E-06	0.022	0.007	0.659

4.4 Discussion

Our study for the first time shows a correlation between gene expression bias during larval development and changes in gene sequence evolution in a polymorphic insect, while similar studies thus far focused on adult (BG Hunt et al. 2010; BG Hunt, Ometto, et al. 2013b; Purandare et al. 2014) or adult and pupal individuals (BG Hunt et al. 2011). By using data from unpooled, replicated samples and by accounting for the confounding effect of expression level on molecular evolutionary rates in the correlations, our analyses advances on our understanding of the evolution of polyphenism and biased gene expression. Overall, our detailed analyses of genes expressed in a developmental context support previous studies suggesting a prominent role for relaxed selection in accelerating rates of molecular evolution in biased genes. However, accurately disentangling selection regimes under which biased genes evolve poses a particular challenge to evolutionary biology and additional efforts are required to improve our understanding of the complex mechanisms directing a genes molecular evolution.

Polyphenisms are invaluable for studying phenotypic plasticity. However, most if not all of an organism's traits are plastic to some extent as the environment has an effect on every biological process (DeWitt and Scheiner 2004). Fluctuating gene expression in response to changing environmental conditions ("allelic sensitivity") accounts for modulations or passive plasticity of an organism's phenotype (Schlichting and Pigliucci 1995). Passive plasticity is usually considered to be evolutionarily neutral and simply a consequence of the genotype's responsiveness to a heterogeneous environment (Whitman and Agrawal 2009; Chevin et al. 2012). Thus, genes with fluctuating expression accounting for passive plasticity should evolve faster due to relaxed selection (Helanterä and Uller 2014). Prior to the evolution of reproductive division of labour in eusocial insects, their solitary ancestors putatively expressed genes under relaxed selection involved in regulation of passively plastic phenotypes thereby setting the stage for the evolution of discrete morphs (Nijhout 2003; Linksvayer and Wade 2005; Amdam et al. 2006). Consequently, it seems likely that by establishing discrete expression profiles for these genes under relaxed selection, the decisive step in the initial evolution of active plasticity and polyphenism was made (Ghalambor et al. 2007; Ruden et al. 2015). In support, relaxed selection has been suggested as a precursor to the evolution of polyphenism in social hymenoptera (BG Hunt et al. 2011) and in general, genes under relaxed selection are expected to be more prone to become co-opted in the expression of a plastic trait (BG Hunt et al. 2010; BG Hunt, Ometto, et al. 2013b; Helanterä and Uller 2014; Purandare et al. 2014).

By analysing intraspecific divergence between populations we show that biased genes continue to evolve at a higher rate even after biased expression patterns have evolved in a species. If natural and sexual selection pressures are similar in the two studied populations from Brazil and Japan, sequence divergence of biased genes could be attributed to relaxed selection and drift. Alternatively, population-specific adaptations under positive selection could lead to sequence divergence. Our

findings on strong inter-individual variation of gene expression are in support of a prominent role of relaxed selection in accelerating sequence evolution in biased genes, because genes under strong positive selection should be simultaneously subject to precise gene regulation where expressed (Helanterä and Uller 2014). However, that biased genes tend to be least variable in their expression in the morph towards which they are biased might indicate that restriction of spurious transcription can occur, potentially as a result of stabilizing selection on caste specific functionality.

Several theoretical and empirical studies suggest that passive, non-adaptive plasticity that produces a phenotype close enough to a new fitness optimum, can form the basis for the evolution of adaptive and eventually active plasticity (Denver 1997; Nijhout 2003; Ghalambor et al. 2007; Gomez-Mestre et al. 2008; Leichty et al. 2012), resulting in genetic accommodation, release from pleiotropic constraint for the genes with a biased expression pattern, and an increase of directional positive selection of genes co-opted for caste- or morph-specific function. In support, a study on adaptive evolution in the honeybee revealed that genes with worker-biased brain expression evolve under strong positive selection (Harpur et al. 2014). Similarly, studies in *Drosophila* showed signatures of strong positive selection in sex-biased genes (Ellegren and Parsch 2007). These apparently conflicting results on the role of relaxed and positive selection in molecular evolution of biased genes emphasize that there is not a simple model for explaining underlying selection regimes, and teasing apart adaptive and neutral contributions to sequence evolution of morph-biased genes is difficult.

In Hymenoptera, gene-body methylation is assumed to play a role in epigenetically regulating alternative splicing and gene expression by supressing spurious transcription (Bonasio 2012; Chittka et al. 2012). Exon-wide o/e CpG rates are used as an indirect measure of epigenetic regulation in the germ line (Glastad et al. 2012). Since methylated CpG dinucleotides are particularly prone to mutate to TpG, reduction in o/e CpG rates are expected to occur over time in germ line methylated genes over several generations. In honeybees, CpG rates are reduced in genes involved in basic biological processes, and increased in genes involved in development and caste-biased genes (Elango et al. 2009). In accordance, we found a significant positive correlation between CpG o/e ratios and expression bias in the *C. obscurior* genome, supporting the idea that plastically expressed genes tend to be less methylated in social Hymenoptera. In addition, we found differences between genes biased towards a particular morph, with WO-biased genes having the lowest and EM-biased genes having the highest average CpG_{exon} o/e rates. These between-morph differences suggest that morphs require different levels of epigenetic regulation by methylation to develop properly. In favour of this concept, a study in honeybees revealed lower methylation levels in queen than worker larvae (Shi et al. 2012).

EM-biased genes appear to be significantly less methylated in the germline than WO-biased genes, suggesting less restrictive transcriptional regulation in WO-biased genes. This however is in contrast to our analysis of inter-individual variation in gene expression, which showed that WO-biased gene

expression is significantly more variable compared to EM-biased genes. By comparing evolutionary rates in highly morph-biased genes, we found a significant difference in intra-specific non-synonymous SNV rates between EM- and WM-biased genes indicating differences in selection regimes. WMs only occur rarely, thus genes biased towards this morph should evolve under relaxed selection as they are only rarely expressed and less exposed to natural selection (Van Dyken and Wade 2010; Purandare et al. 2014). EMs in contrast are produced regularly and have evolved several adaptations to rivalry and multiple mating that are unique in the social Hymenoptera, indicating strong sexual selection. We hypothesize that this effect results in only mild population genetic divergence in EMs, while at the same time acts on relative stronger regulation of expression.

In eusocial insects, queen fitness depends on performance of the non-reproductive workers, so that selection acts indirectly on the non-reproducing workers (Hamilton 1964). In general, this indirect mode of selection should increase relaxed selection in worker-biased genes (Linksvayer and Wade 2009), even though no study, including ours, so far succeeded to verify this prediction. Even though we did not find a significant correlation between expression bias and overall SNV rates, our analysis of highly morph-biased gene sets revealed higher SNV rates in WO-biased genes. Given that increased mutation rates are known to co-occur with increased recombination rates in genomes (Cutter and Payseur 2013), the increased SNV rates in WO-biased genes can be explained if they are predominantly localized in high-recombining regions (H Liu et al. 2014a). Intriguingly in the non-reproductive workers, CpG to TpG mutations in methylated genes are not transmitted, as workers do not harbour a germline. Thus, transcriptional regulation by methylation to reduce gene expression variation could occur in workers without any effects on the frequency of CpG sites in subsequent generations. In accordance, our analysis of inter-individual variation in gene expression revealed that WO-biased genes tend to be least variable among workers and inter-individual variation of WO-biased genes was much higher in queens and males, suggesting that expression of WO-biased genes is less tightly controlled in these morphs.

Conclusion

A particular challenge in our and similar studies is functional pleitropy of genes. Even though we find a strong positive correlation of gene expression biases in two alternative contexts, genes experiencing expression bias in one context might possibly be constitutively expressed in a different, unaccounted context. Thus, deductions about a gene's evolutionary trajectory will ultimately remain impaired due to incomplete data, until extensive maps of gene expression profiles across life stages, tissues and environmental conditions are available. In addition, biased genes can simultaneously be subject to purifying, positive and relaxed selection depending on the respective context and it is difficult to

resolve these selection regimes (Helanterä and Uller 2014). Where expressed, biased genes should evolve under purifying or positive selection, depending on the imposed selection pressures. In contrast, if expressed weakly, these genes are expected to be under relaxed selection, and the expression pattern itself may potentially change the selective regime (Linksvayer and Wade 2009; Van Dyken and Wade 2010).

Even though functional pleiotropy and the complexity and flexibility of the selection regimes remain difficult to account for in studies focusing on the evolutionary forces acting on biased genes, transcriptome-wide analyses have proven to be useful in illuminating the forces affecting the sequence evolution of biased genes. A deeper understanding of the fidelity of gene expression variation and sequence evolution will help to disentangle the roles of positive and relaxed selection in the evolutionary trajectory of biased genes. With additional high-resolution transcriptomic data sets becoming available, we will be able to address the phenomenon of increased evolutionary rates of biased genes in detail, substantiating our understanding of the evolution of adaptive plasticity and polyphenism.

4.5 Acknowledgements

LS and JO are funded by the Deutsche Forschungsgesellschaft (He1623/31). HH is funded by the Academy of Finland (grant number 135970 to HH and 252411 to the Centre of Excellence in Biological Interactions)

Chapter 5

5. A Novel Intracellular Mutualistic Bacterium in the Invasive Ant *Cardiocondyla obscurior*

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

[1] Institut für Zoologie, Universität Regensburg, 93053 Regensburg, Germany

[2] Institut Canvanilles de Biodiversitat i Biologia Evolutiva (ICBiBE), Parc Cientific de la Universitat de Valencia, C/ Catedratico José Beltrán 2, 46980 Paterna (Valencia), Spain

[3] Max Planck Institute for Chemical Ecology, Hans-Knoell-Str. 8, 07745 Jena, Germany

[4] Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

[5] Department of Biology, University Rochester, River Campus Box 270211, Rochester NY 14627, USA

[6] Área de Genómica y Salud de la Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (FISABIO)–Salud Pública, Avda. de Catalunya 21, 46020 Valencia, Spain

[7] Contributed equally

[8] Corresponding author:
Jan Oettler
Institut für Zoologie,
Universität Regensburg,
Universitätsstr. 31,
93053 Regensburg, Germany Tel +49 9419432996, Fax +499419433304
joettler@gmail.com

Running title:

A novel intracellular symbiont in an invasive ant

Keywords: Bacteriome / *Cardiocondyla obscurior* / Genome reduction / Mutualism / Shikimate pathway / Symbiosis

Subject category: Microbe-microbe and microbe-host interactions

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 *Westeberhardia* is found in gut-associated bacteriomes that deteriorate following eclosion. Only queens maintain *Westeberhardia* 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 *Westeberhardia* (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 titres of *Westeberhardia* during pupal stage, this suggests a contribution of *Westeberhardia* to cuticle formation. Despite a widespread occurrence of *Westeberhardia* across host populations, one ant lineage was found to be naturally symbiont-free, pointing to the loss of an otherwise obligate endosymbiont. This study yields insights into a novel intracellular mutualist that could play a role in the invasive success of *C. obscurior*.

5.1 Introduction

Interactions between organisms drive biological complexity (Maynard Smith and Szathmáry 1997), shaping life as we know it. Symbioses with prokaryotes are considered to promote eukaryote diversification (Brucker and Bordenstein 2012), particularly in insects (Moya et al. 2008; Gil et al. 2010). Some bacterial symbionts provide novel ecological traits to their insect hosts, e.g. defence against pathogens or parasitoids (KM Oliver et al. 2003; Kaltenpoth et al. 2005), enhanced stress tolerance (JA Russell and NA Moran 2006), or nutrients (Douglas 2009). Nutrient-providing symbionts are commonly found in hosts with restricted diets, e.g. aphids feeding on phloem sap (Baumann 2005), blood-feeding Diptera (Jingwen 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 harboured in bacteriocytes, specialized host cells that sometimes form special organ-like structures, the bacteriomes (Baumann 2005), or are confined to the insect gut (Engel and NA Moran 2013). Provisioning with nutrients can lead to increased fitness (Himler et al. 2011; 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, which mate with closely related queens in their maternal nest (Oettler et al. 2010). New colonies are established via colony splitting ("budding") (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 *Westeberhardia*). We analysed its distribution within and across host populations, and compared infection of individual ants depending on morph and age. Furthermore, we localized *Westeberhardia* in the host and scrutinized its genome focusing on its metabolic functions. *Westeberhardia* 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 titres during pupal development point to a role for *Westeberhardia* in cuticle formation.

5.2 Material and Methods

Ant colonies

We keep *C. obscurior* colonies from Brazil (BR), Japan (JP) and Spain (SP) in the laboratory. The BR colonies originated from two collection sites approximately 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 (HGT) (David 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. 2010). 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 Material).

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 was 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. 1999). Inconsistencies were broken and manually re-assembled. Base-calling correction was done using POLISHER (LaButti 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 v2.5 (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 v1.3.1 (Lowe and Eddy 1997), and checked with TFAM v1.4 (Tåquist et al. 2007). tmRNAs and their tag-peptides were predicted using ARAGORN v1.2.36 (Laslett and Canback 2004). The genome was searched against Rfam v11 (Burge et al. 2013) using Infernal v1.1 (Nawrocki and Eddy 2013), and the resulting ncRNAs were manually integrated into the annotation following the INSDC's conventions (insdc.org/files/feature_table.html). Ribosome-binding sites were predicted using RBSfinder (Suzek et al. 2001) and signal peptides were predicted using SignalP v4.1 (Petersen et al. 2011). The resulting annotation was manually curated in Artemis (Rutherford et al. 2000).

Metabolic functions were automatically predicted and analysed using Pathway Tools v17.5 (Karp et al. 2009) against BioCyc and MetaCyc databases (Caspi et al. 2012), following manual curation. Functional information was retrieved from the EcoCyc (Keseler et al. 2013), KEGG (Ogata et al. 2000), and BRENDA (Scheer et al. 2010) 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 (H 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 and DW Russell 2001) (Table 5.1). To verify infection with the same *Westeberhardia* species, we sequenced a 917-bp fragment of the 16S rRNA gene of *Westeberhardia* (*WEOB_122*) (*WE16Sfor*: 5'-CATTTGAATATGTAGAATGGACC-3', *WE16Srev*: 5'-AACTTTTACAAGATCGCTTCTC-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 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 analysed workers and 517 queens. Worker and queen DNA was extracted using the hotshot method (Alasaad et al. 2008) and the NucleoSpin®Tissue XS Kit (Machery-Nagel), respectively. To control for age-effects on infection (see below) we selected freshly hatched 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, 48 days; workers= 2, 14, 28 days), and fertilization state of queens (virgin, mated) for only the BR population (Una, 2012).

Fluorescence *in situ* hybridization (FISH)

To localize *Westeberhardia*, we performed FISH as described previously (Kaltenpoth et al. 2012; 2014) 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'-ATCAGTTTCGAACGCCATTC-3') and Wcard2 (5'-CGGAAGCCACAATTCAAGAT-3'), targeting the 16S rRNA gene. Probes were labelled 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 hatched virgin queens were initiated in a mating arena overnight. The following day we dissected and macerated the spermathecae of queens (n=8) in dH_2O and isolated DNA using the NucleoSpin®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 analysed infection status of two worker pupae each emerging from four of the above crosses using the *nrdB* PCR assay.

5.3 Results

Microbial associates of *Cardiocondyla 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 (JA Russell et al. 2009).

BLASTx analyses further revealed a 360-bp intron-less gene of putative prokaryotic origin encoding a xanthine-guanine phosphoribosyltransferase (XPRT, EC: 2.4.2.22), which is incorporated into the host genome and has its closest ortholog in *Enterobacter cloacae* (WP_023478997). XPRT plays 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 ($median_{larvae}$= 1,140.1, $median_{queens}$= 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 5.1). The genome of *Westeberhardia* has been deposited on Genbank and is hosted on antgenomes.org. 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 S3.1).

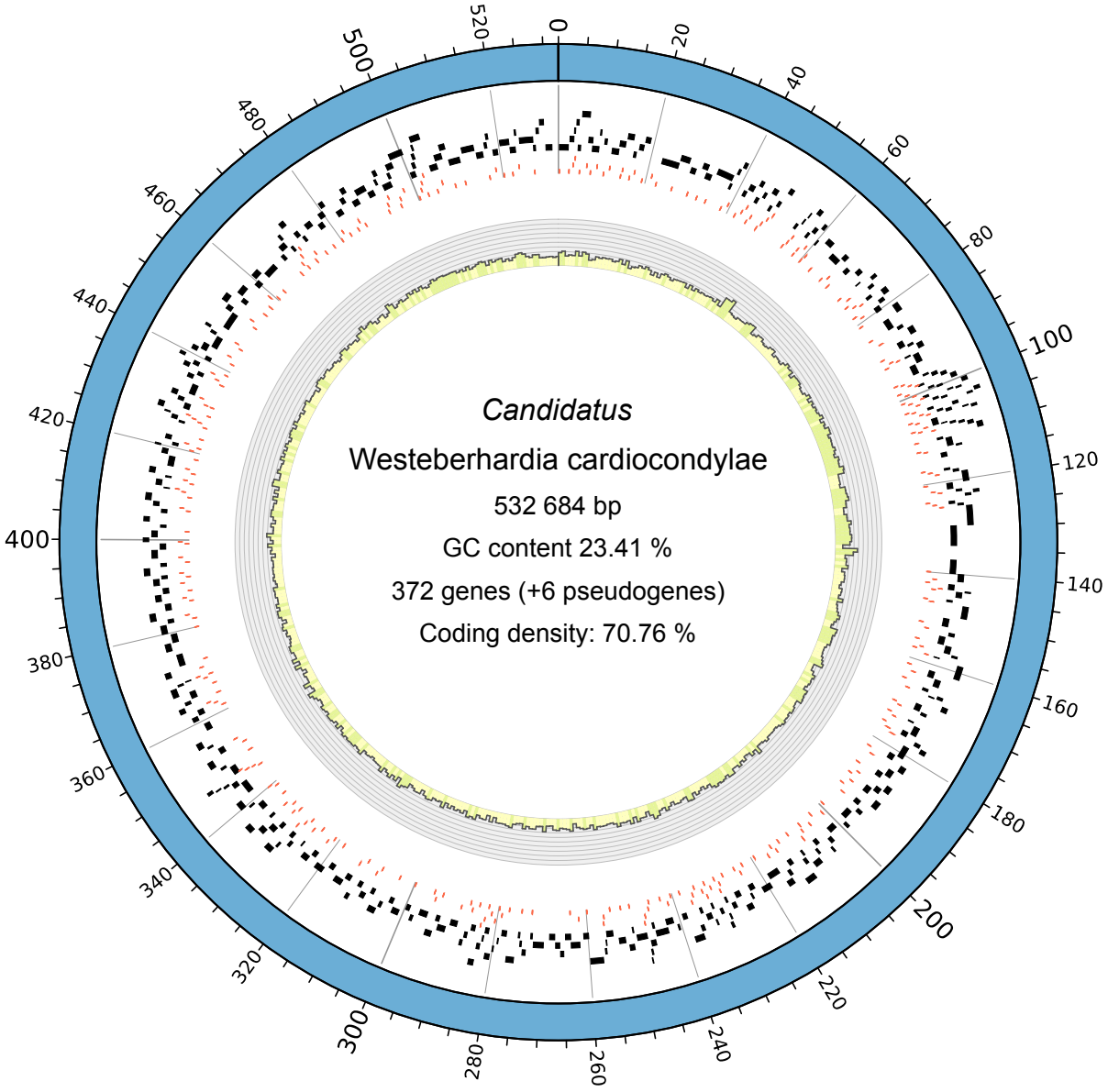


Figure 5.1: Genomic structure of *Westeberhardia*. Representational circos plot (Krzywinski et al. 2009) illustrating genomic properties of *Westeberhardia*. Tile plots show the distribution of protein coding genes (black bars) and ribosomal binding sites (red bars). The histogram in the inner circle shows GC-content in percent for 1-kb windows.

The 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 5.2). *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 glycerophosholipid 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 the pathway for synthesizing 4-hydroxyphenylpyruvate, a precursor of phenylalanine, tyrosine and tryptophan in the shikimate pathway (Hopkins and Kramer 1992; Andersen 2012). While *Westeberhardia* cannot complete the last step of the shikimate pathway to synthesize tyrosine, the host genomes codes for tyrosine aminotransferase (EC 2.6.1.5) converting 4-hydroxyphenylpyruvate to tyrosine (*Cobs_01567*). Further conversion of tyrosine to DOPA (3,4-dihyroxyphenylalanine), 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*).

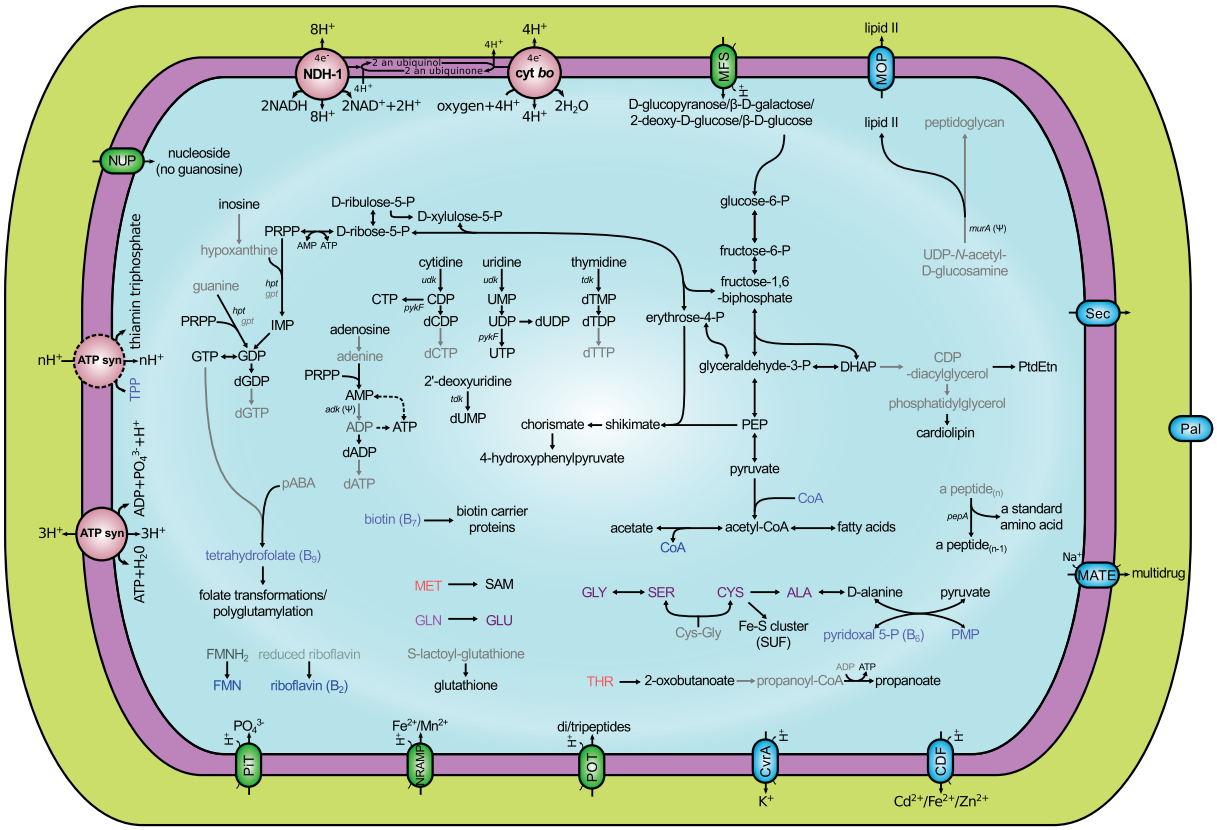


Figure 5.2*: Westeberhardia* metabolic reconstruction. Intact pathways are shown in black lines, unclear pathways (missing a specific gene or having it pseudogenized) are shown in grey lines, and the ones that are already represented elsewhere with another line are shown as dotted lines. Exporters are 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 once with dotted lines to represent another metabolic reaction. 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 BR (Ilhéus, 2009), males from a JP lineage (OypB) are devoid of *Westeberhardia* (coverage BR: 30x, JP: 0.21, Figure 5.3a). qPCR of the *Westeberhardia* specific *nrdB* gene in female pupae as well as larvae and prepupae of unknown sex and caste verified that *Westeberhardia* is completely absent in OypB (Figure 5.3b). Accordingly, *Westeberhardia* was not detected by FISH in sections of adult OypB queens.

Analyses of *C. obscurior* samples collected worldwide showed that *Westeberhardia* is present in 34 of 42 tested samples (81.0 %), including all samples from BR, but absent in some JP populations and in material from Egypt and Sri Lanka (Table 5.1). 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.*

Table 5.1: Prevalence of *Westeberhardia* across populations of *Cardiocondyla obscurior* and the closely related species *Cardiocondyla wroughtonii*, based on a diagnostic PCR screen using the *nrdB* gene. W= worker, Q= queen, Y= *Westeberhardia* present, N= *Westeberhardia* absent. Sample sizes are given in brackets.

Sampling site (year)	Sample description	Morph	*Westeberhardia*
	Cardiocondyla obscurior		
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 et al. (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 et al. (2010)	W	N (1)
FIJI (2007)	EtOH material	W	Y (1)
MALAYSIA: Ulu Gombak (2002)	Oettler et al. (2010)	W	Y (1)
SRI LANKA (2006)	Oettler et al. (2010)	W	N (1)
USA: Lake Alfred, Florida (2004)	Oettler et al. (2010)	W	Y (3)
C. cf obscurior SINGAPORE (2014)	EtOH material	W	Y (2)
	Cardiocondyla wroughtonii		
JAPAN: Nakijin (2013)	Laboratory colonies	W	Y (2)

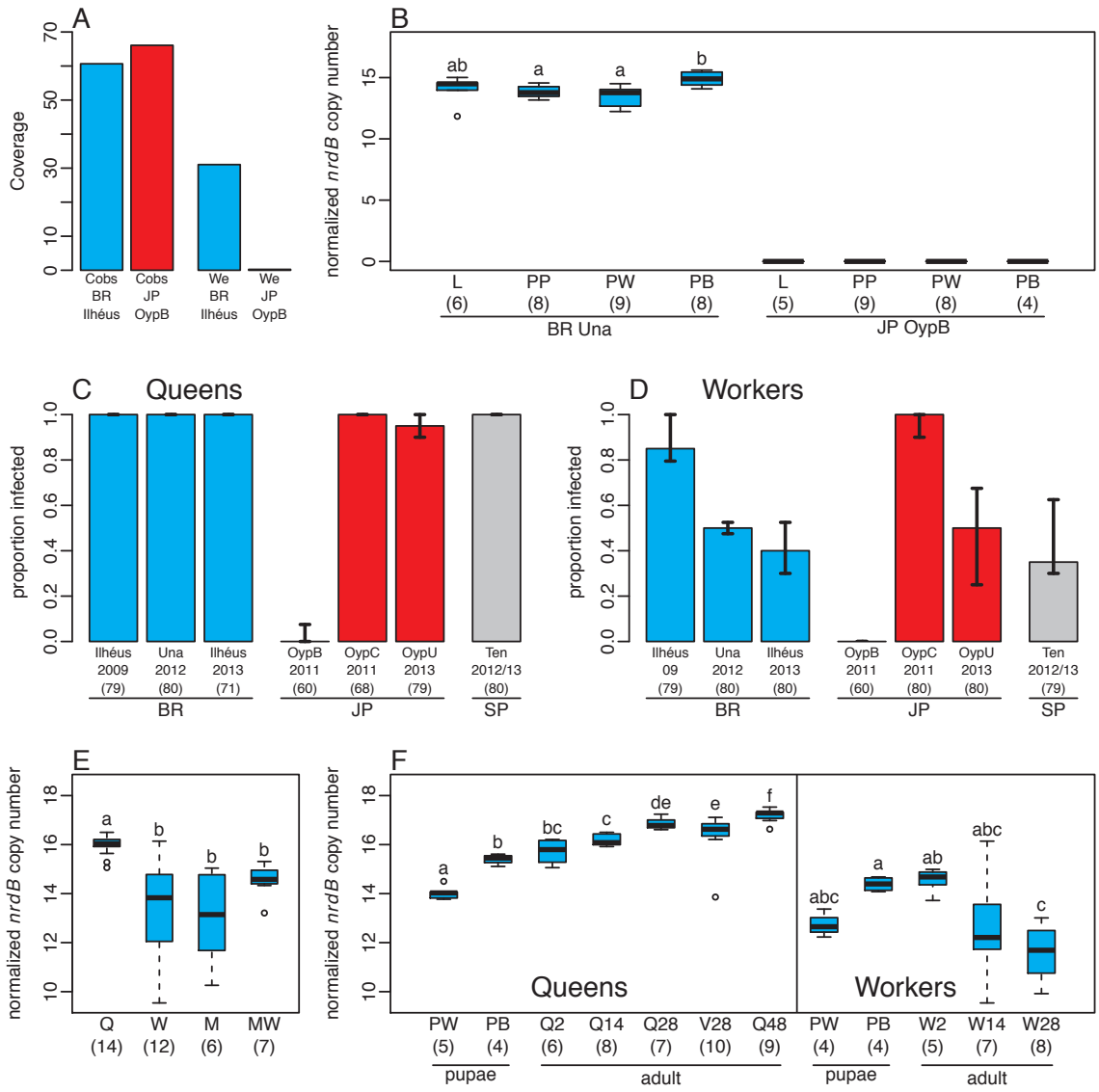


Figure 5.3: Intraspecific and temporal dynamics of *Westeberhardia* infection. (a) In genomic coverage data for pooled haploid males mapped against the *Westeberhardia* reference, *Westeberhardia* reads (We) were exclusively present in the Brazil, Ilhéus (2009) sample (BR, blue) and no reads could be 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 *Westeberhardia* 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 *Westeberhardia* in queens (c) and workers (d) across different populations of *C. obscurior* from Brazil (BR, blue), Japan (JP, red) and Tenerife, Spain (SP, grey), 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, whiskers denote quartiles. Note that while *Westeberhardia* 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 *Westeberhardia* in *C. obscurior* individuals from Brazil (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 difference for within-caste comparisons) (f). Sample sizes are given in parentheses.

Westeberhardia infection of workers varies considerably within and among infected lineages, ranging from 42.5 to 96.3 % (Figure 5.3a), whereas queen infection is almost fixed in all populations (88.6 to 100 %, Figure 5.3c). 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 ($N_{pupa\ white}= 9$, $N_{pupa\ brown}= 8$; t-test: $t(14.7)= -4.3$, $p< 0.001$), but is not different between larvae, prepupae and early pupae (Figure 5.3b, Supplementary Table S3.1). *Westeberhardia* titres are higher in queen compared to worker pupae (t-test: $t(13.3)= 2.6$, $p= 0.023$). A comparison of two to 14 day old adults of each morph (queens, workers, winged and wingless males) shows that *Westeberhardia* titres differ significantly across castes (Kruskal-Wallis: $X^2= 24.2$, $df= 3$, $p< 0.001$, Figure 5.3e), with queens having significantly more *Westeberhardia* than other morphs, which are not different from each other (pairwise Mann-Whitney U-tests with Benjamini-Hochberg correction for multiple testing, Supplementary Table S3.2). We calculated generalized linear models (GLMs) with Gaussian distribution and identity link function to model age dependency of *Westeberhardia* in adult females (Figure 5.3f). In workers, infection decreases with age (GLM: $df= 18$, $F= 12.7$, $p= 0.002$; Supplementary Table S3.3) 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 two after eclosion to day 48 (GLM: $df= 38$, $F= 29.4$, $p< 0.001$, Supplementary Table S3.4). Virgin and mated 28 days old queens show no significant difference in infection (Wilcoxon rank sum test: $W= 49$, $p> 0.05$).

Localization of *Westeberhardia*

Westeberhardia is localized intracellularly in bacteriomes connected to the gut in queen, worker and wingless male pupae (Figure 5.4a-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 5.4d). However, both pupal and adult queens show high *Westeberhardia* abundances in the ovaries (Figure 5.4a,d-f). In particular, *Westeberhardia* is localized predominantly in the nurse cells. Several sections captured trans-generational infection events of the symbiont from the maternal nurse cells into late stages of the developing oocyte (Figure 5.4f).

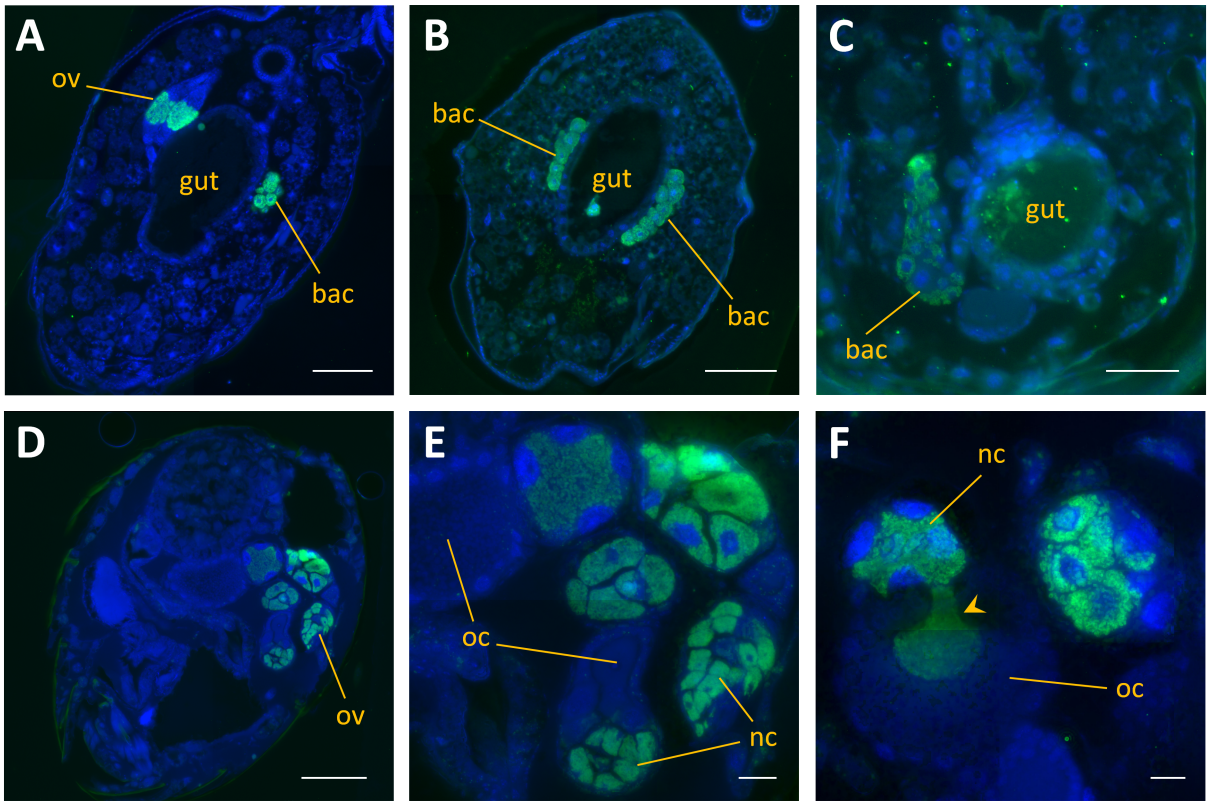


Figure 5.4: Localization of *Westeberhardia* in adults and pupae of *C. obscurior* (from Brazil, Ilhéus, 2009). Symbionts were stained in longitudinal sections through the abdomen with the *Westeberhardia*-specific probe Wcard2-Cy5 (green), and host cell nuclei were counterstained with DAPI (blue). (a-c) 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 µm (a,b,d), 50 µm (c), and 20 µm (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, indicating that *Westeberhardia* does not cause strong CI. Furthermore, as infected females generally also produce male offspring, PI is also unlikely to be elicited by the symbiont. 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 and Stackebrandt 1995). We propose the name '*Candidatus* Westeberhardia cardiocondylae' strain obscurior for this newly discovered gamma3-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.

5.4 Discussion

The 16S sequence of an unknown Enterobacteriaceae isolated from *C. obscurior* was previously published (JA Russell et al. 2009), 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, due to long branch attraction. As already observed by (Husník et al. 2010), 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 ovarial 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 channelled into the oocyte (Mahajan-Miklos and Cooley 1994), cytoplasmic *Westeberhardia* are swept into the developing oocyte, ensuring complete vertical transmission (Figure 5.4f).

CI is a widespread phenotype induced by some bacterial endosymbionts (Gotoh et al. 2007; Werren et al. 2008). In haplodiploids such as social Hymenoptera, CI affects only diploid offspring, while arrhenotokous parthenogenesis results in unfertilized and thereby unaffected haploid male 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, can also be excluded, as this would lead to exclusive female offspring in infected queens.

In contrast to some other intracellular symbionts (NA Moran and 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 horizontal gene transfer 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* titres. 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 size in host-associated bacteria compared to free-living relatives, small effective population size (N_e) of the eusocial host could lead to even faster genome degeneration. With a coding density of 70.8 %, the genome is surprisingly loosely packed, compared to other endosymbionts with similar-length genomes (88 % coding density on average) (McCutcheon and NA Moran 2011). 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* for replication initiation suggests that bacterial replication could underlie host control similar to some mitochondria and other endosymbionts without *dna*A (Shigenobu et al. 2000; Gil et al. 2003). 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 extrinsic 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 *B. 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 and 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). Likewise, *B. floridanus* can synthesize tyrosine, and increased tyrosine biosynthesis during the host's pupal stage (Zientz et al. 2006) coincides with elevated *Blochmannia* titres (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. However, *Westeberhardia* may also provide precursors for tryptophan or phenylalanine synthesis in the host. Although the precise metabolites provided to the host are unclear at this stage, we propose that the retention of the shikimate pathway despite a severe genome reduction shows that here the mutualistic nature of *Westeberhardia* must be most efficient.

After hatching, *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 labour in a eusocial host, most of the *Westeberhardia* population encounters 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, while queens retain *Blochmannia* in their ovarioles (Sauer et al. 2000; Wolschin et al. 2004). Likewise, 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 how 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 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 fact could indicate a facultative status of *Westeberhardia* as it occurs with facultative endosymbionts in *Acyrthosiphon pisum* (NA Moran et al. 2005). However, *Westeberhardia* lacks the main characteristics shared by facultative symbionts, even those of a facultative symbiont in the transition to become 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 and Latorre 2014)). The occurrence of *Westeberhardia* in *C. wroughtonii* and in worldwide collected samples of *C. obscurior* indicates an ancestral infection, and may suggest that secondary loss of *Westeberhardia* in OypB could have occurred through drift (Reuter et al. 2005). Alternatively the impact of facultative symbionts may depend on the particular environmental conditions (Dale and Welburn 2001; Hansen et al. 2007; Haine 2008), indicating that a shift in diet or different gut microbiota could explain symbiont loss. 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 for the first time describes an intracellular mutualist that maintains an obligate relationship with its host but can be lost in some conditions. Its genomic organization, 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 endosymbionts facilitating cuticle build-up during insect development (Zientz et al. 2006; Vigneron et al. 2014) suggests a single origin of metamorphosis-based symbiosis.

Due 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. While 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 2005a).

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

Chapter 6

6. General Discussion

6.1 *Cardiocondyla obscurior* and other Ant Species as Laboratory Models in the Genomics Era

The genus *Cardiocondyla* has raised scientific interest because of its fascinating fighting males that are so profoundly different from the males of other ant species (see Chapter 1.5). In addition, the robustness of several species under laboratory conditions led to the application of *Cardiocondyla* and *C. obscurior* in particular to different fields of research including the study of immunity (e.g. Ugelvig et al. 2010), sexual conflict (e.g. Schrempf et al. 2005), local mate competition (e.g. Suefuji et al. 2008), reproductive skew (e.g. Yamauchi et al. 2007), ageing (e.g. Schrempf et al. 2011), alternative reproductive tactics (e.g. Cremer et al. 2010), sex/caste determination (e.g. Frohschammer and Heinze 2008), and chemical communication (e.g. Will et al. 2012).

Part of the motivation of this thesis was to substantiate the available resources for studies using the model *C. obscurior*. The technological progress in high-throughput sequencing techniques over the last decade allowed us to sequence the genome of *C. obscurior* (Chapter 2) and the endosymbiont *Westeberhardia* (Chapter 5), hence providing a solid foundation for future genetic studies. The analysis of extensive gene expression data from larvae of known developmental fate (Chapter 3, Chapter 4) adds further grounding by demonstrating the suitability of *C. obscurior* as a laboratory model for developmental plasticity and gene expression studies.

Together, the genetic resources and the established laboratory populations are powerful tools and provide a promising framework to extend the use of *C. obscurior* as a model. Several projects applying the here-developed genetic resources have already been initiated during this thesis. For example, *C. obscurior* is currently used to study behavioural genetics of foraging and aggression, the genetic basis of ageing, sex determination in haplodiploid species, and the consequences of sexual conflict. In addition, further efforts are being made to improve the available genomic resources and to extend research on *C. obscurior* to population genomics and experimental evolution. Furthermore, the findings presented in Chapter 2 provide sufficient ground to further analyse the genomic architecture and flexibility of *C. obscurior* with the intention of increasing our understanding of TEs in rapid adaptation of invasive species (Stapley et al. 2015).

While *C. obscurior* offers many advantages for laboratory use and, with the developed genomic resources, also genetic work, its use is currently restricted to very few research groups. Such restrictions also apply to other ant species that are studied in laboratories around the world and the establishment of a model species that is widely distributed and used in the myrmecological community is still not in sight. Species that meet the requirements to become a consistently used model are rare. However, in addition to *C. obscurior* there are other promising candidates such as the pharaoh ant *Monomorium pharaonis* or the clonal raider ant *Cerapachys biroi*.

Similar to *C. obscurior*, colonies of *M. pharaonis* and *C. biroi* can be reared under laboratory conditions with relative ease and their genome sequences have been published recently (Oxley et al. 2014; Mikheyev and Linksvayer 2015). The originally tropical *M. pharaonis* has been spread globally by human commerce and constitutes a massive pest threat in urban environments because of its ability to nest in artificial and highly disturbed habitats. Pharaoh ants are highly polygynous with a distinct queen-worker polyphenism and colonies regularly produce new reproductives, which mate intranidally. In the laboratory, colonies can be manipulated extensively, allowing the application of *M. pharaonis* to a wide range of research approaches (Schmidt et al. 2010; Tay et al. 2014).

Like *C. obscurior* and *M. pharaonis*, *C. biroi* is a successful invader of new habitats in the tropics and subtropics and even fragmented colonies can suffice to establish stable populations (Wetterer et al. 2012). The motivation to sequence the genome of *C. biroi* was explicitly "to establish the clonal raider ant […] as a model eusocial organism" (Oxley et al. 2014) and some biological features of *C. biroi* are indeed advantageous to this endeavour. For one, colonies of *C. biroi* comprise only totipotent workers that reproduce clonally in a colony cycle with reproductive and brood-care phases. Furthermore, colonies of *C. biroi* consist of genetically virtually identical workers and different clonal lines can easily be maintained under laboratory conditions (Teseo et al. 2013). Both phases of the colony cycle can be induced experimentally and colonies consisting of phenotypically and genetically different clonal lines can be compiled, allowing for precise control over colony and individual phenotypes. However, because of its highly derived biology *C. biroi* cannot be used to study key features of eusocial insects, such as the discrete queen-worker polyphenism or effects of indirect fitness.

In spite of their large potential as laboratory models, *C. obscurior*, *M. pharaonis* and *C. biroi* are only used in few myrmecological research institutes and the key to successfully establishing a consistently used model likely lies in the distribution and exchange of colonies of different species between laboratories.

6.2 Phenotypic Plasticity and Transposable Elements as Potential Sources for Rapid Adaptation in *Cardiocondyla obscurior*

Traditionally, evolution by natural selection was considered a slowly advancing process in which adaptations and evolutionary innovations emerge through gradual changes over many generations. Contrasting this classic evolutionary concept, many empirical studies now provide evidence for rapid adaptive evolution over few generations in natural populations facing environmental change (Reznick and Ghalambor 2001). In most cases studied so far, such environmental changes are a consequence of anthropogenic disturbance such as introduction of alien species, soil contamination with heavy metals, or pesticide use (Reznick and Ghalambor 2001). In the face of climate change and ever-increasing anthropogenic disturbance of natural habitats, research on rapid adaptation is becoming an increasingly important field in ecology, conservation biology, and evolutionary biology (Lee 2002; Prentis et al. 2008; Lee and Gelembiuk 2008; Shimada et al. 2009).

This thesis comprises a first assessment of the potential for rapid adaptive evolution in the invasive ant *C. obscurior*. In general, invasive species are well-suited for studies on rapid adaptation (Prentis et al. 2008) because successful establishment of new, isolated populations requires rapid evolution of locally adapted phenotypes. However, not all species are able to adapt to new environments following displacement (Sakai et al. 2001), suggesting that a species' invasive potential is an evolving phenotypic trait under indirect selection (Lee and Gelembiuk 2008). Thus, the "evolution of evolvability" (Lee and Gelembiuk 2008) allowing species to rapidly adapt is suggested to be a product of natural selection on different traits in a natively fluctuating environment. The biology of *C. obscurior* features several traits that potentially facilitate the establishment of stable populations in a changed environment. For one, intranidal mating, colony reproduction by budding, and short generation time allow colonies of *C. obscurior* to reproduce and expand rapidly after displacement. In addition, the tendency to nest at easily disturbed, ephemeral sites indicates a high level of flexibility of colonies and adaptation to novel food sources is expected to be simple for generalist species like *C. obscurior*. Furthermore, for what is known from laboratory experience, colonies of *C. obscurior* are highly robust and have a high potential for recovery after population decline. Both, robustness (Lee and Gelembiuk 2008) and the ability/possibility to recover a population after decline (Reznick and Ghalambor 2001) have been suggested as beneficial pre-adaptations to invasiveness and rapid adaptability. Finally, if local conditions become too harsh winged males of *C. obscurior* potentially enable dispersal and outbreeding to further increase genetic variation, even though empiric evidence for the adaptive value of winged males is still lacking.

Like any other form of natural selection, rapid adaptation generally requires genetic variation within a population. Such variation can either result from formerly cryptic genetic variation (CGV, see below)

that is exposed to selection by shifting conditions or arise from *de novo* mutation (Prentis et al. 2008). In addition to ecological and life-history traits, high levels of developmental plasticity and the peculiar genomic distribution of TEs might render *C. obscurior* particularly well adapted to endure and prosper in a changing environment.

In general, stress and fluctuating selection pressures in a novel environment are expected to increase mutation rate, either as an adaptive response or as a simple by-product of physiological stress (reviewed in Lee and Gelembiuk 2008). Among the factors affecting the mutation rate under stressful conditions, TEs are particularly powerful in restructuring the genome and transcriptome (see Chapter 1). Thus, the increase in TE activity following exposure to a new environment could facilitate rapid adaptations to novel environments by generating a substantial amount of genetic and phenotypic variation in developing populations (Prentis et al. 2008; Barrón et al. 2014; Stapley et al. 2015). In accordance, the genomic analysis presented in Chapter 2 suggests that TEs are of vital importance to the adaptability of *C. obscurior*. The accumulation of TEs in discrete regions of the genome furthermore indicates that their localization is not entirely random but shaped by evolutionary constraints and selection. Thus in *C. obscurior*, TEs appear to be partially domesticated, allowing the host to confine the disruptive potential to certain genomic regions and decrease the likelihood of disruptive mutations. Whether this biased distribution of TEs evolved in response to displacement to novel environments or whether it constitutes a pre-adaptation to invasiveness remains elusive as long as *C. obscurior* has not been studied in its native range. In general however, contemporary concepts assume that a species' invasive potential is shaped by its evolutionary history and that adaptations, which evolved in a species' native range may also facilitate rapid adaptations to novel environments (Lee and Gelembiuk 2008).

In addition to *de novo* arising mutations, the exposure of formerly cryptic genetic variation is an important source of genetic and phenotypic variation in rapid adaptations. Under constant conditions, a large fraction of the standing genetic variation between individuals has little effect on the range of phenotypes expressed in a population. As a consequence, such effectively neutral polymorphisms can accumulate in a population, constituting a rich source of evolutionary potential that can get uncovered by changed environmental conditions (Gibson and Dworkin 2004).

Key mechanisms to the accumulation of CGV are phenotypic plasticity and canalization (phenotypic robustness) (Schlichting 2004; Schlichting and Wund 2014). By adaptively adjusting the expression of a trait in a compensatory response, phenotypic plasticity buffers the effects of novel genetic variants, thus enabling the accumulation of CGV (Moczek 2008; Pfennig et al. 2010). In addition, plastic trait expression should promote the accumulation of genetic variation because conditionally expressed

traits are less exposed to selection compared to constitutively expressed traits (Kawecki 1994; Snell-Rood et al. 2010; Ledon-Rettig et al. 2014). Similarly, canalization by which phenotypic traits remain constant even under changing genetic or environmental conditions promotes the accumulation of CGV (Gibson and Dworkin 2004; Schlichting 2008).

However, environmental change disturbs the equilibrium of canalized systems and previously unexpressed, aberrant phenotypic traits can become exposed by plasticity. By translating formerly cryptic genetic variation into phenotypic variation, exposure of CGV can provide a rich substrate for natural selection by exerting strong phenotypic effects. Thus, plasticity and CGV hold much potential for rapid adaptation under new environmental conditions (Schlichting 2008). Another important contribution of phenotypic plasticity to rapid evolution is the potential to produce adapted phenotypes independent of genetic variation (phenotypic accommodation; West-Eberhard 2005b), which on the one hand allows organism to respond to changed environmental conditions ("buying time", Schlichting 2004) and on the other hand enable the successive manifestation of adaptive phenotypes through genetic accommodation (West-Eberhard 2005a; Pigliucci et al. 2006).

The importance of phenotypic plasticity in rapid adaptations raises the question to what extent plasticity contributes to the adaptability of *C. obscurior* to novel habitats. Many ant species are successful invaders and tramps (Suarez et al. 2010), suggesting an increased potential for contemporary adaptation in these species. In general, the astonishing ecological diversity of ants (Hölldobler and Wilson 1990; Guénard 2013) implies a high level of adaptability over evolutionary time, although this may not be related to generally higher levels of plasticity in ants. However, in invasive species plasticity tends to be significantly increased compared to non-invasive species, as shown in plants (Davidson et al. 2011). Thus, the invasive success of *C. obscurior* and other ants might at least in part be due to higher levels of phenotypic plasticity.

If phenotypic plasticity and TEs are involved in the evolution of rapid adaptations in *C. obscurior*, their contributions likely differ in timing and onset. Following the introduction to an altered environment, phenotypic plasticity likely buffers changed selection pressures through phenotypic accommodation, allowing incipient populations to persevere the initial exposure to new conditions. In addition, exposed CGV in incipient populations may generate phenotypic variability on which natural selection can act to select locally adapted variants. Concomitantly, exposure to environmental stress leads to the liberation of TEs, which then begin to generate genetic variation in the population, thus producing additional potential for adaptive phenotypic variants.

As addressed in Chapter 5, endosymbionts are an additional source of phenotypic novelty for a host and exposure to a novel environment can substantially affect host-symbiont interactions (Ferrari and Vavre 2011; White 2011). For example, novel environments can render a formerly obligate symbiosis obsolete for the host if symbiont-conferred traits are no longer beneficial under a changed selection regime. Similarly, acquisition of alternative symbionts through horizontal transfer can cause dissociation of previously well-established, ancestral symbiosis. The association of *C. obscurior* with *Westeberhardia* offers an interesting system to study the dynamics of host-symbiont interactions under the influence of regular displacement of the host.

The current model of the *Westeberhardia*-*Cardiocondyla* symbiosis suggests a role for the endosymbiont in provisioning metabolites for proper development to the host. The lability of this association however is revealed by the discovery of a naturally endosymbiont-free strain of *C. obscurior* in Japan. Two different scenarios can explain the disruption of the symbiosis: either the endosymbiont was lost through drift, which would require the subsequent evolution of compensatory mechanisms in the host or, *vice versa*, the host successively became independent from the symbiont, eventually resulting in the disruption of the symbiosis and the loss of the endosymbiont. Loss through drift would require rapid compensatory evolution in the host, which could involve mechanisms outlined above or the acquisition of a novel endosymbiont (Koga et al. 2003). Alternatively, the loss of *Westeberhardia* could present a case of endosymbiont replacement, which is known to occur in invasive species (Lefèvre et al. 2004). Other mechanisms by which *Westeberhardia* could have become obsolete in the Japanese strain involve horizontal gene transfer (Lefèvre et al. 2004; McNulty et al. 2010; Hotopp 2013) or a shift in the ecological niche occupied by the host strain.

6.3 Future Prospects

In spite of several years of experience from laboratory studies on *C. obscurior*, many aspects of its biology remain enigmatic. For example, while the expression of WMs has been suspected to be a response to stressful environmental conditions (Cremer and Heinze 2003), they are never encountered in the introduced populations studied so far. Hence, under natural conditions WMs are apparently not expressed, raising the question whether WMs constitute an atavistic, no longer expressed phenotype that is a mere remnant of ancestral developmental potential (Rajakumar et al. 2012). Alternatively, expression of WMs might only be triggered by specific cues that do not occur in the habitats studied so far (e.g. sudden temperature drop, Schrempf and Heinze 2006). If so, the WM phenotype should be under relaxed selection leading to the decay of associated traits by accumulation of CGV in underlying loci. However, the analyses of gene expression patterns in larvae presented in Chapters 3 and 4 did not reveal an increased mutational rate in WM-biased genes. According to these findings relaxed selection is not increased in genes underlying the WM phenotype, suggesting that under certain natural

conditions WMs are likely expressed and adaptive. Since WMs are common in other species of the genus (Oettler et al. 2010) and only introduced populations of *C. obscurior* have been studied so far, one might speculate whether WMs are adaptive and regularly produced in native populations of *C. obscurior* as well. Similar to the hidden supersoldier phenotype in *Pheidole* (Rajakumar et al. 2012) WM expression can be induced in *C. obscurior* by aberrant rearing conditions, allowing us to study development and life history of WMs in detail. However, many of the open questions regarding the evolution of WM remain unanswered until *C. obscurior* can be studied in its native range.

Access to native populations of *C. obscurior* would not only help to uncover the evolutionary significance of the WM phenotype, it would also help to assess the impact of this species' invasion history on its biology. Even though *C. obscurior* is well-equipped to colonize new habitats, it remains elusive which adaptive traits evolved prior to its worldwide spread – as pre-adaptations in its native range (Lee and Gelembiuk 2008) – and which traits evolved in response to frequent displacement of population fragments.

While the study of genome dynamics presented in Chapter 2 revealed TEs as a potential agent for rapid adaptation, further studies are necessary to deepen our understanding of the contribution of TEs to phenotypic evolution in *C. obscurior*. Current efforts to resolve the genomic structure of *C. obscurior* to chromosome-level using RAD sequencing (Klein et al. unpublished) will provide insight into the chromosomal distribution of TE islands. Certain chromosomal regions tend to accumulate TEs due to reduced recombination rates and low gene content (Rizzon et al. 2002). If TE islands in *C. obscurior* are in fact localized in such regions it is likely that they are a product of relaxed selection and thus unlikely to have evolved as an adaptation to repeated exposure to novel environments.

In general, improving the available genomic resources is key to a better understanding of adaptive evolution in *C. obscurior*. Applying recently generated gene expression data (e.g. Chapter 3) to the annotation of coding and non-coding genes, and active TEs will be a first step to uncovering the transcript diversity of *C. obscurior*. The annotation of TEs can furthermore be applied to phylogenetic reconstruction of the evolutionary history of mobile elements in *C. obscurior* (Le Rouzic et al. 2013), allowing insight into short- and long-term TE dynamics. Unravelling contemporary dynamics of TEs and their potential contribution to rapid adaptations is possible by experimental evolution. Due to the relatively short generation time of *C. obscurior*, colonies can be reared under varying selection regimes and environmental stress over several generations such as changes in rearing temperature and humidity, or even pesticide exposure. On the one hand, by performing transposon display assays of stressed and unstressed lines in each generation differences in TE activity can be quantified (Vandenbussche et al. 2013), allowing insight into genomic stress responses. On the other hand, phenotypic responses to different selection regimes can be quantified with regard to the potential emergence of contemporary adaptations.

Overall, the available resources specific to *C. obscurior* and general technical progress in the analysis of genome dynamics now allow for various research approaches. However, to resolve the evolutionary history and the evolutionary perspective of *C. obscurior* in detail, it will be important to study the species with regard to its native environment. Thus, acquisition of live colonies from the native range as well as other introduced populations will be of substantial value for future research.

Chapter 7

7. Summary

Contemporary evolutionary concepts emphasize the importance of environment, development and genetics in shaping an organism's phenotype. Natural selection requires phenotypic differences between individuals, which arise by interactions of environmental factors with genetic systems and are channelled by developmental constraints. This thesis on phenotypic evolution in the invasive ant *C. obscurior* focuses on different sources for phenotypic variation. A brief overview of genetic and environmental mechanisms underlying the evolution of phenotypic novelty is provided in **Chapter 1**, followed by an introduction of the study organism with an emphasis on its feasibility for laboratory studies.

Chapter 2 comprises an assessment of the genomic and phenotypic divergence of two separated *C. obscurior* populations from Brazil and Japan. The genomic structure and genetic differences between the populations suggest transposable elements (TEs) as important agents in generating genetic novelty and potential drivers of rapid adaptation. The discrete genomic distribution patterns of TEs and gene families, differences in gene expression regulation, and signatures of differential genetic divergence reveal a functional sub-structuring of the genome into TE-rich "TE islands" and TE-depleted "low density regions".

In **Chapter 3**, the developmental genetic basis of polyphenism is studied in *C. obscurior*. Based on gene expression data from 28 individually sequenced larvae from four different developmental trajectories (queens, workers, winged males, ergatoid males), this chapter reveals that functionally distinct sets of genes are recruited to the four different developmental pathways, with an emphasis on sphingolipid metabolic genes.

Chapter 4 applies the gene expression data compiled in Chapter 3 in a different context, addressing consequences of biased gene expression on gene evolution. The conducted analyses show that gene expression bias in larvae correlates negatively with average expression levels, but positively with overall expression variation, larvae-to-adult expression bias, sequence divergence within and between species, and relative CpG content. The chapter concludes with a discussion on changing selection regimes and implications for the evolution of biased gene expression.

The study presented in **Chapter 5** contains the first description of the putatively mutualistic endosymbiosis between *Cardiocondyla* and the enterobacterium *Candidatus* Westeberhardia

cardiocondylae ("*Westeberhardia*"). The symbiosis is characterized based on functional and phylogenetic analyses of the endosymbiont's genome and an assessment of infection dynamics across developmental stages and different populations. The results suggest that *Westeberhardia* facilitates development and metamorphosis of its host by provisioning 4-hydroxyphenylpyruvate or derived metabolites that are required for cuticle formation. In spite of the apparent benefits of hosting *Westeberhardia*, a sub-population of *C. obscurior* in Japan was found to be devoid of the endosymbiont, indicating that the symbiosis is not strictly obligatory.

The general discussion in **Chapter 6** begins with a critical assessment and comparison of *C. obscurior* and two other well-suited ant species for their potential to become widely used model systems. The chapter continues with a discussion of potential sources of rapid adaptation in *C. obscurior* in an attempt to synthesize different topics addressed in Chapters 2 to 5. Chapter 6 concludes with discussing possible challenges and directions of future studies on the evolutionary potential of *C. obscurior*.

8. References

Abbot P, Abe J, Alcock J, Alizon S, Alpedrinha JAC, Andersson M, Andre J-B, van Baalen M, Balloux F, Balshine S, Barton N, Beukeboom LW, Biernaskie JM, Bilde T, Borgia G, Breed M, Brown S, Bshary R, Buckling A, Burley NT, Burton-Chellew MN, Cant MA, Chapuisat M, Charnov EL, Clutton-Brock T, Cockburn A, Cole BJ, Colegrave N, Cosmides L, Couzin ID, Coyne JA, Creel S, Crespi B, Curry RL, Dall SRX, Day T, Dickinson JL, Dugatkin LA, Mouden El C, Emlen ST, Evans J, Ferriere R, Field J, Foitzik S, Foster K, Foster WA, Fox CW, Gadau J, Gandon S, Gardner A, Gardner MG, Getty T, Goodisman MAD, Grafen A, Grosberg R, Grozinger CM, Gouyon P-H, Gwynne D, Harvey PH, Hatchwell BJ, Heinze J, Helanterä H, Helms KR, Hill K, Jiricny N, Johnstone RA, Kacelnik A, Kiers ET, Kokko H, Komdeur J, Korb J, Kronauer D, Kümmerli R, Lehmann L, Linksvayer TA, Lion S, Lyon B, Marshall JAR, McElreath R, Michalakis Y, Michod RE, Mock D, Monnin T, Montgomerie R, Moore AJ, Mueller UG, Noë R, Okasha S, Pamilo P, Parker GA, Pedersen JS, Pen I, Pfennig D, Queller DC, Rankin DJ, Reece SE, Reeve HK, Reuter M, Roberts G, Robson SKA, Roze D, Rousset F, Rueppell O, Sachs JL, Santorelli L, Schmid-Hempel P, Schwarz MP, Scott-Phillips T, Shellmann-Sherman J, Sherman PW, Shuker DM, Smith J, Spagna JC, Strassmann B, Suarez AV, Sundström L, Taborsky M, Taylor P, Thompson G, Tooby J, Tsutsui ND, Tsuji K, Turillazzi S, Úbeda F, Vargo EL, Voelkl B, Wenseleers T, West SA, West-Eberhard MJ, Westneat DF, Wiernasz DC, Wild G, Wrangham R, Young AJ, Zeh DW, Zeh JA, Zink A. 2011. Inclusive fitness theory and eusociality. *Nature* **471**: E1–4– author reply E9–10.

Abouheif E, Wray GA. 2002. Evolution of the gene network underlying wing polyphenism in ants. *Science* **297**: 249–252.

Abouheif E, Fave M-J, Ibarraran-Viniegra AS, Lesoway MP, Rafiqi AM, Rajakumar R. 2013. Eco-Evo-Devo: The Time Has Come.Landry CR, Aubin-Horth N eds. *Ecological Genomics*. New York: Springer Science & Business Media, 107–125.

Adachi-Yamada T, Gotoh T, Sugimura I, Tateno M, Nishida Y, Onuki T, Date H. 1999. De novo synthesis of sphingolipids is required for cell survival by down-regulating c-Jun N-terminal kinase in *Drosophila* imaginal discs. *Molecular and Cellular Biology* **19**: 7276–7286.

Akagi K, Li J, Symer DE. 2013. How do mammalian transposons induce genetic variation? A conceptual framework. *BioEssays* **35**: 397–407.

Alasaad S, Rossi L, Maione S, Sartore S, Soriguer RC, Pérez JM, Rasero R, Zhu XQ, Soglia D. 2008. HotSHOT Plus ThermalSHOCK, a new and efficient technique for preparation of PCR-quality mite genomic DNA. *Parasitology Research* **103**: 1455–1457.

Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and Environmental Microbiology* **56**: 1919–1925.

Amdam GV, Csondes A, Fondrk MK, Page RE. 2006. Complex social behaviour derived from maternal reproductive traits. *Nature* **439**: 76–78.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology* **11**: R106.

Anders S, Pyl PT, Huber W. 2015. HTSeq - a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Andersen SO. 2012. Cuticular Sclerotization and Tanning.Denlinger DL, Yocum GD, Rinehart JL eds. *Comprehensive Molecular Insect Science*. Elsevier, 167–192.

AntWeb. – antweb.org/description.do?rank=species&genus=cardiocondyla&name=obscurior (accessed 24.03.2015)

Anxolabéhère D, Kidwell MG, Periquet G. 1988. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P* elements. *Molecular biology and evolution* **5**: 252–269.

Auer PL, Doerge RW. 2010. Statistical design and analysis of RNA sequencing data. *Genetics* **185**: 405–416.

Azevedo SV, Caranton OAM, de Oliveira TL, Hartfelder K. 2011. Differential expression of hypoxia pathway genes in honey bee (*Apis mellifera* L.) caste development. *Journal of Insect Physiology* **57**: 38–45.

Badisco L, Van Wielendaele P, Vanden Broeck J. 2013. Eat to reproduce: a key role for the insulin signaling pathway in adult insects. *Frontiers in physiology* **4**: 202.

Badyaev AV. **2005**. Stress-induced variation in evolution: from behavioural plasticity to genetic assimilation. *Proceedings of the Royal Society B: Biological Sciences* **272**: 877–886.

Bailey JA, Liu G, Eichler EE. **2003**. An *Alu* transposition model for the origin and expansion of human segmental duplications. *American Journal of Human Genetics* **73**: 823–834.

Baker KD, Thummel CS. **2007**. Diabetic larvae and obese flies—Emerging studies of metabolism in *Drosophila*. *Cell Metabolism* **6**: 257–266.

Balmand S, Lohs C, Aksoy S, Heddi A. **2013**. Tissue distribution and transmission routes for the tsetse fly endosymbionts. *Journal of invertebrate pathology* **112**: S116–S122.

Barchuk AR, Cristino AS, Kucharski R, Costa LF, Simões ZLP, Maleszka R. **2007**. Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC developmental biology* **7**: 70.

Barrick JE, Lenski RE. **2013**. Genome dynamics during experimental evolution. *Nature Reviews Genetics* **14**: 827–839.

Barrón MG, Fiston-Lavier A-S, Petrov DA, González J. **2014**. Population genomics of transposable elements in *Drosophila*. *Annual Review of Genetics* **48**: 561–581.

Baumann P. **2005**. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Review of Microbiology* **59**: 155–189.

Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, Luo L, Zhang XH, Xiao J, Alekseyev YO, Wang S, Levy S, Massion PP, Lenburg M, Spira A. **2011**. Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq. *Cancer Prevention Research* **4**: 803–817.

Begna D, Fang Y, Feng M, Li J. **2011**. Mitochondrial proteins differential expression during honeybee (*Apis mellifera* L.) queen and worker larvae caste determination. *Journal of proteome research* **10**: 4263–4280.

Beldade P, Mateus ARA, Keller RA. **2011**. Evolution and molecular mechanisms of adaptive developmental plasticity. *Molecular Ecology* **20**: 1347–1363.

Berens AJ, Hunt JH, Toth AL. **2015**. Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Molecular biology and evolution* **32**: 690–703.

Bergé JB, Feyereisen R, Amichot M. **1998**. Cytochrome P450 monooxygenases and insecticide resistance in insects. *Philosophical Transactions of the Royal Society B: Biological Sciences* **353**: 1701–1705.

Blomquist GJ, Nelson DR, de Renobales M. **1987**. Chemistry, biochemistry, and physiology of insect cuticular lipids. *Archives of insect biochemistry and physiology* **6**: 227–265.

Boersma M, Spaak P, De Meester L. **1998**. Predator-mediated plasticity in morphology, life history, and behavior of *Daphnia*: The uncoupling of responses. *The American Naturalist* **152**: 237–248.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. **2011**. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**: 578–579.

Bohbot J, Pitts RJ, Kwon HW, Rützler M, Robertson HM, Zwiebel LJ. **2007**. Molecular characterization of the *Aedes aegypti* odorant receptor gene family. *Insect Molecular Biology* **16**: 525–537.

Bonasio R. **2012**. Emerging topics in epigenetics: ants, brains, and noncoding RNAs. *Annals of the New York Academy of Sciences* **1260**: 14–23.

Bonasio R, Li Q, Lian J, Mutti NS, Jin L, Zhao H, Zhang P, Wen P, Xiang H, Ding Y, Jin Z, Shen SS, Wang Z, Wang W, Wang J, Berger SL, Liebig J, Zhang G, Reinberg D. **2012**. Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Current biology* **22**: 1755–1764.

Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, Qin N, Donahue G, Yang P, Li Q, Li C, Zhang P, Huang Z, Berger SL, Reinberg D, Wang J, Liebig J. **2010**. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329**: 1068–1071.

Boomsma JJ, Beekman M, Cornwallis CK, Griffin AS, Holman L, Hughes WOH, Keller L, Oldroyd BP, Ratnieks FLW. **2011**. Only full-sibling families evolved eusociality. *Nature* **471**: E4–E5.

Bourke AF. **2011**. *Principles of social evolution*. Oxford University Press Oxford.

Boutin TS, Le Rouzic A, Capy P. **2011**. How does selfing affect the dynamics of selfish transposable elements? *Mobile DNA* **3**: 5–5.

Bowsher JH, Wray GA, Abouheif E. **2007**. Growth and patterning are evolutionarily dissociated in the vestigial wing discs of workers of the red imported fire ant, *Solenopsis invicta*. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **308**: 769–776.

Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. **2008**. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**: 311–322.

Brown TA. **2002**. How Genomes Evolve.Brown TA ed. *Genomes*. Wiley-Liss.

Brucker RM, Bordenstein SR. **2012**. Speciation by symbiosis. *Trends in Ecology & Evolution* **27**: 443–451.

Buchon N, Vaury C. **2006**. RNAi: a defensive RNA-silencing against viruses and transposable elements. *Heredity* **96**: 195–202.

Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A. **2013**. Rfam 11.0: 10 years of RNA families. *Nucleic acids research* **41**: D226–D232.

Burruel V, Klooster KL, Chitwood J, Ross PJ, Meyers SA. **2013**. Oxidative damage to rhesus macaque spermatozoa results in mitotic arrest and transcript abundance changes in early embryos. *Biology of reproduction* **89**: 72.

Cameron RC, Duncan EJ, Dearden PK. **2013**. Biased gene expression in early honeybee larval development. *BMC Genomics* **14**: 903.

Capy P, Gasperi G, Biémont C, Bazin C. **2000**. Stress and transposable elements: co-evolution or useful parasites? *Heredity* **85**: 101–106.

Carroll SB. **2000**. Endless forms: the evolution of gene regulation and morphological diversity. *Cell* **101**: 577–580.

Casacuberta E, González J. **2013**. The impact of transposable elements in environmental adaptation. *Molecular Ecology* **22**: 1503–1517.

Casida JE, Durkin KA. **2013**. Neuroactive insecticides: targets, selectivity, resistance, and secondary effects. *Annual Review of Entomology* **58**: 99–117.

Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Pujar A, Shearer AG, Travers M, Weerasinghe D, Zhang P, Karp PD. **2012**. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic acids research* **40**: D742–D753.

Charlesworth D, Charlesworth B. **1987**. Inbreeding depression and its evolutionary consequences. *Annual review of ecology and systematics* **18**: 237–268.

Charlesworth D, Charlesworth B. **1995**. Transposable elements in inbreeding and outbreeding populations. *Genetics* **140**: 415–417.

Charlesworth D, Wright SI. **2001**. Breeding systems and genome evolution. *Current opinion in genetics & development* **11**: 685–690.

Chen H, Liu Z, Gong S, Wu X, Taylor WL, Williams RW, Matta SG, Sharp BM. **2011**. Genome-wide gene expression profiling of nucleus accumbens neurons projecting to ventral pallidum using both microarray and transcriptome sequencing. *Frontiers in neuroscience* **5**: 98.

Chen M-E, Lewis DK, Keeley LL, Pietrantonio PV. **2004**. cDNA cloning and transcriptional regulation of the vitellogenin receptor from the imported fire ant, *Solenopsis invicta* Buren (Hymenoptera: Formicidae). *Insect Molecular Biology* **13**: 195–204.

Chen X, Hu Y, Zheng H, Cao L, Niu D, Yu D, Sun Y, Hu S, Hu F. **2012**. Transcriptome comparison between honey bee queen- and worker-destined larvae. *Insect Biochemistry and Molecular Biology* **42**: 665–673.

Chevin L-M, Collins S, Lefèvre F. **2012**. Phenotypic plasticity and evolutionary demographic responses to climate change: taking theory out to the field (A Hoffmann, Ed.). *Functional Ecology* **27**: 967–979.

Chevreux B, Wetter T, Suhai S. **1999**. Genome sequence assembly using trace signals and additional sequence information. *German conference on bioinformatics*: 45–56.

Chittka A, Wurm Y, Chittka L. **2012**. Epigenetics: The making of ant castes. *Current biology* **22**: R835–R838.

Conant GC, Wolfe KH. **2008**. Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics* **9**: 938–950.

Conceição IC, Aguade M. **2008**. High incidence of interchromosomal transpositions in the evolutionary history of a subset of or genes in *Drosophila*. *Journal of Molecular Evolution* **66**: 325–332.

Conrad B, Antonarakis SE. **2007**. Gene duplication: A drive for phenotypic diversity and cause of human disease. *Annual Review of Genomics and Human Genetics* **8**: 17–35.

Cordaux R, Bouchon D, Grève P. **2011**. The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends in Genetics* **27**: 332–341.

Corona M, Estrada E, Zurita M. **1999**. Differential expression of mitochondrial genes between queens and workers during caste determination in the honeybee *Apis mellifera*. *The Journal of experimental biology* **202**: 929–938.

Cremer S, Heinze J. **2003**. Stress grows wings: Environmental induction of winged dispersal males in *Cardiocondyla* ants. *Current biology* **13**: 219–223.

Cremer S, Lautenschläger B, Heinze J. **2002a**. A transitional stage between the ergatoid and winged male morph in the ant *Cardiocondyla obscurior*. *Insectes Sociaux* **49**: 221–228.

Cremer S, Schrempf A, Heinze J. **2010**. Competition and opportunity shape the reproductive tactics of males in the ant *Cardiocondyla obscurior*. *PLoS ONE* **6**: e17323–e17323.

Cremer S, Sledge MF, Heinze J. **2002b**. Chemical mimicry: male ants disguised by the queen's bouquet. *Nature* **419**: 897–897.

Cutter AD, Payseur BA. **2013**. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics* **14**: 262–274.

Dale C, Welburn SC. **2001**. The endosymbionts of tsetse flies: manipulating host-parasite interactions. *International journal for parasitology* **31**: 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. *Current biology* **18**: R1087–R1088.

Davidson AM, Jennions M, Nicotra AB. **2011**. Do invasive species show higher phenotypic plasticity than native species and, if so, is it adaptive? A meta-analysis. *Ecology Letters* **14**: 419–431.

de Boer JG, Yazawa R, Davidson WS, Koop BF. **2007**. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* **8**: 422.

Denver RJ. **1997**. Environmental stress as a developmental cue: Corticotropin-releasing hormone is a proximate mediator of adaptive phenotypic plasticity in amphibian metamorphosis. *Hormones and behavior* **31**: 169–179.

DeWitt TJ, Scheiner SM. **2004**. *Phenotypic Plasticity: Functional and Conceptual Approaches*. New York: Oxford University Press.

DeWitt TJ, Sih A, Wilson DS. **1998**. Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution* **13**: 77–81.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See L-H, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigo R, Gingeras TR. **2012**. Landscape of transcription in human cells. *Nature* **489**: 101–108.

Douglas AE. **2009**. The microbial dimension in insect nutritional ecology. *Functional Ecology* **23**: 38–47.

Drapeau MD, Albert S, Kucharski R, Prusko C, Maleszka R. **2006**. Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honey bees. *Genome research* **16**: 1385–1394.

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. **2005**. Why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 14338–14343.

Du Y, Schrempf A, Heinze J. **2007**. Environmental determination of the male morph in the ant *Cardiocondyla obscurior* (Hymenoptera: Formicidae). *European Journal of Entomology* **104**: 243–246.

Dubnau J, Chiang AS, Grady L, Barditch J, Gossweiler S, McNeil J, Smith P, Buldoc F, Scott R, Certa U, Broger C, Tully T. **2003**. The *staufen*/*pumilio* pathway is involved in *Drosophila* long-term memory. *Current*

biology **13**: 286–296.

Duvernell DD, Schmidt PS, Eanes WF. **2003**. Clines and adaptive evolution in the *methuselah* gene region in *Drosophila melanogaster*. *Molecular Ecology* **12**: 1277–1285.

Dyall SD. **2004**. Ancient invasions: From endosymbionts to organelles. *Science* **304**: 253–257.

Eddy SR. **2001**. Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics* **2**: 919–929.

Edger PP, Pires JC. **2009**. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research* **17**: 699–717.

Elango N, Hunt BG, Goodisman MAD, Yi SV. **2009**. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 11206–11211.

Ellegren H, Parsch J. **2007**. The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics* **8**: 689–698.

Engel P, Moran NA. **2013**. The gut microbiota of insects - diversity in structure and function. *FEMS microbiology reviews* **37**: 699–735.

Evans JD, Wheeler D E. **2001**. Expression profiles during honeybee caste determination. *Genome Biology* **2**: research0001.1–0001.6.

Eyre-Walker A, Keightley PD. **2007**. The distribution of fitness effects of new mutations. *Nature Reviews Genetics* **8**: 610–618.

Fatica A, Bozzoni I. **2013**. Long non-coding RNAs: new players in cell differentiation and development. *Nature Reviews Genetics* **15**: 7–21.

Fedoroff NV. **2012**. Transposable elements, epigenetics, and genome evolution. *Science* **338**: 758–767.

Fedoroff NV. **2013**. *Plant Transposons and Genome Dynamics in Evolution* (NV Fedoroff, Ed.). Hoboken, NJ: Wiley-Blackwell.

Feldhaar H. **2011**. Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecological Entomology* **36**: 533–543.

Feldhaar H, Straka J, Krischke M, Berthold K, Stoll S, Mueller MJ, Gross R. **2007**. Nutritional upgrading for omnivorous carpenter ants by the endosymbiont *Blochmannia*. *BMC biology* **5**: 48.

Feldmeyer B, Elsner D, Foitzik S. **2013**. Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. *Molecular Ecology* **23**: 151–161.

Ferrari J, Vavre F. **2011**. Bacterial symbionts in insects or the story of communities affecting communities. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**: 1389–1400.

Feschotte C. **2008**. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics* **9**: 397–405.

Feschotte C, Pritham EJ. **2007**. DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics* **41**: 331–368.

Feschotte C, Jiang N, Wessler SR. **2002**. Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics* **3**: 329–341.

Flagel LE, Wendel JF. **2009**. Gene duplication and evolutionary novelty in plants. *New Phytologist* **183**: 557–564.

Foley B, Chenoweth SF, Nuzhdin SV, Blows MW. **2006**. Natural genetic variation in cuticular hydrocarbon expression in male and female *Drosophila melanogaster*. *Genetics* **175**: 1465–1477.

Fontdevila A. **2011**. *The Dynamic Genome: A Darwinian Approach* (A Fontdevila, Ed.). Oxford: Oxford University Press.

Forsman A. **2014**. Rethinking phenotypic plasticity and its consequences for individuals, populations and species. *Heredity*.

Frohschammer S, Heinze J. **2008**. A heritable component in sex ratio and caste determination in a *Cardiocondyla* ant. *Frontiers in Zoology* **6**: 27–33.

Fusco G, Minelli A. **2010**. Phenotypic plasticity in development and evolution: facts and concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 547–556.

Gadau J, Helmkampf M, Nygaard S, Roux J, Simola DF, Smith CR, Suen G, Wurm Y, Smith CD. **2011**. The genomic impact of 100 million years of social evolution in seven ant species. *Trends in Genetics* **28**: 14–21.

Galiegue S, Jbilo O, Combes T, Bribes E, Carayon P, Le Fur G, Casellas P. **1999**. Cloning and characterization of PRAX-1: A new protein that specifically interacts with the peripheral Benzodiazepine receptor. *The Journal of biological chemistry* **274**: 2938–2952.

Gault CR, Obeid LM, Hannun YA. **2010**. An overview of sphingolipid metabolism: from synthesis to breakdown. *Advances in experimental medicine and biology* **688**: 1–23.

Ghalambor CK, McKay JK, Carroll SP, Reznick DN. **2007**. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology* **21**: 394–407.

Gibson G, Dworkin I. **2004**. Uncovering cryptic genetic variation. *Nature Reviews Genetics* **5**: 681–690.

Gil R, Latorre A, Moya A. **2010**. Evolution of Prokaryote-Animal Symbiosis from a Genomics Perspective. Microbiology Monographs. *Microbiology Monographs*. Berlin, Heidelberg: Springer Berlin Heidelberg, 207–233.

Gil R, Sabater-Munoz B, Latorre A, Silva FJ, Moya A. **2002**. Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 4454–4458.

Gil R, Silva FJ, Peretó J, Moya A. **2004**. Determination of the core of a minimal bacterial gene set. *Microbiology and Molecular Biology Reviews* **68**: 518–537.

Gil R, Silva FJ, Zientz E, Delmotte F, González-Candelas F, Latorre A, Rausell C, Kamerbeek J, Gadau J, Hölldobler B, van Ham RCHJ, Gross R, Moya A. **2003**. The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 9388–9393.

Gilbert LI. **2012a**. *Insect Molecular Biology and Biochemistry*. London: Academic Press.

Gilbert SF. **2001**. Ecological developmental biology: Developmental biology meets the real world. *Developmental Biology* **233**: 1–12.

Gilbert SF. **2005**. Mechanisms for the environmental regulation of gene expression. *Birth Defects Research Part C: Embryo Today: Reviews* **72**: 291–299.

Gilbert SF. **2012b**. Ecological developmental biology: environmental signals for normal animal development. *Evolution & development* **14**: 20–28.

Gilbert SF, Epel D. **2009**. *Ecological Developmental Biology: Integrating Epigenetics, Medicine, and Evolution*. Sunderland, MA: Sinauer Associates.

Glastad KM, Hunt BG, Goodisman MAD. **2012**. Evidence of a conserved functional role for DNA methylation in termites. *Insect Molecular Biology* **22**: 143–154.

Gomez-Mestre I, Wiens JJ, Warkentin KM. **2008**. Evolution of adaptive plasticity: Risk-sensitive hatching in neotropical leaf-breeding treefrogs. *Ecological Monographs* **78**: 205–224.

González J, Karasov TL, Messer PW, Petrov DA. **2010**. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLoS genetics* **6**: e1000905.

Gopinath SCB, Matsugami A, Katahira M, Kumar PKR. **2005**. Human vault-associated non-coding RNAs bind to mitoxantrone, a chemotherapeutic compound. *Nucleic acids research* **33**: 4874–4881.

Gotoh T, Noda H, Ito S. **2007**. *Cardinium* symbionts cause cytoplasmic incompatibility in spider mites. *Heredity* **98**: 13–20.

Gould SB, Waller RF, McFadden GI. **2008**. Plastid evolution. *Annual Review of Plant Biology* **59**: 491–517.

Gould SJ. **1980**. Is a new and general theory of evolution emerging. *Paleobiology* **6**: 119–130.

Guénard B. **2013**. *An Overview of the Species and Ecological Diversity of Ants*. Chichester, UK: eLS. John Wiley & Sons, Ltd.

Guo S, Kim J. **2007**. Molecular evolution of *Drosophila* odorant receptor genes. *Molecular biology and evolution* **24**: 1198–1207.

Haine ER. **2008**. Symbiont-mediated protection. *Proceedings of the Royal Society B: Biological Sciences* **275**: 353–361.

Hamdoun A, Epel D. **2007**. Embryo stability and vulnerability in an always changing world. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 1745–1750.

Hamel S, Fantini J, Schweisguth F. **2010**. Notch ligand activity is modulated by glycosphingolipid membrane composition in *Drosophila melanogaster*. *The Journal of Cell Biology* **188**: 581–594.

Hamilton WD. **1964**. The genetical evolution of social behaviour. I. *Journal of theoretical biology* **7**: 1–16.

Hansen AK, Jeong G, Paine TD, Stouthamer R. **2007**. Frequency of secondary symbiont infection in an invasive psyllid relates to parasitism pressure on a geographic scale in California. *Applied and Environmental Microbiology* **73**: 7531–7535.

Harpur BA, Kent CF, Molodtsova D, Lebon JMD, Alqarni AS, Owayss AA, Zayed A. **2014**. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 2614–2619.

Heddi A, Grenier A-M, Khatchadourian C, Charles H, Nardon P. **1999**. Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and *Wolbachia*. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 6814–6819.

Heinze J. **1998**. Intercastes, intermorphs, and ergatoid queens: who is who in ant reproduction? *Insectes Sociaux* **45**: 113–124.

Heinze J, Hölldobler B. **1993**. Fighting for a harem of queens: physiology of reproduction in *Cardiocondyla* male ants. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 8412–8414.

Heinze J, Schrempf A. **2012**. Terminal investment: individual reproduction of ant queens increases with age. *PLoS ONE* **7**: e35201.

Heinze J, Aumeier V, Bodenstein B, Crewe RM, Schrempf A. **2013**. Wingless and intermorphic males in the ant *Cardiocondyla venustula*. *Insectes Sociaux* **60**: 43–48.

Heinze J, Cremer S, Eckl N, Schrempf A. **2006**. Stealthy invaders: the biology of *Cardiocondyla* tramp ants. *Insectes Sociaux* **53**: 1–7.

Helanterä H, Uller T. **2014**. Neutral and adaptive explanations for an association between caste-biased gene expression and rate of sequence evolution. *Front Genet* **5**: 1–10.

Hemingway J, Ranson H. **2000**. Insecticide resistance in insect vectors of human disease. *Annual Review of Entomology* **45**: 371–391.

Herr DR, Fyrst H, Creason MB, Phan VH, Saba JD, Harris GL. **2004**. Characterization of the *Drosophila* sphingosine kinases and requirement for Sk2 in normal reproductive function. *The Journal of biological chemistry* **279**: 12685–12694.

Herr DR, Fyrst H, Phan V, Heinecke K, Georges R, Harris GL, Saba JD. **2003**. Sply regulation of sphingolipid signaling molecules is essential for *Drosophila* development. *Development* **130**: 2443–2453.

Herre EA, Knowlton N, Mueller UG, Rehner SA. **1999**. The evolution of mutualisms: exploring the paths between conflict and cooperation. *Trends in Ecology & Evolution* **14**: 49–53.

Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, Cravchik A, Collins FH, Robertson HM, Zwiebel LJ. **2002**. G protein-coupled receptors in *Anopheles gambiae*. *Science* **298**: 176–178.

Himler AG, Adachi-Hagimori T, Bergen JE, Kozuch A, Kelly SE, Tabashnik BE, Chiel E, Duckworth VE, Dennehy TJ, Zchori-Fein E, Hunter MS. **2011**. Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science* **332**: 254–256.

Hoffman EA, Goodisman MAD. **2007**. Gene expression and the evolution of phenotypic diversity in social wasps. *BMC biology* **5**: 23.

Holt C, Yandell M. **2010**. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics* **12**: 491–491.

Hopkins TL, Kramer KJ. **1992**. Insect cuticle sclerotization. *Annual Review of Entomology* **37**: 273–302.

Hotopp JCD. **2013**. Lateral Gene Transfer in Multicellular Organisms.Gophna U ed. *Lateral Gene Transfer in Evolution*. New York: Springer.

Hölldobler B, Wilson EO. **1990**. *The Ants*. Berlin: Springer.

Hölldobler B, Wilson EO. **2009**. *The Superorganism*. New York: W. W. Norton & Company.

Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P. **2010**. The struggle for life of the genome's selfish architects. *Biology Direct* **6**: 19–48.

Huang DW, Sherman BT, Lempicki RA. **2008**. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**: 44–57.

Huang DW, Sherman BT, Lempicki RA. **2009**. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**: 1–13.

Hunt BG, Glastad KM, Yi SV, Goodisman MAD. **2013a**. The function of intragenic DNA methylation: Insights from insect epigenomes. *Integrative and Comparative Biology* **53**: 319–328.

Hunt BG, Ometto L, Keller L, Goodisman MAD. **2013b**. Evolution at two levels in fire ants: the relationship between patterns of gene expression and protein sequence evolution. *Molecular biology and evolution* **30**: 263–271.

Hunt BG, Ometto L, Wurm Y, Shoemaker D, Yi SV, Keller L, Goodisman MAD. **2011**. Relaxed selection is a precursor to the evolution of phenotypic plasticity. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 15936–15941.

Hunt BG, Wyder S, Elango N, Werren JH, Zdobnov EM, Yi SV, Goodisman MAD. **2010**. Sociality is linked to rates of protein evolution in a highly social insect. *Molecular biology and evolution* **27**: 497–500.

Hunt JH, Amdam GV. **2005**. Bivoltinism as an antecedent to eusociality in the paper wasp genus *Polistes*. *Science* **308**: 264–267.

Hunt JH, Kensinger BJ, Kossuth JA, Henshaw MT, Norberg K, Wolschin F, Amdam GV. **2007**. A diapause pathway underlies the gyne phenotype in *Polistes* wasps, revealing an evolutionary route to caste-containing insect societies. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 14020–14025.

Hurlbut GD, Kankel MW, Lake RJ, Artavanis-Tsakonas S. **2007**. Crossing paths with Notch in the hyper-network. *Current Opinion in Cell Biology* **19**: 166–175.

Hurst G, Werren JH. **2001**. The role of selfish genetic elements in eukaryotic evolution. *Nature Reviews Genetics* **2**: 597–606.

Husník F, Chrudimský T, Hypša V. **2010**. Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches. *BMC biology* **9**: 87–104.

Husník F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, Tanaka M, Satoh N, Bachtrog D, Wilson ACC, Dohlen von CD, Fukatsu T, McCutcheon JP. **2013**. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* **153**: 1567–1578.

Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. **2010**. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* **11**: 119–130.

Jiménez N, González-Candelas F, Silva FJ. **2000**. Prephenate dehydratase from the aphid endosymbiont (*Buchnera*) displays changes in the regulatory domain that suggest its desensitization to inhibition by phenylalanine. *Journal of Bacteriology* **182**: 2967–2969.

Johnson BR, Linksvayer TA. **2010**. Deconstructing the superorganism: social physiology, groundplans, and sociogenomics. *The Quarterly review of biology* **85**: 57–79.

Jones PA. **2012**. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* **13**: 484–492.

Kaltenpoth M, Engl T. **2014**. Defensive microbial symbionts in Hymenoptera. *Functional Ecology* **28**: 315–327.

Kaltenpoth M, Gottler W, Herzner G, Strohm E. **2005**. Symbiotic bacteria protect wasp larvae from fungal infestation. *Current biology* **15**: 475–479.

Kaltenpoth M, Roeser-Mueller K, Koehler S, Peterson A, Nechitaylo TY, Stubblefield JW, Herzner G, Seger J, Strohm E. **2014**. Partner choice and fidelity stabilize coevolution in a Cretaceous-age defensive symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 6359–6364.

Kaltenpoth M, Yildirim E, Gürbüz MF, Herzner G, Strohm E. **2012**. Refining the roots of the beewolf-*Streptomyces* symbiosis: antennal symbionts in the rare genus *Philanthinus* (Hymenoptera, Crabronidae). *Applied and Environmental Microbiology* **78**: 822–827.

Kamakura M. **2011**. Royalactin induces queen differentiation in honeybees. *Nature* **473**: 478–483.

Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Frise E, Wheeler DA, Lewis SE, Rubin GM, Ashburner M, Celniker SE. **2002**. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biology* **3**: research0084–0084.20.

Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. **2013**. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs (HE Hoekstra, Ed.). *PLoS genetics* **9**: e1003470.

Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, Kaipa P, Gilham F, Spaulding A, Popescu L, Altman T, Paulsen I, Keseler IM, Caspi R. **2009**. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics* **11**: 40–79.

Kawecki TJ. **1994**. Accumulation of deleterious mutations and the evolutionary cost of being a generalist. *American Naturalist*: 833–838.

Kazazian HH. **2004**. Mobile elements: drivers of genome evolution. *Science* **303**: 1626–1632.

Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martínez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M, Muñiz-Rascado L, Ong Q, Paley S, Schröder I, Shearer AG, Subhraveti P, Travers M, Weerasinghe D, Weiss V, Collado-Vides J, Gunsalus RP, Paulsen I, Karp PD. **2013**. EcoCyc: fusing model organism databases with systems biology. *Nucleic acids research* **41**: D605–12.

Kikuchi Y. **2009**. Endosymbiotic bacteria in insects: their diversity and culturability. *Microbes and Environments* **24**: 195–204.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. **2013**. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**: R36.

Kim S-H, Yi SV. **2006**. Understanding relationship between sequence and functional evolution in yeast proteins. *Genetica* **131**: 151–156.

Kinomura K, Yamauchi K. **1987**. Fighting and mating behaviors of dimorphic males in the ant *Cardiocondyla wroughtoni*. *Journal of Ethology* **5**: 75–81.

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. *Proceedings of the National Academy of Sciences of the United States of America* **109**: E1230–E1237.

Koga R, Tsuchida T, Fukatsu T. **2003**. Changing partners in an obligate symbiosis: a facultative endosymbiont can compensate for loss of the essential endosymbiont *Buchnera* in an aphid. *Proceedings of the Royal Society B: Biological Sciences* **270**: 2543–2550.

Kopp A, Duncan I, Godt D, Carroll SB. **2000**. Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature* **408**: 553–559.

Kraut R. **2011**. Roles of sphingolipids in *Drosophila* development and disease. *Journal of Neurochemistry* **116**: 764–778.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. **2009**. Circos: an information aesthetic for comparative genomics. *Genome research* **19**: 1639–1645.

Kucharski R, Maleszka J, Foret S, Maleszka R. **2008**. Nutritional control of reproductive status in honeybees via DNA methylation. *Science* **319**: 1827–1830.

Kugler J. **1983**. The males of *Cardiocondyla emery* (Hymenoptera: Formicidae) with the description of the winged male of *Cardiocondyla wroughtoni* (Forel). *Israel Journal of Entomology* **17**: 1–21.

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**: 538–547.

LaButti K, Foster B, Lowry S, Trong S, Goltsman E, Lapidus A. **2008**. POLISHER: a tool for using ultra short reads in microbialgenome finishing

Lai C-Y, Baumann L, Baumann P. **1994**. Amplification of trpEG: adaptation of *Buchnera aphidicola* to an endosymbiotic association with aphids. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 3819–3823.

Lanfear R, Ho SYW, Love D, Bromham L. **2010**. Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 20423–20428.

Langerhans RB, DeWitt TJ. **2002**. Plasticity constrained: over-generalized induction cues cause maladaptive phenotypes. *Evolutionary Ecology Research* **4**: 857–870.

Langmead B, Salzberg SL. **2012**. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**: 357–359.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. **2007**. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.

Lartillot N, Lepage T, Blanquart S. **2009**. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**: 2286–2288.

Laslett D, Canback B. **2004**. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic acids research* **32**: 11–16.

Latchman DS. **2010**. *Eukaryotic transcription factors*. New York: Academic press.

Law CW, Chen Y, Shi W, Smyth GK. **2014**. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* **15**: R29.

Le Rouzic A, Payen T, Hua-Van A. **2013**. Reconstructing the evolutionary history of transposable elements. *Genome Biology and Evolution* **5**: 77–86.

Leal WS, Choo Y-M, Xu P, da Silva CS, Ueira-Vieira C. **2013**. Differential expression of olfactory genes in the southern house mosquito and insights into unique odorant receptor gene isoforms. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 18704–18709.

Lebman DA, Spiegel S. **2008**. Thematic Review Series: Sphingolipids. Cross-talk at the crossroads of sphingosine-1-phosphate, growth factors, and cytokine signaling. *The Journal of Lipid Research* **49**: 1388–1394.

Ledon-Rettig CC, Pfennig DW, Chunco AJ, Dworkin I. **2014**. Cryptic Genetic Variation in Natural Populations: A Predictive Framework. *Integrative and Comparative Biology* **54**: 783–793.

Lee CE. **2002**. Evolutionary genetics of invasive species. *Trends in Ecology & Evolution* **17**: 386–391.

Lee CE, Gelembiuk GW. **2008**. Evolutionary origins of invasive populations. *Evolutionary Applications* **1**: 427–448.

Lefèvre C, Charles H, Vallier A, Delobel B, Farrell B, Heddi A. **2004**. Endosymbiont phylogenesis in the dryophthoridae weevils: evidence for bacterial replacement. *Molecular biology and evolution* **21**: 965–973.

Leichty AR, Pfennig DW, Jones CD, Pfennig KS. **2012**. Relaxed genetic constraint is ancestral to the evolution of phenotypic plasticity. *Integrative and Comparative Biology* **52**: 16–30.

Levin HL, Moran JV. **2011**. Dynamic interactions between transposable elements and their hosts. *Nature Reviews Genetics* **12**: 615–627.

Levy S, Hannenhalli S. **2002**. Identification of transcription factor binding sites in the human genome sequence. *Mammalian Genome* **13**: 510–514.

Li H. **2011**. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993.

Li J, Wu J, Begna Rundassa D, Song F, Zheng A, Fang Y. **2010**. Differential protein expression in honeybee (*Apis mellifera* L.) larvae: underlying caste differentiation. *PLoS ONE* **5**: e13455.

Linksvayer TA, Wade MJ. **2005**. The evolutionary origin and elaboration of sociality in the aculeate Hymenoptera: maternal effects, sib-social effects, and heterochrony. *The Quarterly review of biology* **80**: 317–336.

Linksvayer TA, Wade MJ. **2009**. Genes with social effects are expected to harbor more sequence variation within and between species. *Evolution* **63**: 1685–1696.

Liu H, Zhang X, Huang J, Chen J-Q, Tian D, Hurst LD, Yang S. **2014a**. Causes and consequences of crossing-over evidenced via a high-resolution recombinational landscape of the honey bee. *Genome Biology* **16**: 15–15.

Liu Y, Zhou J, White KP. **2014b**. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* **30**: 301–304.

Liu Y-L, Lu W-C, Brummel TJ, Yuh C-H, Lin P-T, Kao T-Y, Li F-Y, Liao P-C, Benzer S, Wang H-D. **2009**. Reduced expression of *alpha-1,2-mannosidase I* extends lifespan in *Drosophila melanogaster* and *Caenorhabditis elegans*. *Aging Cell* **8**: 370–379.

Loewe L, Hill WG. **2010**. The population genetics of mutations: good, bad and indifferent. *Philosophical Transactions*

of the Royal Society B: Biological Sciences **365**: 1153–1167.

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

Lowe TM, Eddy SR. **1997**. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research* **25**: 0955–0964.

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 DW, 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**: 18.

Lynch M. **2007**. *The Origins of Genome Architecture*. Sunderland, MA: Sinauer Associates.

Lynch M. **2010**. Evolution of the mutation rate. *Trends in Genetics* **26**: 345–352.

Macgown J. **2012**. First record of the tramp ant *Cardiocondyla obscurior* (Hymenoptera: Formicidae) for Mississippi. *Midsouth Entomologist* **5**: 9–10.

Madlung A, Comai L. **2004**. The effect of stress on genome regulation and structure. *Annals of Botany* **94**: 481–495.

Mahajan-Miklos S, Cooley L. **1994**. Intercellular cytoplasm transport during *Drosophila* oogenesis. *Developmental Biology* **165**: 336–351.

Malboeuf CM, Yang X, Charlebois P, Qu J, Berlin AM, Casali M, Pesko KN, Boutwell CL, DeVincenzo JP, Ebel GD, Allen TM, Zody MC, Henn MR, Levin JZ. **2013**. Complete viral RNA genome sequencing of ultra-low copy samples by sequence-independent amplification. *Nucleic acids research* **41**: e13.

Mank JE, Hultin-Rosenberg L, Zwahlen M, Ellegren H. **2008**. Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. *The American Naturalist* **171**: 35–43.

Manzano-Marín A, Latorre A. **2014**. Settling down: the genome of *Serratia symbiotica* from the aphid *Cinara tujafilina* zooms in on the process of accommodation to a cooperative intracellular life. *Genome Biology and Evolution* **6**: 1683–1698.

Martin M. **2011**. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**: pp. 10–12.

Mattick JS. **2006**. Non-coding RNA. *Human Molecular Genetics* **15**: R17–R29.

Maynard Smith J, Szathmáry E. **1997**. *The major transitions in evolution*. New York: Oxford University Press.

McClintock B. **1948**. Mutable loci in maize. *Carnegie Institution of Washington Yearbook* **47**: 155–169.

McCutcheon JP, Moran NA. **2011**. Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology* **10**: 13–26.

McDonald JH, Kreitman M. **1991**. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.

McNulty SN, Foster JM, Mitreva M, Dunning Hotopp JC, Martin J, Fischer K, Wu B, Davis PJ, Kumar S, Brattig NW, Slatko BE, Weil GJ, Fischer PU. **2010**. Endosymbiont DNA in endobacteria-free filarial nematodes indicates ancient horizontal genetic transfer (L Rénia, Ed.). *PLoS ONE* **5**: e11029.

Medvedev P, Stanciu M, Brudno M. **2009**. Computational methods for discovering structural variation with next-generation sequencing. *Nature methods* **6**: S13–S20.

Menuz V, Howell KS, Gentina S, Epstein S, Riezman I, Fornallaz-Mulhauser M, Hengartner MO, Gomez M, Riezman H, Martinou J-C. **2009**. Protection of *C. elegans* from anoxia by HYL-2 ceramide synthase. *Science* **324**: 381–384.

Merrill AH, Nikolova-Karakashian M, Schmelz EM, Morgan ET, Stewart J. **1999**. Regulation of cytochrome P450 expression by sphingolipids. *Chemistry and Physics of Lipids* **102**: 131–139.

Mi S, Lee X, Li XP, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, Keith JC, McCoy JM. **2000**. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**: 785–789.

Michalkova V, Benoit JB, Weiss BL, Attardo GM, Aksoy S. **2014**. Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies. *Applied and Environmental Microbiology* **80**: 5844–5853.

Mikheyev AS, Linksvayer TA. **2015**. Genes associated with ant social behavior show distinct transcriptional and evolutionary patterns. *eLife* **4**: e04775–e04775.

Millar NS, Denholm I. **2007**. Nicotinic acetylcholine receptors: targets for commercially important insecticides. *Invertebrate Neuroscience* **7**: 53–66.

Moczek AP. **2008**. On the origins of novelty in development and evolution. *BioEssays* **30**: 432–447.

Moczek AP, Sultan S, Foster S, Ledon-Rettig C, Dworkin I, Nijhout HF, Abouheif E, Pfennig DW. **2011**. The role of developmental plasticity in evolutionary innovation. *Proceedings of the Royal Society B: Biological Sciences* **278**: 2705–2713.

Moran NA, Dunbar HE. **2006**. Sexual acquisition of beneficial symbionts in aphids. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 12803–12806.

Moran NA, Russell JA, Koga R, Fukatsu T. **2005**. Evolutionary relationships of three new species of *Enterobacteriaceae* living as symbionts of aphids and other insects. *Applied and Environmental Microbiology* **71**: 3302–3310.

Mosbech M-B, Kruse R, Harvald EB, Olsen ASB, Gallego SF, Hannibal-Bach HK, Ejsing CS, Færgeman NJ. **2013**. Functional loss of two ceramide synthases elicits autophagy-dependent lifespan extension in *C. elegans* (J Sadoshima, Ed.). *PLoS ONE* **8**: e70087.

Moya A, Peretó J, Gil R, Latorre A. **2008**. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nature Reviews Genetics* **9**: 218–229.

Murat S, Hopfen C, McGregor AP. **2010**. The function and evolution of *Wnt* genes in arthropods. *Arthropod Structure & Development*: 1–7.

Murray RG, Stackebrandt E. **1995**. Taxonomic note: implementation of the provisional status *Candidatus* for incompletely described procaryotes. *International journal of systematic bacteriology* **45**: 186–187.

Mutti NS, Dolezal AG, Wolschin F, Mutti JS, Gill KS, Amdam GV. **2011**. IRS and TOR nutrient-signaling pathways act via juvenile hormone to influence honey bee caste fate. *The Journal of experimental biology* **214**: 3977–3984.

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*. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 5477–5482.

Nawrocki EP, Eddy SR. **2013**. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**: 2933–2935.

Nei M, Maruyama T, Chakraborty R. **1975**. The bottleneck effect and genetic variability in populations. *Evolution*: 1–10.

Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK, Maurano MT, Humbert R, Rynes E, Wang H, Vong S, Lee K, Bates D, Diegel M, Roach V, Dunn D, Neri J, Schafer A, Hansen RS, Kutyavin T, Giste E, Weaver M, Canfield T, Sabo P, Zhang M, Balasundaram G, Byron R, MacCoss MJ, Akey JM, Bender MA, Groudine M, Kaul R, Stamatoyannopoulos JA. **2012**. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**: 83–90.

Nijhout HF. **2003**. Development and evolution of adaptive polyphenisms. *Evolution & development* **5**: 9–18.

Nowak MA, Tarnita CE, Wilson EO. **2010**. The evolution of eusociality. *Nature* **466**: 1057–1062.

Nuzhdin SV, Pasyukova EG, Mackay TF. **1997**. Accumulation of transposable elements in laboratory lines of *Drosophila melanogaster*. *Genetica* **100**: 167–175.

Nygaard S, Zhang G, Schiøtt M, Li C, Wurm Y, Hu H, Zhou J, Ji L, Qiu F, Rasmussen M, Pan H, Hauser F, Krogh A, Grimmelikhuijzen CJP, Wang J, Boomsma JJ. **2011**. The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome research* **21**: 1339–1348.

O'Reilly C, Turner PD, Smith-Keary PF, McConnell DJ. **1984**. Molecular cloning of genes involved in purine biosynthetic and salvage pathways of *Salmonella typhimurium*. *Molecular and General Genetics MGG* **196**: 152–157.

Oettler J, Suefuji M, Heinze J. **2010**. The evolution of alternative reproductive tactics in male *Cardiocondyla* ants.

Evolution **64**: 3310–3317.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. **2000**. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**: 27–30.

Ohno S. **1970**. Duplication for the Sake of Producing More of the Same. *Evolution by gene duplication*. Springer Berlin, Heidelberg and New York, 59–65.

Ohno S. **1972**. So much "junk" DNA in our genome. *Brookhaven symposia in biology* **23**: 366–370.

Oliver KM, Russell JA, Moran NA, Hunter MS. **2003**. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 1803–1807.

Oliver KR, Greene WK. **2009**. Transposable elements: powerful facilitators of evolution. *BioEssays* **31**: 703–714.

Ometto L, Shoemaker D, ROSS KG, Keller L. **2011**. Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species. *Molecular biology and evolution* **28**: 1381–1392.

Oxley PR, Ji L, Fetter-Pruneda I, McKenzie SK, Li C, Hu H, Zhang G, Kronauer DJC. **2014**. The genome of the clonal raider ant *Cerapachys biroi*. *Current biology* **24**: 451–458.

Pace JK, Feschotte C. **2007**. The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. *Genome research* **17**: 422–432.

Page RE, Peng C. **2001**. Aging and development in social insects with emphasis on the honey bee, *Apis mellifera* L. *Experimental Gerontology* **36**: 695–711.

Palade GE. **1955**. A small particulate component of the cytoplasm. *The Journal of biophysical and biochemical cytology* **1**: 59–68.

Pardue M-L, DeBaryshe PG. **2003**. Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annual Review of Genetics* **37**: 485–511.

Patel A, Fondrk MK, Kaftanoglu O, Emore C, Hunt G, Frederick K, Amdam GV. **2007**. The making of a queen: TOR pathway is a key player in diphenic caste development. (P Michalak, Ed.). *PLoS ONE* **2**: e509.

Patino-Navarrete R, Piulachs M-D, Bellés 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. *Biology Letters* **10**: 20140407.

Penick CA, Liebig J. **2012**. Regulation of queen development through worker aggression in a predatory ant. *Behavioral Ecology* **23**: 992–998.

Penick CA, Prager SS, Liebig J. **2012**. Juvenile hormone induces queen development in late-stage larvae of the ant *Harpegnathos saltator*. *Journal of Insect Physiology* **58**: 1643–1649.

Pepperl J, Reim G, Lüthi U, Kaech A, Hausmann G, Basler K. **2013**. Sphingolipid depletion impairs endocytic traffic and inhibits Wingless signaling. *Mechanisms of Development* **130**: 493–505.

Petersen TN, Brunak S, Heijne von G, Nielsen H. **2011**. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**: 785–786.

Pérez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A. **2006**. A small microbial genome: the end of a long symbiotic relationship? *Science* **314**: 312–313.

Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD, Moczek AP. **2010**. Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution* **25**: 459–467.

Phan VH, Herr DR, Panton D, Fyrst H, Saba JD, Harris GL. **2007**. Disruption of sphingolipid metabolism elicits apoptosis-associated reproductive defects in *Drosophila*. *Developmental Biology* **309**: 329–341.

Pigliucci M, Murren CJ, Schlichting CD. **2006**. Phenotypic plasticity and evolution by genetic assimilation. *The Journal of experimental biology* **209**: 2362–2367.

Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ. **2008**. Adaptive evolution in invasive species. *Trends in Plant Science* **13**: 288–294.

Price TD, Qvarnstrom A, Irwin DE. **2003**. The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society B: Biological Sciences* **270**: 1433–1440.

Purandare SR, Bickel RD, Jaquiery J, Rispe C, Brisson JA. **2014**. Accelerated evolution of morph-biased genes in

pea aphids. *Molecular biology and evolution* **31**: 2073–2083.

Rajakumar R, San Mauro D, Dijkstra MB, Huang MH, Wheeler DE, Hiou-Tim F, Khila A, Cournoyea M, Abouheif E. **2012**. Ancestral developmental potential facilitates parallel evolution in ants. *Science* **335**: 79–82.

Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, Mason CE, Socci ND, Betel D. **2013**. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology* **14**: R95.

Ratzka C, Gross R, Feldhaar H. **2013**. Gene expression analysis of the endosymbiont-bearing midgut tissue during ontogeny of the carpenter ant *Camponotus floridanus*. *Journal of Insect Physiology* **59**: 611–623.

Ravindran S. **2012**. Barbara McClintock and the discovery of jumping genes. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 20198–20199.

Rebollo R, Horard B, Begeot F, Delattre M, Gilson E, Viera C. **2012**. A snapshot of histone modifications within transposable elements in *Drosophila* wild type strains. *PLoS ONE* **7**: e44253.

Reuter M, Pedersen JS, Keller L. **2005**. Loss of *Wolbachia* infection during colonisation in the invasive Argentine ant *Linepithema humile*. *Heredity* **94**: 364–369.

Reznick DN, Ghalambor CK. **2001**. The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* **112-113**: 183–198.

Rizzon C, Marais G, Gouy M, Biemont C. **2002**. Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome research* **12**: 400–407.

Romiguier J, Lourenco J, Gayral P, Faivre N, Weinert LA, 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. *Journal of Evolutionary Biology* **27**: 593–603.

Rostant WG, Wedell N, Hosken DJ. **2012**. Transposable elements and insecticide resistance.Goodwin SF, Friedmann T, Dunlap JC eds. *Advances in Genetics*. Burlington: Elsevier, 169–201.

Ruden DM, Cingolani PE, Sen A, Qu W, Wang L, Senut M-C, Garfinkel MD, Sollars VE, Lu X. **2015**. Epigenetics as an answer to Darwin's 'special difficulty,' Part 2: natural selection of metastable epialleles in honeybee castes. *Frontiers in Genetics* **6**: 60.

Ruppell O, Heinze J. **1999**. Alternative reproductive tactics in females: the case of size polymorphism in winged ant queens. *Insectes Sociaux* **46**: 6–17.

Russell JA, Moran NA. **2006**. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. *Proceedings of the Royal Society B: Biological Sciences* **273**: 603–610.

Russell JA, Moreau CS, Goldman-Huertas B, Fujiwara M, Lohman DJ, Pierce NE. **2009**. Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 21236–21241.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B. **2000**. Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944–945.

Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J, With KA, Baughman S, Cabin RJ, Cohen JE, Ellstrand NC, McCauley DE, O'Neil P, Parker IM, Thompson JN, Weller SG. **2001**. The population biology of invasive species. *Annual review of ecology and systematics* **32**: 305–332.

Sambrook J, Russell DW. **2001**. *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.

Sasamura T, Matsuno K, Fortini ME. **2013**. Disruption of *Drosophila melanogaster* lipid metabolism genes causes tissue overgrowth associated with altered developmental signaling. (SJ Bray, Ed.). *PLoS genetics* **9**: e1003917.

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. *International Journal of Systematic and Evolutionary Microbiology* **50**: 1877–1886.

Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, Sohngen C, Stelzer M, Thiele J, Schomburg D. **2010**. BRENDA, the enzyme information system in 2011. *Nucleic acids research* **39**: D670–D676.

Schlichting CD. **2004**. The Role of Phenotypic Plasticity in Diversification.DeWitt TJ, Scheiner SM eds. *Phenotypic Plasticity: Functional and Conceptual Approaches*. New York: Oxford University Press.

Schlichting CD. **2008**. Hidden reaction norms, cryptic genetic variation, and evolvability. *Annals of the New York Academy of Sciences* **1133**: 187–203.

Schlichting CD, Pigliucci M. **1995**. Gene regulation, quantitative genetics and the evolution of reaction norms. *Evolutionary Ecology* **9**: 154–168.

Schlichting CD, Pigliucci M. **1998**. *Phenotypic Evolution: A Reaction Norm Perspective*. Sunderland, MA: Sinauer Associates, Incorporated.

Schlichting CD, Wund MA. **2014**. Phenotypic plasticity and epigenetic marking: an assessment of evidence for genetic accommodation. *Evolution* **68**: 656–672.

Schmidt AM, Linksvayer TA, Boomsma JJ, Pedersen JS. **2010**. Queen–worker caste ratio depends on colony size in the pharaoh ant (*Monomorium pharaonis*). *Insectes Sociaux* **58**: 139–144.

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

Schrader L, Simola DF, Heinze J, Oettler J. **2015**. Sphingolipids, transcription factors, and conserved toolkit genes: developmental plasticity in the ant *Cardiocondyla obscurior*. *Molecular biology and evolution*.

Schrempf A, Heinze J. **2006**. Proximate mechanisms of male morph determination in the ant *Cardiocondyla obscurior*. *Evolution & development* **8**: 266–272.

Schrempf A, Cremer S, Heinze J. **2011**. Social influence on age and reproduction: reduced lifespan and fecundity in multi-queen ant colonies. *Journal of Evolutionary Biology* **24**: 1455–1461.

Schrempf A, Heinze J, Cremer S. **2005**. Sexual cooperation: mating increases longevity in ant queens. *Current biology* **15**: 267–270.

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. stambuloffii, C. wroughtonii, C. emeryi,* and *C. minutior* species groups. *Annalen des Naturhistorischen Museums in Wien. Serie B für Botanik und Zoologie*: 203–338.

Shabalina SA, Koonin EV. **2008**. Origins and evolution of eukaryotic RNA interference. *Trends in Ecology & Evolution* **23**: 578–587.

Shee C, Gibson JL, Rosenberg SM. **2012**. Two mechanisms produce mutation hotspots at DNA breaks in *Escherichia coli*. *Cell reports* **2**: 714–721.

Shi YY, Yan WY, Huang ZY, Wang ZL, Wu XB, Zeng ZJ. **2012**. Genomewide analysis indicates that queen larvae have lower methylation levels in the honey bee (*Apis mellifera*). *Naturwissenschaften* **100**: 193–197.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. **2000**. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**: 81–86.

Shimada M, Ishii Y, Shibao H. **2009**. Rapid adaptation: a new dimension for evolutionary perspectives in ecology. *Population Ecology* **52**: 5–14.

Simola DF, Wissler L, Donahue G, Waterhouse RM, Helmkampf M, Roux J, Nygaard S, Glastad KM, Hagen DE, Viljakainen L, Reese JT, Hunt BG, Graur D, Elhaik E, Kriventseva EV, Wen J, Parker BJ, Cash E, Privman E, Childers CP, Muñoz-Torres MC, Boomsma JJ, Bornberg-Bauer E, Currie CR, Elsik CG, Suen G, Goodisman MAD, Keller L, Liebig J, Rawls A, Reinberg D, Smith CD, Smith CR, Tsutsui N, Wurm Y, Zdobnov EM, Berger SL, Gadau J. **2013a**. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome research* **23**: 1235–1247.

Simola DF, Ye C, Mutti NS, Dolezal K, Bonasio R, Liebig J, Reinberg D, Berger SL. **2013b**. A chromatin link to caste identity in the carpenter ant *Camponotus floridanus*. *Genome research* **23**: 486–496.

Singh RS, Xu J, Kulathinal RJ. **2012**. *Rapidly Evolving Genes and Genetic Systems*. Oxford: Oxford University Press.

Slotkin RK, Martienssen R. **2007**. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* **8**: 272–285.

Slotkin T, Seidler F. **2009**. Transcriptional profiles reveal similarities and differences in the effects of developmental neurotoxicants on differentiation into neurotransmitter phenotypes in PC12 cells. *Brain Research Bulletin* **78**:

211–225.

Smith CD, Zimin A, Holt C, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E, Elsik CG, Favé M-J, Fernandes V, Gadau J, Gibson JD, Graur D, Grubbs KJ, Hagen DE, Helmkampf M, Holley J-A, Hu H, Viniegra ASI, Johnson BR, Johnson RM, Khila A, Kim JW, Laird J, Mathis KA, Moeller JA, Muñoz-Torres MC, Murphy MC, Nakamura R, Nigam S, Overson RP, Placek JE, Rajakumar R, Reese JT, Robertson HM, Smith CR, Suarez AV, Suen G, Suhr EL, Tao S, Torres CW, van Wilgenburg E, Viljakainen L, Walden KKO, Wild AL, Yandell M, Yorke JA, Tsutsui ND. **2011a**. Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proceedings of the National Academy of Sciences of the United States of America* **108**: 5673–5678.

Smith CR, Smith CD, Robertson HM, Helmkampf M, Zimin A, Yandell M, Holt C, Hu H, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E, Elsik CG, Favé M-J, Fernandes V, Gibson JD, Graur D, Gronenberg W, Grubbs KJ, Hagen DE, Viniegra ASI, Johnson BR, Johnson RM, Khila A, Kim JW, Mathis KA, Muñoz-Torres MC, Murphy MC, Mustard JA, Nakamura R, Niehuis O, Nigam S, Overson RP, Placek JE, Rajakumar R, Reese JT, Suen G, Tao S, Torres CW, Tsutsui ND, Viljakainen L, Wolschin F, Gadau J. **2011b**. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 5667–5672.

Smith-Gill SJ. **1983**. Developmental plasticity - developmental conversion versus phenotypic modulation. *American Zoologist* **23**: 47–55.

Smyth GK, Ritchie M, Thorne N, Wettenhall J, Shi W, Hu Y. **2002**. Limma: linear models for microarray data. *Bioconductor*.

Snell-Rood EC, Van Dyken JD, Cruickshank T, Wade MJ, Moczek AP. **2010**. Toward a population genetic framework of developmental evolution: the costs, limits, and consequences of phenotypic plasticity. *BioEssays* **32**: 71–81.

Spiegel S, Merrill AH. **1996**. Sphingolipid metabolism and cell growth regulation. *Faseb Journal* **10**: 1388–1397.

Stabe HA. **1930**. The rate of growth of worker, drone and queen larvae of the honeybee, *Apis mellifera* Linn. *Journal of economic entomology* **23**: 447–453.

Staden R, Beal KF, Bonfield JK. **1999**. The Staden package, 1998.Misener S, Krawetz SA eds. *Bioinformatics methods and protocols*. Totowa, NJ: Humana Press, 115–130.

Stapley J, Santure AW, Dennis SR. **2015**. Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Molecular Ecology*.

Stoll S, Feldhaar H, Fraunholz MJ, Gross R. **2010**. Bacteriocyte dynamics during development of a holometabolous insect, the carpenter ant *Camponotus floridanus*. *BMC Microbiology* **10**: 308.

Stuart RJ, Francoeur A, Loiselle R. **1987**. Lethal fighting among dimorphic males of the ant, *Cardiocondyla wroughtonii*. *Naturwissenschaften* **74**: 548–549.

Suarez AV, McGlynn TP, Tsutsui ND. **2010**. Biogeographic and Taxonomic Patterns of Introduced Ants. *Ant Ecology*. New York: Oxford University Press, USA, 233–244.

Suefuji M, Cremer S, Oettler J, Heinze J. **2008**. Queen number influences the timing of the sexual production in colonies of *Cardiocondyla* ants. *Biology Letters* **4**: 670–673.

Suen G, Teiling C, Li L, Holt C, Abouheif E, Bornberg-Bauer E, Bouffard P, Caldera EJ, Cash E, Cavanaugh A, Denas O, Elhaik E, Favé M-J, Gadau J, Gibson JD, Graur D, Grubbs KJ, Hagen DE, Harkins TT, Helmkampf M, Hu H, Johnson BR, Kim J, Marsh SE, Moeller JA, Muñoz-Torres MC, Murphy MC, Naughton MC, Nigam S, Overson R, Rajakumar R, Reese JT, Scott JJ, Smith CR, Tao S, Tsutsui ND, Viljakainen L, Wissler L, Yandell MD, Zimmer F, Taylor J, Slater SC, Clifton SW, Warren WC, Elsik CG, Smith CD, Weinstock GM, Gerardo NM, Currie CR. **2011**. The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle. *PLoS genetics* **7**: e1002007.

Sun Z, Asmann YW, Nair A, Zhang Y, Wang L, Kalari KR, Bhagwate AV, Baker TR, Carr JM, Kocher J-PA, Perez EA, Thompson EA. **2013**. Impact of library preparation on downstream analysis and interpretation of RNA-Seq data: comparison between Illumina PolyA and NuGEN Ovation protocol. (K Wang, Ed.). *PLoS ONE* **8**: e71745.

Suzek BE, Ermolaeva MD, Schreiber M, Salzberg SL. **2001**. A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics* **17**: 1123–1130.

Tay J-W, Neoh K-B, Lee C-Y. **2014**. The roles of the queen, brood, and worker castes in the colony growth dynamics

of the pharaoh ant *Monomorium pharaonis* (Hymenoptera: Formicidae). *Myrmecological News* **20**: 87–94.

Tåquist H, Cui Y, Ardell DH. **2007**. TFAM 1.0: an online tRNA function classifier. *Nucleic acids research* **35**: W350–W353.

Tedesco P, Jiang J, Wang J, Jazwinski SM, Johnson TE. **2008**. Genetic analysis of *hyl-1*, the *C. elegans* homolog of *LAG1/LASS1*. *AGE* **30**: 43–52.

Teseo S, Kronauer DJC, Jaisson P, Châline N. **2013**. Enforcement of reproductive synchrony via policing in a clonal ant. *Current biology* **23**: 328–332.

Toft C, Andersson SGE. **2010**. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nature Reviews Genetics* **11**: 465–475.

Treangen TJ, Salzberg SL. **2012**. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics* **13**: 36–46.

Tsuji N, Yamauchi K, Yamamura N. **1994**. A mathematical model for wing dimorphism in male *Cardiocondyla* ants. *Journal of Ethology* **12**: 19–24.

Tsutsui ND, Suarez AV, Grosberg RK. **2003**. Genetic diversity, asymmetrical aggression, and recognition in a widespread invasive species. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 1078–1083.

Tsutsui ND, Suarez AV, Spagna JC, Johnston JS. **2008**. The evolution of genome size in ants. *BMC Evolutionary Biology* **8**: 64–73.

Ugelvig LV, Kronauer DJC, Schrempf A, Heinze J, Cremer S. **2010**. Rapid anti-pathogen response in ant societies relies on high genetic diversity. *Proceedings of the Royal Society B: Biological Sciences* **277**: 2821–2828.

Ungerer MC, Strakosh SC, Zhen Y. **2006**. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Current biology* **16**: R872–R873.

Van de Peer Y, Maere S, Meyer A. **2009**. The evolutionary significance of ancient genome duplications. *Nature Publishing Group* **10**: 725–732.

Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, Lu P, Szafron D, Greiner R, Wishart DS. **2005**. BASys: a web server for automated bacterial genome annotation. *Nucleic acids research* **33**: W455–W459.

Van Dyken JD, Wade MJ. **2010**. The genetic signature of conditional expression. *Genetics* **184**: 557–570.

van Zon A, Mossink MH, Scheper RJ, Sonneveld P, Wiemer EAC. **2003**. The vault complex. *Cellular and Molecular Life Sciences* **60**: 1828–1837.

van Zweden JS, van Zweden JS, D'Ettorre P, dEttorre P. **2010**. Nestmate recognition in social insects and the role of hydrocarbons.Blomquist GJ, Bagnères A-G eds. *Insect Hydrocarbons: Biology, Biochemistry, and Chemical Ecology*. Insect Hydrocarbons: Biology, Biochemistry, and Chemical Ecology, 222–243.

Vandenbussche M, Zethof J, Gerats T. **2013**. Transposon Display: A Versatile Method for Transposon Tagging.Peterson T ed. *Plant Transposable Elements: Methods and Protocols*. New York: Springer.

Veerla S, Ringnér M, Höglund M. **2010**. Genome-wide transcription factor binding site/promoter databases for the analysis of gene sets and co-occurrence of transcription factor binding motifs. *BMC Genomics* **11**: 145.

Veluchamy A, Lin X, Maumus F, Rivarola M, Bhavsar J, Creasy T, O'Brien K, Sengamalay NA, Tallon LJ, Smith AD, Rayko E, Ahmed I, Le Crom S, Farrant GK, Sgro J-Y, Olson SA, Bondurant SS, Allen AE, Allen A, Rabinowicz PD, Sussman MR, Bowler C, Tirichine L. **2013**. Insights into the role of DNA methylation in diatoms by genome-wide profiling in *Phaeodactylum tricornutum*. *Nature Communications* **4**: 2091.

Via S, Gomulkiewicz R, De Jong G, Scheiner SM, Schlichting CD, Van Tienderen PH. **1995**. Adaptive phenotypic plasticity: consensus and controversy. *Trends in Ecology & Evolution* **10**: 212–217.

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 Microbiology* **12**: 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. *Current biology* **24**: 2267–2273.

Volff J-N. **2006**. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays* **28**: 913–922.

Vosshall LB, Wong AM, Axel R. **2000**. An olfactory sensory map in the fly brain. *Cell* **102**: 147–159.

Walls SM, Attle SJ, Brulte GB, Walls ML, Finley KD, Chatfield DA, Herr DR, Harris GL. **2013**. Identification of sphingolipid metabolites that induce obesity via misregulation of appetite, caloric intake and fat storage in *Drosophila*. *PLoS genetics* **9**: e1003970.

Walter P, Blobel G. **1982**. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* **299**: 691–698.

Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BG, Dong X, Kundaje A, Cheng Y, Rando OJ, Birney E, Myers RM, Noble WS, Snyder M, Weng Z. **2012**. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome research* **22**: 1798–1812.

Wang Jingwen, Weiss BL, Aksoy S. **2013a**. Tsetse fly microbiota: form and function. *Frontiers in cellular and infection microbiology* **3**: 1–6.

Wang John, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang Y-C, Shoemaker D, Keller L. **2013b**. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**: 664–668.

Wasserman WW, Sandelin A. **2004**. Applied bioinformatics for the identification of regulatory elements. *Nature Reviews Genetics* **5**: 276–287.

Watanabe K, Yukuhiro F, Matsuura Y, Fukatsu T, Noda H. **2014**. Intrasperm vertical symbiont transmission. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 7433–7437.

Weinstock GM, Robinson GE, Gibbs RA, Weinstock GM, Weinstock GM, Robinson GE, Worley KC, Evans JD, Maleszka R, Robertson HM, Weaver DB, Beye M, Bork P, Elsik CG, Evans JD, Hartfelder K, Hunt GJ, Robertson HM, Robinson GE, Maleszka R, Weinstock GM, Worley KC, Zdobnov EM, Hartfelder K, Amdam GV, Bitondi MMG, Collins AM, Cristino AS, Evans JD, Michael H, Lattorff G, Lobo CH, Moritz RFA, Nunes FMF, Page RE, Simões ZLP, Wheeler D, Carninci P, Fukuda S, Hayashizaki Y, Kai C, Kawai J, Sakazume N, Sasaki D, Tagami M, Maleszka R, Amdam GV, Albert S, Baggerman G, Beggs KT, Bloch G, Cazzamali G, Cohen M, Drapeau MD, Eisenhardt D, Emore C, Ewing MA, Fahrbach SE, Foret S, Grimmelikhuijzen CJP, Hauser F, Hummon AB, Hunt GJ, Huybrechts J, Jones AK, Kadowaki T, Kaplan N, Kucharski R, Leboulle G, Linial M, Littleton JT, Mercer AR, Page RE, Robertson HM, Robinson GE, Richmond TA, RodriguezZas SL, Rubin EB, Sattelle DB, Schlipalius D, Schoofs L, Shemesh Y, Sweedler JV, Velarde R, Verleyen P, Vierstraete E, Williamson MR, Beye M, Ament SA, Brown SJ, Corona M, Dearden PK, Dunn WA, Elekonich MM, Elsik CG, Foret S, Fujiyuki T, Gattermeier I, Gempe T, Hasselmann M, Kadowaki T, Kage E, Kamikouchi A, Kubo T, Kucharski R, Kunieda T, Lorenzen M, Maleszka R, Milshina NV, Morioka M, Ohashi K, Overbeek R, Page RE, Robertson HM, Robinson GE, Ross CA, Schioett M, Shippy T, Takeuchi H, Toth AL, Willis JH, Wilson MJ, Robertson HM, Zdobnov EM, Bork P, Elsik CG, Gordon KHJ, Letunic I, Hackett K, Peterson J, Felsenfeld A, Guyer M, Solignac M, Agarwala R, Cornuet JM, Elsik CG, Emore C, Hunt GJ, Monnerot M, Mougel F, Reese JT, Schlipalius D, Vautrin D, Weaver DB, Gillespie JJ, Cannone JJ, Gutell RR, Johnston JS, Elsik CG, Cazzamali G, Eisen MB, Grimmelikhuijzen CJP, Hauser F, Hummon AB, Iyer VN, Iyer V, Kosarev P, Mackey AJ, Maleszka R, Reese JT, Richmond TA, Robertson HM, Solovyev V, Souvorov A, Sweedler JV, Weinstock GM, Williamson MR, Zdobnov EM, Evans JD, Aronstein KA, Bilikova K, Chen YP, Clark AG, Decanini LI, Gelbart WM, Hetru C, Hultmark D, Imler J-L, Jiang H, Kanost M, Kimura K, Lazzaro BP, Lopez DL, Simuth J, Thompson GJ, Zou Z, De Jong P, Sodergren E, Csürös M, Milosavljevic A, Johnston JS, Osoegawa K, Richards S, Shu C-L, Weinstock GM, Elsik CG, Duret L, Elhaik E, Graur D, Reese JT, Robertson HM, Robertson HM, Elsik CG, Maleszka R, Weaver DB, Amdam GV, Anzola JM, Campbell KS, Childs KL, Collinge D, Crosby MA, Dickens CM, Elsik CG, Gordon KHJ, Gramates LS, Grozinger CM, Jones PL, Jorda M, Ling X, Matthews BB, Miller J, Milshina NV, Mizzen C, Peinado MA, Reese JT, Reid JG, Robertson HM, Robinson GE, Russo SM, Schroeder AJ, St Pierre SE, Wang Y, Zhou P, Robertson HM, Agarwala R, Elsik CG, Milshina NV, Reese JT, Weaver DB, Worley KC, Childs KL, Dickens CM, Elsik CG, Gelbart WM, Jiang H, Kitts P, Milshina NV, Reese JT, Ruef B, Russo SM, Venkatraman A, Weinstock GM, Zhang L, Zhou P, Johnston JS, Aquino-Perez G, Cornuet JM, Monnerot M, Solignac M, Vautrin D, Whitfield CW, Behura SK, Berlocher SH, Clark AG, Gibbs RA, Johnston JS, Sheppard WS, Smith DR, Suarez AV, Tsutsui ND, Weaver DB, Wei X, Wheeler D, Weinstock GM, Worley KC, Havlak P, Li B, Liu Y, Sodergren E, Zhang L, Beye M, Hasselmann M, Jolivet A, Lee S, Nazareth LV, Pu L-L, Thorn R, Weinstock GM, Stolc V, Robinson GE, Maleszka R, Newman T, Samanta M, Tongprasit WA, Aronstein KA, Claudianos C, Berenbaum MR, Biswas S, de Graaf DC, Feyereisen R, Johnson RM, Oakeshott JG, Ranson H, Schuler MA, Muzny D, Gibbs RA,

Weinstock GM, Chacko J, Davis C, Dinh H, Gill R, Hernandez J, Hines S, Hume J, Jackson L, Kovar C, Lewis L, Miner G, Morgan M, Nazareth LV, Nguyen N, Okwuonu G, Paul H, Richards S, Santibanez J, Savery G, Sodergren E, Svatek A, Villasana D, Wright R. **2006**. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**: 931–949.

Wenseleers T, Sundström L, Billen J. **2002**. Deleterious Wolbachia in the ant Formica truncorum. *Proceedings of the Royal Society B: Biological Sciences* **269**: 623–629.

Werren JH, Baldo L, Clark ME. **2008**. *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology* **6**: 741–751.

Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, The Nasonia Genome Working Group, Beukeboom LW, Desplan C, Elsik CG, Grimmelikhuijzen CJP, Kitts P, Lynch JA, Murphy T, Oliveira DCSG, Smith CD, Zande LVD, Worley KC, Zdobnov EM, Aerts M, Albert S, Anaya VH, Anzola JM, Barchuk AR, Behura SK, Bera AN, Berenbaum MR, Bertossa RC, Bitondi MMG, Bordenstein SR, Bork P, Bornberg-Bauer E, Brunain M, Cazzamali G, Chaboub L, Chacko J, Chavez D, Childers CP, Choi JH, Clark ME, Claudianos C, Clinton RA, Cree AG, Cristino AS, Dang PM, Darby AC, de Graaf DC, Devreese B, Dinh HH, Edwards R, Elango N, Elhaik E, Ermolaeva O, Evans JD, Foret S, Fowler GR, Gerlach D, Gibson JD, Gilbert DG, Graur D, Grunder S, Hagen DE, Han Y, Hauser F, Hultmark D, Hunter HC, Hurst GDD, Jhangian SN, Jiang H, Johnson RM, Jones AK, Junier T, Kadowaki T, Kamping A, Kapustin Y, Kechavarzi B, Kim J, Kim J, Kiryutin B, Koevoets T, Kovar CL, Kriventseva EV, Kucharski R, Lee H, Lee SL, Lees K, Lewis LR, Loehlin DW, Logsdon JM, Lopez JA, Lozado RJ, Maglott D, MALESZKA R, Mayampurath A, Mazur DJ, McClure MA, Moore AD, Morgan MB, Muller J, Munoz-Torres MC, Muzny DM, Nazareth LV, Neupert S, Nguyen NB, Nunes FMF, Oakeshott JG, Okwuonu GO, Pannebakker BA, Pejaver VR, Peng Z, Pratt SC, Predel R, Pu LL, Ranson H, Raychoudhury R, Rechtsteiner A, Reid JG, Riddle M, Romero-Severson J, Rosenberg M, Sackton TB, Sattelle DB, Schluns H, Schmitt T, Schneider M, Schüler A, Schurko AM, Shuker DM, Simoes ZLP, Sinha S, Smith Z, Souvorov A, Springauf A, Stafflinger E, Stage DE, Stanke M, Tanaka Y, Telschow A, Trent C, Vattathil S, Viljakainen L, Wanner KW, Waterhouse RM, Whitfield JB, Wilkes TE, Williamson M, Willis JH, Wolschin F, Wyder S, Yamada T, Yi SV, Zecher CN, Zhang L, Gibbs RA. **2010**. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**: 343–348.

West-Eberhard MJ. **1989**. Phenotypic plasticity and the origins of diversity. *Annual review of ecology and systematics*: 249–278.

West-Eberhard MJ. **2003**. *Developmental plasticity and evolution*. New York: Oxford University Press.

West-Eberhard MJ. **2005a**. Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 6543–6549.

West-Eberhard MJ. **2005b**. Phenotypic accommodation: adaptive innovation due to developmental plasticity. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **304B**: 610–618.

Wetterer JK, Kronauer DJ, Borowiec ML. **2012**. Worldwide spread of *Cerapachys biroi* (Hymenoptera: Formicidae: Cerapachyinae). *Myrmecological News* **17**: 1–4.

Wheeler D E. **1986**. Developmental and physiological determinants of caste in social Hymenoptera: evolutionary implications. *The American Naturalist*.

Wheeler David, Redding AJ, Werren JH. **2013**. Characterization of an ancient lepidopteran lateral gene transfer. *PLoS ONE* **8**: e59262.

Wheeler DE. **1991**. The developmental basis of worker caste polymorphism in ants. *The American Naturalist* **138**: 1218–1238.

White JA. **2011**. Caught in the act: Rapid, symbiont-driven evolution. *BioEssays* **33**: 823–829.

Whitman DW, Agrawal AA. **2009**. What is Phenotypic Plasticity and Why is it Important?Whitman DW, ananthakrishnan TN eds. Phenotypic Plasticity of Insects: Mechanisms and Consequences. *Phenotypic Plasticity of Insects: Mechanisms and Consequences*. Enfield, NH: CRC Press, 1–63.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH. **2007**. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* **8**: 973–982.

Will S, Delabie JHC, Heinze J, Ruther J, Oettler J. **2012**. Cuticular lipid profiles of fertile and non-fertile *Cardiocondyla* ant queens. *Journal of Insect Physiology* **58**: 1245–1249.

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. *Applied and Environmental Microbiology* **70**: 4096–4102.

Wolschin F, Mutti NS, Amdam GV. **2011**. Insulin receptor substrate influences female caste development in honeybees. *Biology Letters* **7**: 112–115.

Worby CA, Dixon JE. **2002**. Sorting out the cellular functions of sorting nexins. *Nature Reviews Molecular Cell Biology* **3**: 919–931.

Worning P, Jensen LJ, Hallin PF, Staerfeldt H-H, Ussery DW. **2006**. Origin of replication in circular prokaryotic chromosomes. *Environmental microbiology* **8**: 353–361.

Wray GA. **2007**. The evolutionary significance of *cis*-regulatory mutations. *Nature Reviews Genetics* **8**: 206–216.

Wright SI, Ness RW, Foxe JP, Barrett SCH. **2008**. Genomic consequences of outcrossing and selfing in plants. *International Journal of Plant Sciences* **169**: 105–118.

Wurm Y, Wang J, Riba-Grognuz O, Corona M, Nygaard S, Hunt BG, Ingram KK, Falquet L, Nipitwattanaphon M, Gotzek D, Dijkstra MB, Oettler J, Comtesse F, Shih C-J, Wu W-J, Yang C-C, Thomas J, Beaudoing E, Pradervand S, Flegel V, Cook ED, Fabbretti R, Stockinger H, Long L, Farmerie WG, Oakey J, Boomsma JJ, Pamilo P, Yi SV, Heinze J, Goodisman MAD, Farinelli L, Harshman K, Hulo N, Cerutti L, Xenarios I, Shoemaker D, Keller L. **2011**. The genome of the fire ant *Solenopsis invicta*. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 5679–5684.

Xu L, Wu M, Han Z. **2013**. Overexpression of multiple detoxification genes in deltamethrin eesistant *Laodelphax striatellus* (Hemiptera: Delphacidae) in China (G Smagghe, Ed.). *PLoS ONE* **8**: e79443.

Yamauchi K, Asano Y, Lautenschläger B, Trindl A, Heinze J. **2005**. A new type of male dimorphism with ergatoid and short-winged males in *Cardiocondyla* cf. *kagutsuchi*. *Insectes Sociaux* **52**: 274–281.

Yamauchi K, Furukawa T, Kinomura K, Takamine H, Tsuji K. **1991**. Secondary polygyny by inbred wingless sexuals in the dolichoderine ant *Technomyrmex albipes*. *Behavioral Ecology and Sociobiology* **29**: 313–319.

Yamauchi K, Ishida Y, Hashim R, Heinze J. **2006**. Queen-queen competition by precocious male production in multiqueen ant colonies. *Current biology* **16**: 2424–2427.

Yamauchi K, Ishida Y, Hashim R, Heinze J. **2007**. Queen-queen competition and reproductive skew in a *Cardiocondyla* ant. *Insectes Sociaux* **54**: 268–274.

Yan H, Simola DF, Bonasio R, Liebig J, Berger SL, Reinberg D. **2014**. Eusocial insects as emerging models for behavioural epigenetics. *Nature Reviews Genetics* **15**: 677–688.

Yandell M, Ence D. **2012**. A beginner's guide to eukaryotic genome annotation. *Nature Reviews Genetics* **13**: 329–342.

Yang Q, Gong Z-J, Zhou Y, Yuan J-Q, Cheng J, Tian L, Li S, Lin X-D, Xu R, Zhu Z-R, Mao C. **2010**. Role of *Drosophila* alkaline ceramidase (*Dacer*) in *Drosophila* development and longevity. *Cellular and Molecular Life Sciences* **67**: 1477–1490.

Yang Z. **2007**. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular biology and evolution* **24**: 1586–1591.

Zeh DW, Zeh JA, Ishida Y. **2009**. Transposable elements and an epigenetic basis for punctuated equilibria. *BioEssays* **31**: 715–726.

Zerbino DR, Birney E. **2008**. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* **18**: 821–829.

Zhou X, Slone JD, Rokas A, Berger SL, Liebig J, Ray A, Reinberg D, Zwiebel LJ. **2012**. Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding (NA Moran, Ed.). *PLoS genetics* **8**: e1002930.

Zhu H, Shen H, Sewell AK, Kniazeva M, Han M. **2013**. A novel sphingolipid-TORC1 pathway critically promotes postembryonic development in *Caenorhabditis elegans*. *eLife* **2**: e00429–e00429.

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. *Applied and Environmental Microbiology* **72**: 6027–6033.

9. Acknowledgements

First of all, I would like to express my gratitude to Jürgen Heinze, who gave me the opportunity to work on a remarkable ant and a truly fascinating topic. In addition, I am very grateful for his thorough support, enabling me to experience conferences, courses and collecting trips on four different continents and broaden my horizon in many different ways.

Furthermore, I thank Judith Korb and Stephan Schneuwly for agreeing to review my thesis and to be part of the committee, and thanks to Jürgen Gadau for mentoring me over the past four years.

My gratitude belongs to Jan Oettler, who has been integral in developing this thesis. According to my email inbox, we have exchanged a staggering 1731 emails over roughly four years. That's 1.18 emails per day. His contributions are countless including supervision and liberal guidance, extensive discourse and challenging discussions over TEs, ants, coffee, music and everything else, constant support and constructive criticism, and an enriching friendship. Big up. Give danks.

Several other friendships began at the department but extended beyond the boundaries of the University. Thank you Robert, Antonia, Claudi, Tobi, Eva, Nicky, Daniel, and Richie for a wonderful time in Regensburg.

I would also like to thank the members of the department in general and some people in particular. Abel, Katha, and Jürgen, thank you for the splendid climbing sessions. Sonja and Tina, thank you for the great time in South Africa. I am thankful for the opportunity of getting to know you Ali, Alex, Doris, Agnes, Niko, Adnan, Tommy, Bine, and Marion. I very much enjoyed your company, whenever I left my desk to roam the offices and hallways of the department and at every other occasion.

The last sentences are devoted to my family and my fiancée Loredana.

I am grateful to my family for their endless support, and for believing in me and the decisions I take. Thank you all for visiting me in Regensburg so frequently and for your interest in my temporary home deep in the heart of the Bavarian cosmos.

Loredana. There is little I could have achieved without you. I am grateful for having you by my side. Thank you for your unrestrained support. Thank you for giving me so much energy, bliss and motivation, and for sharing your life with me in Bochum, Regensburg and everywhere else. I will always treasure the time we spent together in Regensburg and I am looking forward to the things that lie ahead of us.

Supplementary Material

Contents

S1.1 Supplementary Material - Chapter 2 ..... S3

S1.1.1 Supplementary Methods S4

S1.1.2 Supplementary Figures S14

S1.1.3 Supplementary Tables S23

S1.2 Supplementary Material - Chapter 3 ..... S37

S1.2.1 Supplementary Data S38

S1.2.2 Supplementary Tables S40

S1.3 Supplementary Material - Chapter 5 ..... S42

S1.3.1 Supplementary Methods S43

S1.3.2 Supplementary Figures S47

S1.3.3 Supplementary Tables S48

S1.4 Supplementary References S49

S1.1 Supplementary Material - Chapter 2

Transposable element islands facilitate adaptation to novel environments in an invasive species

S1.1.1 Supplementary Methods

Organisms

Live colonies of *Cardiocondyla obscurior* were collected from aborted fruits on coconut trees (*Cocos nucifera*) in Brazil (collected in 2009) and from bark cavities in coral trees (*Erythrina* sp.) in Japan (collected in 2010). The colonies were transferred to Regensburg and placed in plastered petri dishes. Food (honey-soaked shreds of paper; *Drosophila* or small chunks of *Periplaneta americana*) and water were provided every three days and colonies were kept in incubators under constant conditions (12h 28° C light/12h 24° C dark). We emphasize that this is one species because recombinant inbred lines have produced viable offspring for over three years (3-4 generation / year) in our lab. Sampled individuals for subsequent DNA/RNA extractions were transferred to Eppendorf tubes, snap-frozen in liquid nitrogen and stored at -80° C.

Colony size

To assess differences in colony structure we used a more recent data set with detailed collection data. Colonies collected and censused immediately in November 2013 (BR) and April 2011 (JP) contained similar numbers of workers (Mann Whitney $U = 778.5$, $Z = -0.634$, $p = 0.526$; BR: median = 28, quartiles 21.75 and 51.25, n = 27 colonies; JP: median = 29, quartiles 16 and 47, n = 64). In contrast queen number was higher in Japan (Mann Whitney $U = 501$, $Z = -3.084$, $p < 0.003$; BR: 5 queens, quartiles 3, 8, n = 27 JP: 10 queens, quartiles 4, 19, n = 64).

Morphometry

We compared body size of workers, queens and males of each population drawn randomly from different source colonies, using four continuous morphological characters (head width (HW), head length (HL), thorax width (TW), and thorax length (TL)), measured under a Keyence VH Z00R. In workers HL and HW were correlated (Pearson's $r = 0.233$, $p = 0.028$, n = 97) as well as TL and TW ($r = 0.257$, $p = 0.012$, n = 96). Workers from the BR lineage had smaller HW (Mann Whitney $U = 394$, $Z = -5.647$, $p < 0.001$) and smaller TW ($U = 36$, $Z = -8179$, $p < 0.001$). In queens all four characters were tightly correlated with each other (minimum Pearson correlation HL – TW, $r = 0.492$, $p < 0.001$, n = 59). Queens from BR and JP did not differ in head size (HW: $U = 378$, $Z = -0.864$, $p = 0.387$; HL: $U = 312$, $Z = -1,865$, $p = 0.062$) but BR queens had smaller thoraces (TL: $U = 171$, $Z = -4.003$, $p < 0.001$; TW: $U = 168$, $Z = -4.048$, $p < 0.001$). In wingless males the characters were also strongly correlated (minimum Pearson correlation HL – TW, $r = 0.571$, $p = 0.002$, n = 27) but did not differ between BR and JP (HL: $U = 84$, $Z = -0.340$, $p = 0.756$; TW: $U = 75$, $Z = -0.776$, $p = 0.458$).

Behavioural assays

We tested the behaviour of experimental colonies towards individual workers or queens from either the same or the other lineage. Experimental colonies consisted of 20 workers, one mated queen, and brood. Colonies were housed in small petri dishes with plaster flooring and a 5-cent sized deep indentation covered by a dark red cover slide. These colonies were allowed to adjust to their new nest for one week prior to the trials. Trials were performed under dimmed ambient red light (6 lux). For each trial we removed the cover slide carefully and waited for five minutes to minimize effects by the disturbance before placing one alien individual into the vicinity of the nest. In addition to workers and mated queens of *C. obscurior*, we also performed trials with individual workers of *Wasmannia auropunctata* (*Waur*), to assess aggression against another ant species. After the introduction, we noted the behaviour for a period of 5 minutes or until the intruder was killed. We scored the behaviour with 1: Light antennation, 2: Antennation, display of mandible threat, 3: Antennation and short biting/pinches, 4: Antennation, short immobilization and biting, 5: Severe biting, occasional stinging and death of the intruder. Trials for which no interaction between intruder and resident occurred within 5 minutes were discarded. We performed a GLM comparing high aggressive interactions (score 5) versus all other categories combined, separately for workers and queens (Supplementary Tables S1.10-S1.11).

Chemical analyses

We analysed 8 BR and 8 JP colonies for differences in cuticular lipid profiles. Ants were extracted for 10 min in batches of 6 individuals in 40 µl Hexane containing 30 ng methyl decanoate as internal standard. Extracts were analysed on a GC2010 gas-chromatograph (GC) connected to a QP2010 plus mass-spectrometer (MS; both Shimadzu, Duisburg, Germany). The GC was equipped with a non-polar capillary column (BPX-5, 30 m length, 0.25 mm inner diameter, 0.25 µm film thickness; SGE Analytical Science, Milton Keynes, UK). Helium was used as carrier gas with a constant linear velocity of 50 cm s^{-1}. The temperature program of the GC-oven started at 80 °C and was raised by 5° C min^{-1} to 300° C. The MS was run in electron impact (EI) mode at 70 eV and set to a scan range from 35 to 600 mz^{-1}. All samples were injected split-less at an injector temperature of 300° C. n-Alkanes were identified by comparing retention times and mass spectra with those of synthetic reference compounds. Methyl-branched CHCs were identified by interpretation of diagnostic ions and comparison of linear retention indices with literature data (Carlson et al. 1998).

For further analysis we used only those peaks that had a minimal area of 1 % in at least 75 % of the samples of at least one lineage. Twenty-two Aitchison-normalized peak areas were subjected to a principal component analysis followed by linear discriminant analysis with leave-one-out cross validation on the first four PCs using the R package vegan (Oksanen et al. 2013).

DNA extraction

The reference genome is based on one colony that was kept under strict inbreeding in the lab for four generations prior to extractions. Sampled ants were ground with disposable micro-tube pestles and whole DNA was extracted with CTAB (Sambrook and Russell 2001). Extracts were treated with proteinase K and RNAse H, washed twice with ethanol, dried, and finally dissolved in sterile water. We extracted DNA from 900 ants, which were pooled to be sequenced with 454 technology. Extracts of five, ten and 30 Brazilian males and 26 Japanese males, respectively were used for Illumina libraries.

DNA library preparation and sequencing

Absorbance measurements at 260 nm and 280 nm (NanoDrop 1000) and Agilent Bioanalyzer traces were obtained for basic quality control of DNA samples designated for paired-end Illumina sequencing. Shearing of extracted DNA was performed on a Covaris S2 AFA system. For Illumina sequencing, we generated 200 and 500 bp insert libraries with Illumina's TruSeq DNA sample preparation kits from 5 µg of total DNA. Quality control and library preparation were carried out by the KFB sequencing centre of the University Regensburg, sequencing runs were performed by Illumina (Hayward, USA) on a HiSeq2000.

Quality control, library preparation, and sequencing of 8-kb and 20-kb long paired end (LPE) libraries (454, Roche) were carried out by Eurofins MWG Operon (Ebersberg, Germany). Extracted DNA was fragmented into the appropriate fragment sizes (8 kb and 20 kb) using the HydroShear DNA Shearing Device (GeneMachine). Further library preparation was performed according to "GS FLX Titanium Paired End Library Prep 20+8kb Span Method Manual" before sequencing on a GS FLX Titanium (Roche).

De novo genome assembly

We generated relatively few genomic 454 reads – about 2.3x genome coverage, a single run of the sequencer. Additional coverage was provided by Illumina reads and connectivity was provided by the 8-kb and 20-kb 454 mate pairs (Supplementary Table S1.2). The resulting N50 scaffold and contig sizes of the assembly show that the data was sufficient for high quality assembly (Supplementary Table S1.13). The assembly was created with MSR-CA version 1.4 open source assembler (University of Maryland genome assembly group at ftp.genome.umd.edu/pub/MSR-CA/). The MSR-CA assembler combines a deBruijn graph strategy with the traditional Overlap-Layout-Consensus employed by various assembly programs for Sanger-based projects (Arachne, PCAP, CABOG, etc.). The MSR-CA uses a modified version of CABOG version 6.1 for contiging and scaffolding. The combined strategy allowed us to natively combine the short 100 bp Illumina reads and longer 454 reads in a single assembly without resorting to an approach that would require one to assemble each

type of data separately and then creating a combined assembly. Total run time for the assembly was approximately 3 days on a 16-core AMD Opteron computer with 128 Gb RAM.

Using CEGMA (Parra et al. 2007) on the genome sequence to assess the completeness of the assembly, we confirmed complete presence of 244 of 248 ultra-conserved genes (98.39 %). We analysed seven other published ant genomes with CEGMA and they all performed similarly well, with the draft genome of *L. humile* containing the highest number (245) of complete ultra-conserved genes. The other genomes contained 228 (*S. invicta*), 241 (*C. floridanus*), 234 (*A. cephalotes*), 243 (*A. echinatior*), 243 (*P. barbatus*), and 242 (*H. saltator*) complete copies. The percentage of core eukaryotic genes with more than one complete ortholog was elevated in *C. obscurior* (23.77 %), compared to the other analysed ant genomes (9.65 % - 12.40 %).

Whole RNA extraction, normalized cDNA library preparation, and transcriptome assembly

We sampled individuals from the same BR colony that was used for the genomic DNA sequencing. Whole RNA was extracted from separate pools of eggs, the three larval stages, prepupae, pupal and different adult stages of queens, workers, ergatoid males and winged males using TRIzol (Life technologies) and subsequent Microcon purification (Millipore). Equal quantities of RNA from each extract were combined in a single pool, which was subsequently used to generate a normalized, random-primed cDNA library for emPCR-based sequencing. Sequencing was carried out on a GS FLX using Titanium series chemistry by Eurofins MWG Operon (Ebersberg, Germany), generating 1 245,994 reads (0.4 Gb).

We used the FastX toolkit (hannonlab.cshl.edu/fastx_toolkit/) for quality control of raw reads and only kept high quality reads (length 10-550 bases, minimum quality scores of 20 for 70 % of the called bases). The remaining 1,122,247 reads were submitted to the reference based transcriptome assembly with Newbler v2.6 (Roche, options "-cdna -gref -ml 60 -mi 95"). We generated a total of 19,325 contigs ranging between 500 and 12,699 bases length (N50 1 155 bases) that were supplied as EST evidence to MAKER in the subsequent gene annotation.

Gene annotation

MAKER version 2.20 (Holt and Yandell 2010) was run on the *C. obscurior* draft genome using the assembled transcriptome, amino acid sequence data from Swiss-prot and the ant genomes portal (Supplementary Table S1.14), in addition to hand-curated amino acid sequence for desaturase proteins in *Acromyrmex echinatior* and *Pogonomyrmex barbatus* (Simola et al. 2013). Repetitive regions were masked using a custom repeat library constructed with RepeatModeler (repeatmasker.org/RepeatModeler.html), all organisms in Repbase (Jurka et al. 2005), and a list of known transposable elements in MAKER. *Ab initio* gene predictors (GeneMark (Ter-Hovhannisyan et

al. 2008), Augustus (Stanke et al. 2006), and SNAP (Korf 2004)) were trained on the assembly and also used by MAKER to generate gene models.

The official gene set contains 17,552 genes, of which 9,552 genes contain a known protein domain as detected by IPRScan (Quevillon et al. 2005) and 72.5 % of the genes in the final gene set have an AED (annotation edit distance) of less than 0.5, which is consistent with a well-annotated genome (Holt and Yandell 2010). The total number of *ab silicio* predicted genes falls within the range of recent estimates for the other sequenced ant species (Libbrecht et al. 2013).

A comparison of the gene set with seven other ant genomes, *Apis mellifera* and *Nasonia vitripennis* using orthologous groups annotated with OrthoDB (Waterhouse et al. 2012) revealed extensive duplication events in *Cobs* and *Nvit* (Supplementary Table S1.15).

Functional annotation of Cobs1.4 genes

Gene Ontology (GO) term annotation for Cobs1.4 genes was done using the Blast2GO pipeline (Conesa and Götz 2008). Predicted protein sequences for each gene were blasted against the non-redundant NCBI protein database nr (retrieved May, 3rd 2013) and parsed through Interpro scan (IPS 5-RC6). BLASTx returned hits with e-values less than 1e-10 for ~70 % of the transcripts and 53,818 Interpro domains were annotated in 9,252 gene models. Using Blast2gPipe (v2.5, default settings), 43,166 GO terms were retrieved for 8,908 genes. We also used the Blast2go BDA system to assign provisional gene aliases for 3,415 genes. All computations were performed on the Queen Mary University of London SBCS-informatics Apocrita compute facility.

Repeat annotation

Our goal was to use existing repeat prediction tools to generate *de novo* repeat libraries for several insect genomes. To this end, we implemented a pipeline that has several repetitive element prediction tools at its core.

Our pipeline combines results from RepeatModeler (v1.04) and PILER-DF (Edgar and Myers 2005; DePristo et al. 2011). RepeatModeler is a wrapper around two *de novo* repetitive element detection algorithms, RECON and RepeatScout. It also uses TandemRepeatsFinder (Benson 1999) to search for simple repeats and RepeatMasker for masking and annotating repeat elements. PILER-DF is not a part of the RepeatModeler package, but is also used for repeat prediction. For both RepeatModeler and PILER-DF, the output consists of consensus sequences corresponding to repetitive elements in the input genome. Essentially, the repetitive elements found throughout the input genome are clustered into distinct repeat "families" based on similarity of sequence. A repeat "family" sequence can be thought of as a best representative consensus for all of its member sequences. Consensus sequences are then pooled into a repetitive element library for this input genome. In a latter part of the pipeline, we will use our consensus repeat library to scan our genome of interest to find repeat elements.

The first part of the pipeline involves generating repeat family sequences using the tools mentioned above. In the next part of the pipeline we (1) combined RepeatModeler and PILER results, (2) ran quality control, (3) added additional annotations for consensus repeat sequences. Combining the results of multiple prediction tools will inevitably result in duplicates. To remove duplicates, we performed an all-by-all sequence comparison of our combined repeat prediction libraries and retained only one from pairs that show 80% identity over 80% of length (length of shorter sequence).
One pitfall of repeat prediction is that false positives are often genes or gene families containing transposable element-like domains or simple repeat domains (such as the calx-beta motif). When we developed this pipeline using the genomes of various fruit flies, *L. humile* and *A. cephalotes*, we found that the native quality controls in RepeatModeler and PILER did not sufficiently filter false positives. Thus, we enforced a stricter threshold. In our current pipeline, we used the genome of *Drosophila melanogaster* as a reference to find false positives (BLASTx hits with at least 50% identity over 50% length), which are removed. We arrived at these Blast parameter thresholds through a combination of Blast searches and manual curation of false positives in several genomes. Although 50/50 is a safe threshold for not including genes into a repeat library, we note that a hard sequence similarity cut-off such as ours serves as a coarse filter.

While the RepeatModeler pipeline annotates its repeat predictions using RepBase (Smith et al. 2007), PILER has no such functionality. We annotated PILER consensus repeats using RepeatMasker, which uses RepBase as a reference. Additionally, we scanned all consensus repeats for the presence of long terminal repeats (LTR) or terminal inverted repeats (TIR) using custom scripts. After all annotations were updated, the final *C. obscurior* repeat library was output in FASTA and EMBL format.

For predicting repetitive elements in the *C. obscurior* genome, we added a library for *C. obscurior* generated in a first run of the pipeline to our master library consisting of the following: the latest RepBase (at the time, this was version 20121104), our *de novo* consensus repeat libraries generated from the genomes of 7 ants, 6 bees/wasps, and 12 drosophilid flies (see below) and reran the pipeline.

1. Ants – *Atta cephalotes, Acromyrmex echinatior, Camponotus floridanus, Cardiocondyla obscurior, Harpegnathos saltator, Linepithema humile, Pogonomyrmex barbatus, Solenopsis invicta*
2. Bees/Wasps – *Apis florea, Apis mellifera, Bombus terrestris, Megachile rotundata, Nasonia vitripennis*
3. Flies – *D. ananassae, D. erecta, D. grimshawi, D. melanogaster, D. mojavensis, D. pseudoobscura, D. persimilis, D. sechellia, D. simulans, D. virilis, D. willistoni, D. yakuba*

We used CENSOR (Kohany et al. 2006) for reference-based annotation of repetitive elements in *C. obscurior*. Our master repeat library was used as a reference. CENSOR uses simplistic filters such as seg, xnu, and dust to search for tandem repeats. We supplemented this with our own TandemRepeatsFinder (TRF) results. All hits to our master library were recorded in a GFF3 format file.

Mapping of genomic reads against the Cobs1.4 reference genome

For each lineage, we randomly sampled 140 M 100 bp reads from libraries generated from 26 (JP) and 30 (BR) male pupae. Raw reads were parsed through quality filtration and adapter trimming (Trimmomatic v0.22 (usadellab.org/cms/?page=trimmomatic), options: HEADCROP:7 LEADING:28 TRAILING:28 SLIDINGWINDOW:10:10) and mapped against the BR reference genome with BWA samse v0.5.9-r16 (Li and Durbin 2010) in single end mode. Ambiguous reads were re-aligned with Stampy v1.0.21 (Lunter and Goodson 2011; Langmead and Salzberg 2012) to reduce misalignments (Nielsen et al. 2011). Aligned reads were stored in SAM format.

De novo assembly of unmapped reads

We extracted 23,054,888 Illumina reads generated from the JP lineage that could not be mapped against the reference genome using custom perl scripts. After filtering with Trimmomatic v0.22 (HEADCROP:7 LEADING:28 TRAILING:28 SLIDINGWINDOW:10:10), we generated *de novo* assemblies of these reads using velvetoptimizer v2.2.4 (bioinformatics.net.au/software.velvetoptimiser.shtml) with velvet 1.2.07 (Zerbino and Birney 2008). The optimised assembly contained 144,664 contigs (N50 8.7 kb, mean length 1.2 kb). We removed short contigs, contigs with extreme coverage, and contigs returning BLASTn hits against the BR raw draft assembly with an e-value < 1e-10, leaving a final set of 4,108 contigs (N50 0.4 kb, mean length 0.34 kb) not present in the BR genome assembly. These contigs were blasted (BLASTx) against NCBI's non-redundant database (retrieved May, 3rd 2013) and against the Cobs1.4 proteins. Contigs without hits below an e-value of 1e-10 in eukaryotes or with hits against a Cobs1.4 protein (e-value below 1e-10) were removed, producing a set of 17 contigs containing an open-reading frame that are only present in the JP genome.

Calculation of sliding windows

One kb windows of different stats (TEs, exons, SNPs, coverage) were calculated for all scaffolds based on GFF, VCF, and SAM files. For GFF and VCF files, custom bash and perl scripts were used to calculated TE and exon bases per 1 kb, and variant calls (see below) per 1 kb. Coverage per 1 kb was calculated from SAM files, using samtools' depth algorithm (Li 2011) and custom bash and perl scripts. Subsequent processing, calculating of 200 kb sliding windows, and plotting of the data was performed with R v3.0.0 (r-project.org).

Detection of small-scale genomic structural variants

To identify differences in the genome affecting genes, we filtered 1-kb windows (see above) where log2 coverage ratio (BR/JP) was below -0.8 or above 0.8. Values below -0.8 suggest either regions of low coverage in BR or regions of elevated coverage in JP, *vice versa* for values above 0.8. We applied a second filter based on exon and TE content of each individual window and selected only those windows containing more annotated exon than TE bases, thus focusing on windows dominated by exonic over transposon sequence. A list of candidate genes was compiled based on intersection of the MAKER annotation with the list of candidate 1-kb windows. The absolute base-wise coverage for BR and JP as well as the $\log_2$ coverage ratio were plotted against the genomic position and candidate genes were manually inspected and classified as either partial or full gene deletions or duplications.

Experimental proof-of-principle was conducted by PCR and Sanger sequencing for two deletion candidates (*Cobs_13563* and *Cobs_01070*) and by real-time quantitative PCR for four duplication candidates (*Cobs_13806*, *Cobs_17872*, *Cobs_13486*, and *Cobs_16853*) (see Supplementary Figure S1.7). For deletion candidates, we designed primers spanning the putative deletion and performed PCR on extracted genomic DNA for both lineages. PCR products were purified and Sanger sequenced to confirm the deletion. For duplication candidates, we designed primers within the putative duplicated genomic sequences and performed qPCR experiments (normalization against a single copy gene (*actin*, *Cobs_04257*)) on genomic DNA, isolated from three different colonies of each population. By calculating the ratio of normalized relative quantities between BR and JP copy number variations were confirmed.

Variant calling

Single nucleotide variant and InDel calling was carried out combining samtools and the GATK (McKenna et al. 2010; DePristo et al. 2011), retaining only those variants called consistently by both tools. Potential PCR duplicates were marked with Picard MarkDuplicates (picard.sourceforge.net/). Raw variant calls were produced with the GATK after local realignment around InDels. Subsequently, all calls were annotated and filtered; producing sets of high and low confidence SNVs and InDels, respectively. The set of high confidence SNVs was used to train the GATK's VariantRecalibrator for variant quality score recalibration to filter additional SNVs from raw variant calls. The final set produced with the GATK consisted of 783,009 called single nucleotide variants and 168,754 InDels.

Raw variant calls produced by samtools were filtered based on mapping quality and genotype quality (Q>29, GQ>31), resulting in a set of 601,214 SNVs and 151,656 InDels.

A total of 567,552 SNVs and 68,430 InDels were called consistently by both tools. The transition from Cobs1.3 to Cobs1.4 removed contaminating endosymbiotic scaffolds, resulting in a final variant set of 553,052 SNVs and 67,987 InDels stored in a single VCF file. Single nucleotide variants were annotated with SNPeff (Cingolani et al. 2012) to identify non-synonymous and synonymous substitutions.

Gene Ontology enrichment

To test for enrichment or depletion of certain GO terms in genes in TE islands, we performed a two-tailed GO enrichment analysis. The Gossip package (Blüthgen et al. 2005), implemented in Blast2GO, uses Fisher's Exact Test for each GO term and corrects for multiple testing. GO terms with FDR<0.05 were considered to be significantly enriched/depleted in the test set.

Enrichment of Transposable Element superfamilies

Similarly to the GO enrichment analyses, we tested all TE superfamilies for enrichment in TE islands. We performed one-tailed Fisher's Exact Tests for each superfamily in TE islands, testing for significant enrichment of copy numbers and in a second test for enrichment of total bases compared to other superfamilies in TE islands. We applied FDR corrections for multiple testing and considered all TE superfamilies to be significantly enriched in copy number or base count with an FDR<0.05.

Gene expression analysis with RNAseq

We extracted whole RNA from seven individual mated queens of the same age (4 weeks after pupal moult) and seven individual developing queens in the early 3rd larval instar (11-13 days). To sample larvae, we set up experimental colonies consisting of 20 workers and 20 to 30 methoprene-treated eggs, as queen development can be induced by treatment with low concentrations of the JH-analogue (Schrempf and Heinze 2006). Unsampled larvae from these colonies were kept alive to confirm the exclusive development of queen pupae. Sampled queens and larvae were placed individually in 1.5 ml Eppendorf tubes, snap-frozen in liquid nitrogen, and kept at -80° C till further processing.

We extracted whole RNA with the RNeasy Plus Micro kit (Qiagen) yielding 27 to 153 ng per individual larvae and 57 to 122 ng per individual queen. Single end Illumina libraries from amplified RNA (Ovation RNAseq system V2) were generated following the manufacturers protocol (Ovation Rapid Multiplexsystem, NuGEN). Sequencing on an Illumina HiSeq1000 at the in-house sequencing centre (KFB, Regensburg, Germany) generated ~20 M 100 bp reads per sample (Supplementary Table S1.16). Raw reads were filtered for adapter contamination (cutadapt, Martin 2011), parsed through quality filtration (Trimmomatic v0.27, options: LEADING:10 TRAILING:10 SLIDING:4:10 MINLEN:15), and mapped against the reference genome using the tophat2 (v2.0.8) and bowtie2 (v2.1.0) package (Langmead and Salzberg 2012; Kim et al. 2013, --b2-sensitive mode, mapping rate ~50 %). Low mapping rates are most likely a consequence of the required amplification step during library preparation. Gene expression analysis was carried out with DESeq2 (Love et al. 2014), based on count tables produced with HTSeq (Anders et al. 2015) against the Cobs1.4 MAKER annotation (Supplementary Table S1.16). Genes were considered to be differentially expressed at an FDR < 0.05 and expression values are reported as untransformed base means of read counts per treatment group, after correcting for library size differences ("size factor normalization").

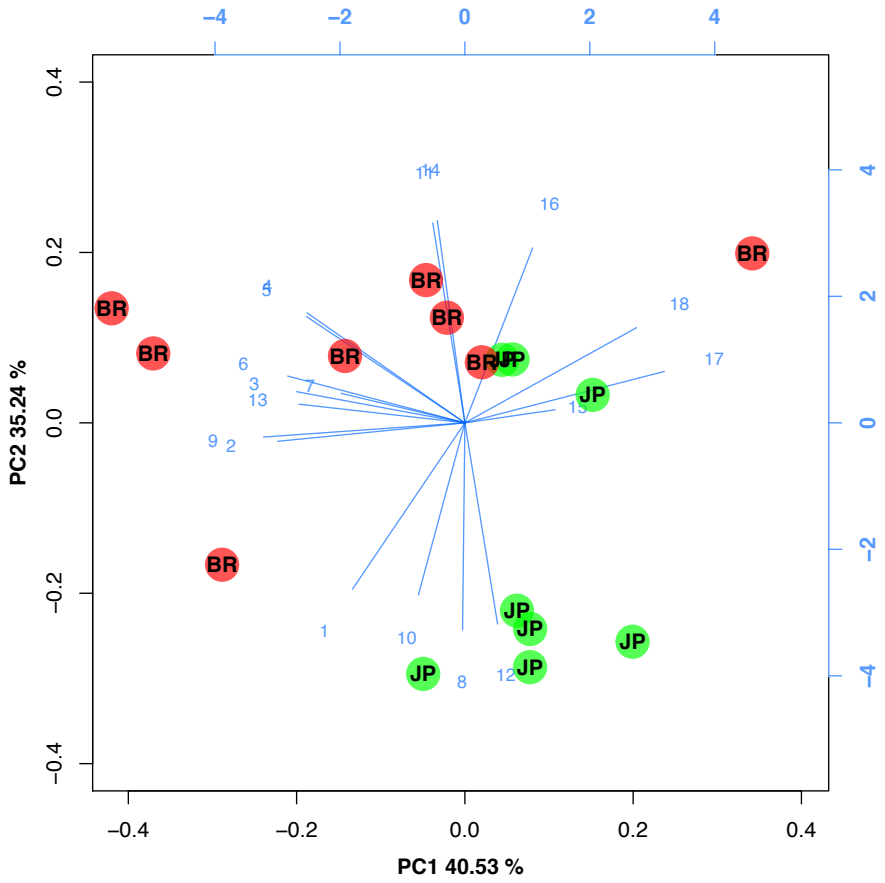
Calculation of exon wide CpG o/e values

Observed to expected CpG values for all exons were calculated as (Glastad et al. 2012):

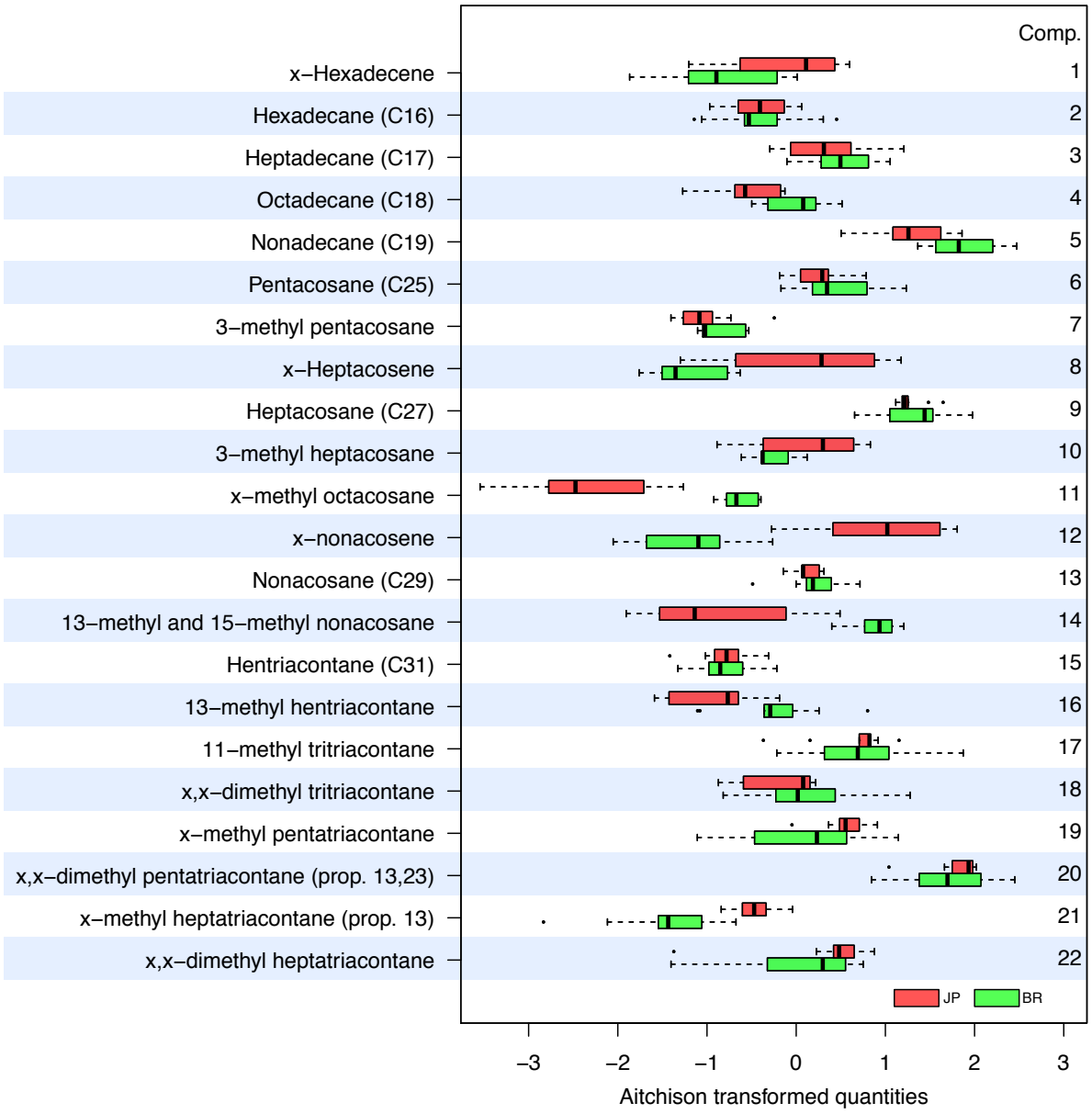
$$\frac{Obs}{Exp}CpG = \frac{n_{CpG}}{n_C \times n_G} \times N \qquad \text{(S1)}$$

where N is the total number of nucleotides in the analysed exon.

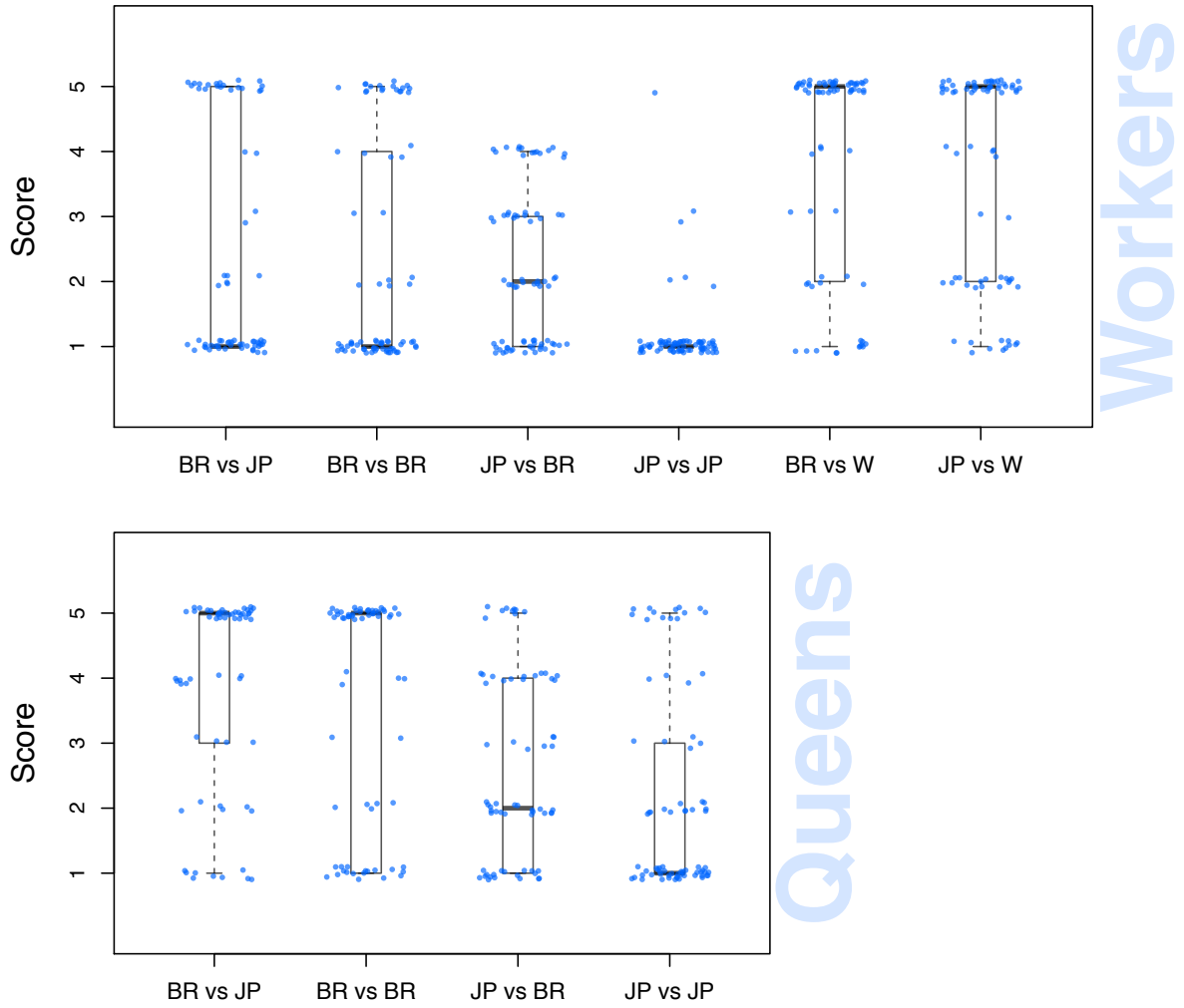
S1.1.2 Supplementary Figures



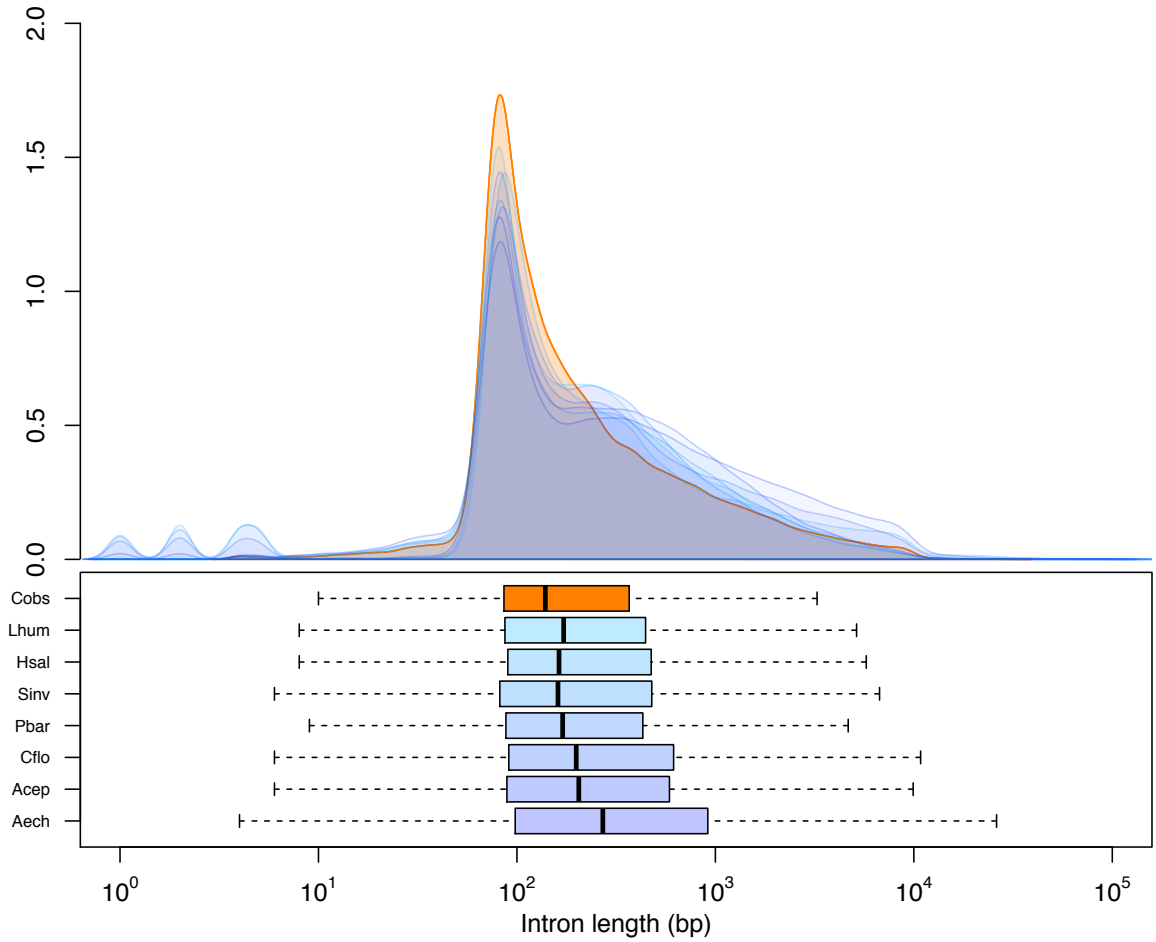
Supplementary Figure S1.1a: PCA plot of 22 Aitchison-transformed peak-areas. PC1 explains 40.53 % of the variance in the data matrix. Blue lines denote peak number.



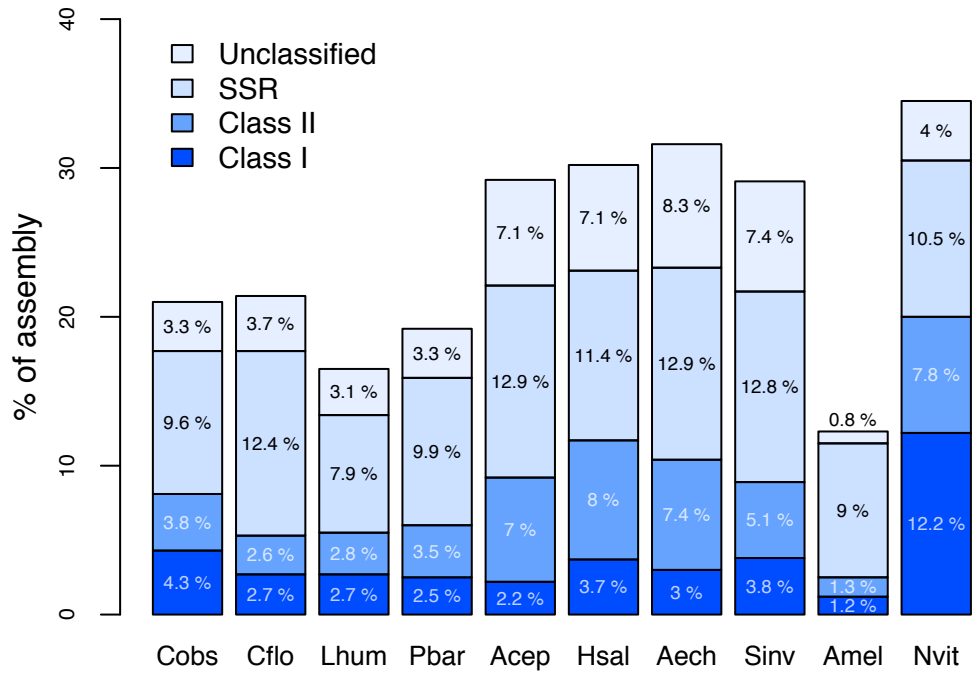
Supplementary Figure S1.1b: Boxplot of relative compound abundance in each lineage. Boxplot of relative compound abundance in each lineage. 'x' denotes unknown position of the double bond or methyl branch. Compound 14 could not be separated with the GC parameters used.



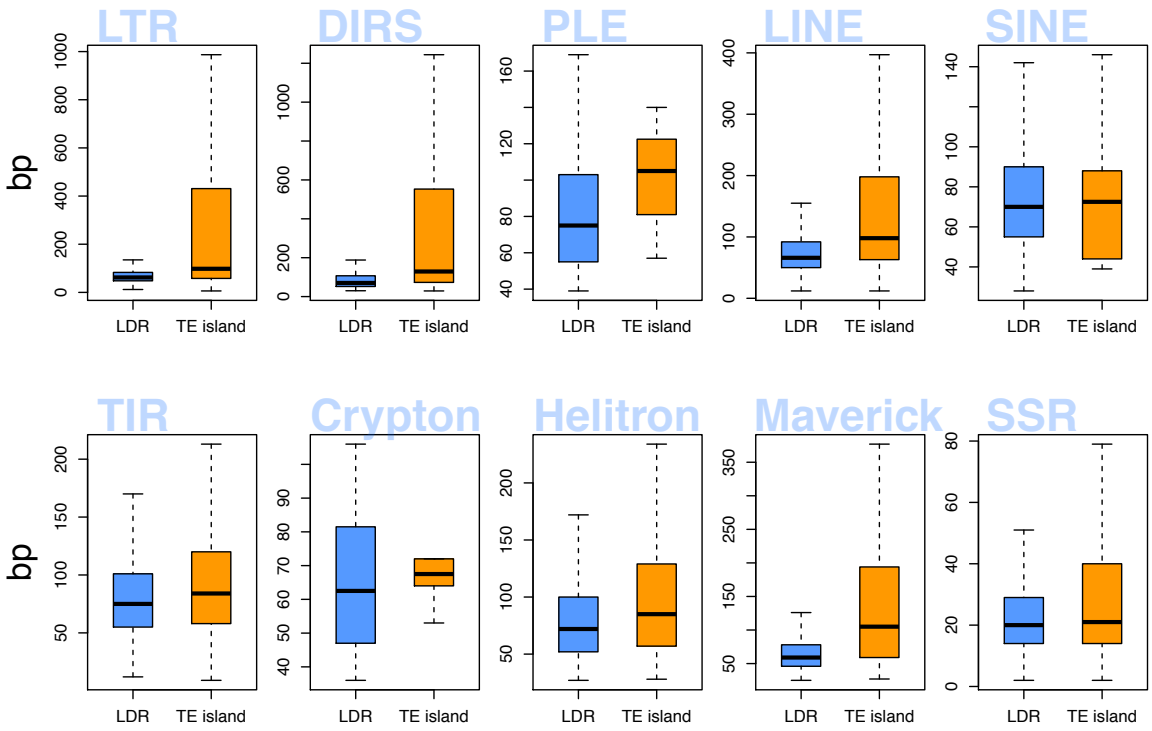
Supplementary Figure S1.2: Aggression indices in behavioural assay. Workers (top) and queens (bottom) of each lineage (and workers of *W. auropunctata*) were introduced to experimental colonies of either JP or BR. We scored the behaviour of the receiving colony based on defined aggression indices and tested for significant differences in potential for high aggressiveness between each of the tested combinations (origin of receiving colony vs. origin of the introduced ant) in a generalized linear model.



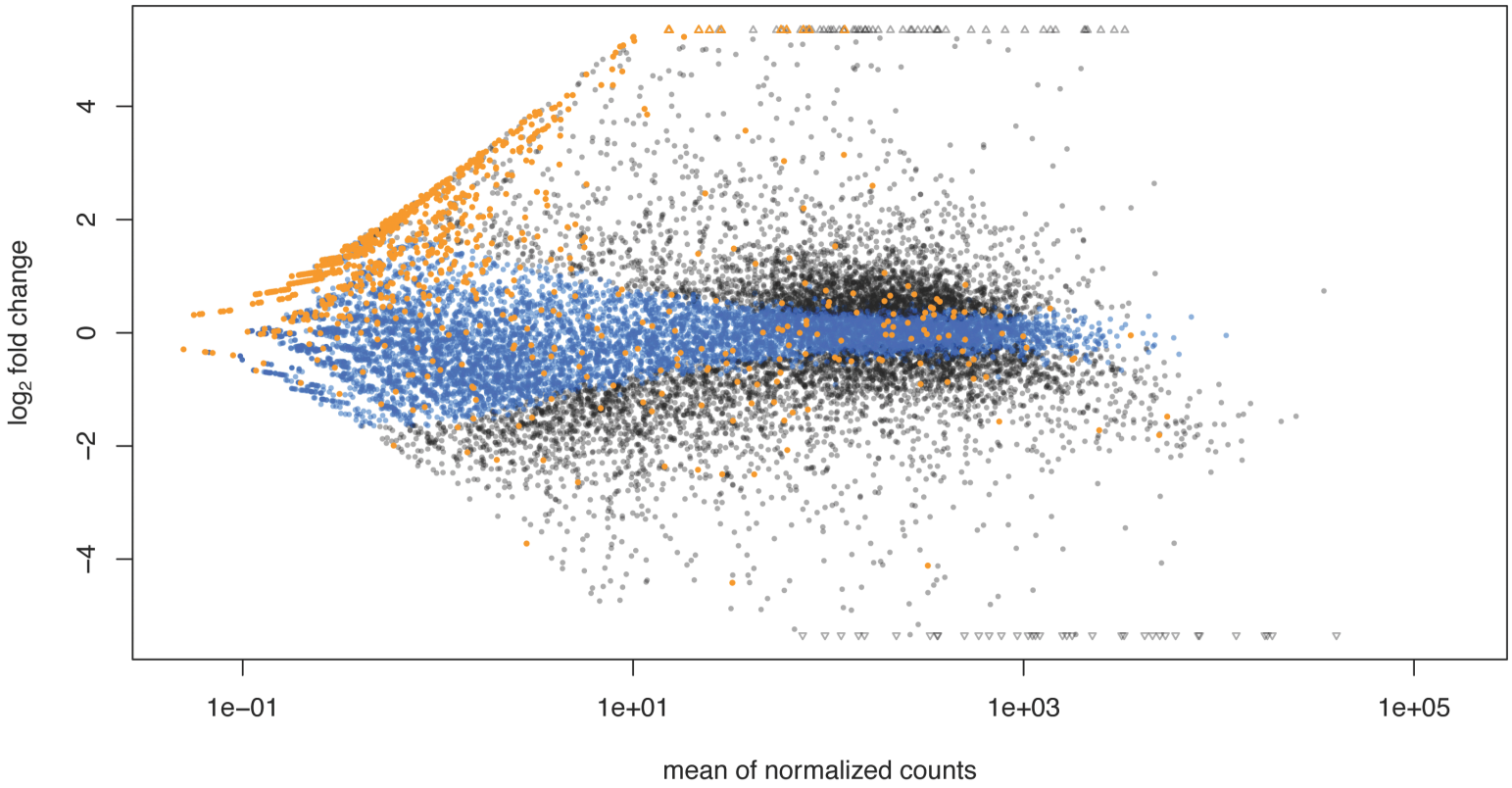
Supplementary Figure S1.3: Density plot of intron lengths for the sequenced ant genomes. The density plots for intron lengths in *C. obscurior* and seven other published ant genomes show that while the distribution is bimodal in other genomes, the introns of *C. obscurior* deviate from this pattern, with a single peak and a median intron length of 139 bp.



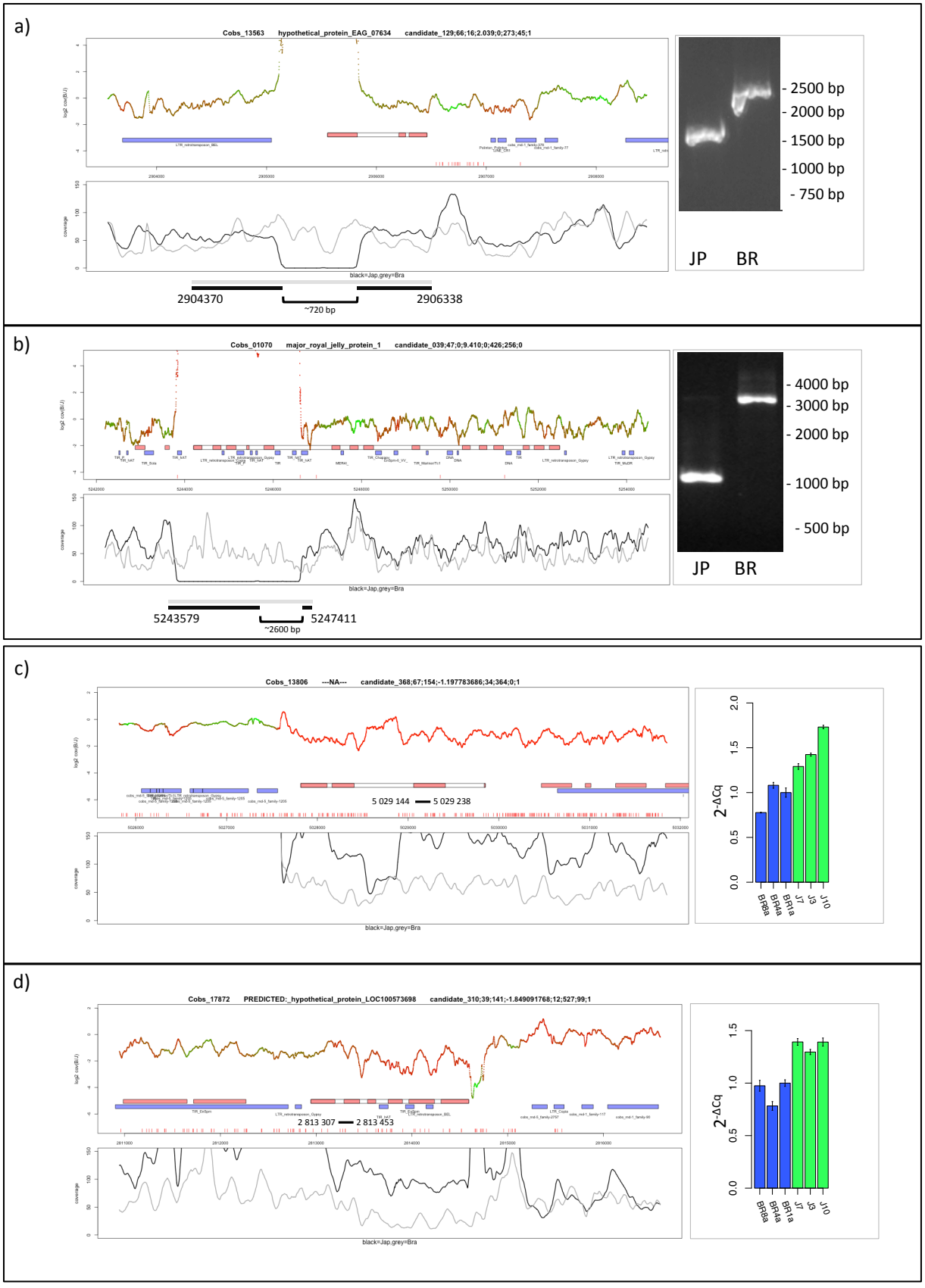
Supplementary Figure S1.4: Repeat content in sequenced ant genomes relative to assembled genome size. Relative repeat content of *C. obscurior* and nine hymenopteran genomes as calculated from the repeat annotations presented in this study. Across the analysed ant genomes, repeat content ranges between 16.5 % in *L. humile* to 31.5 % in *A. echinatior*. Relative class I content is higher in *C. obscurior* (4.3 %) than in any of the other ant genomes, yet overall relative repeat content in not different from the smaller genomes (*Cflo*, *Lhum*, *Pbar*). The genomes of *A. mellifera* (*Amel*) and *N. vitripennis* (*Nvit*) are distinct from the analysed ant genomes in having either much less (*Amel*) or much more (*Nvit*) annotated TEs. SSR = Short simple repeats.

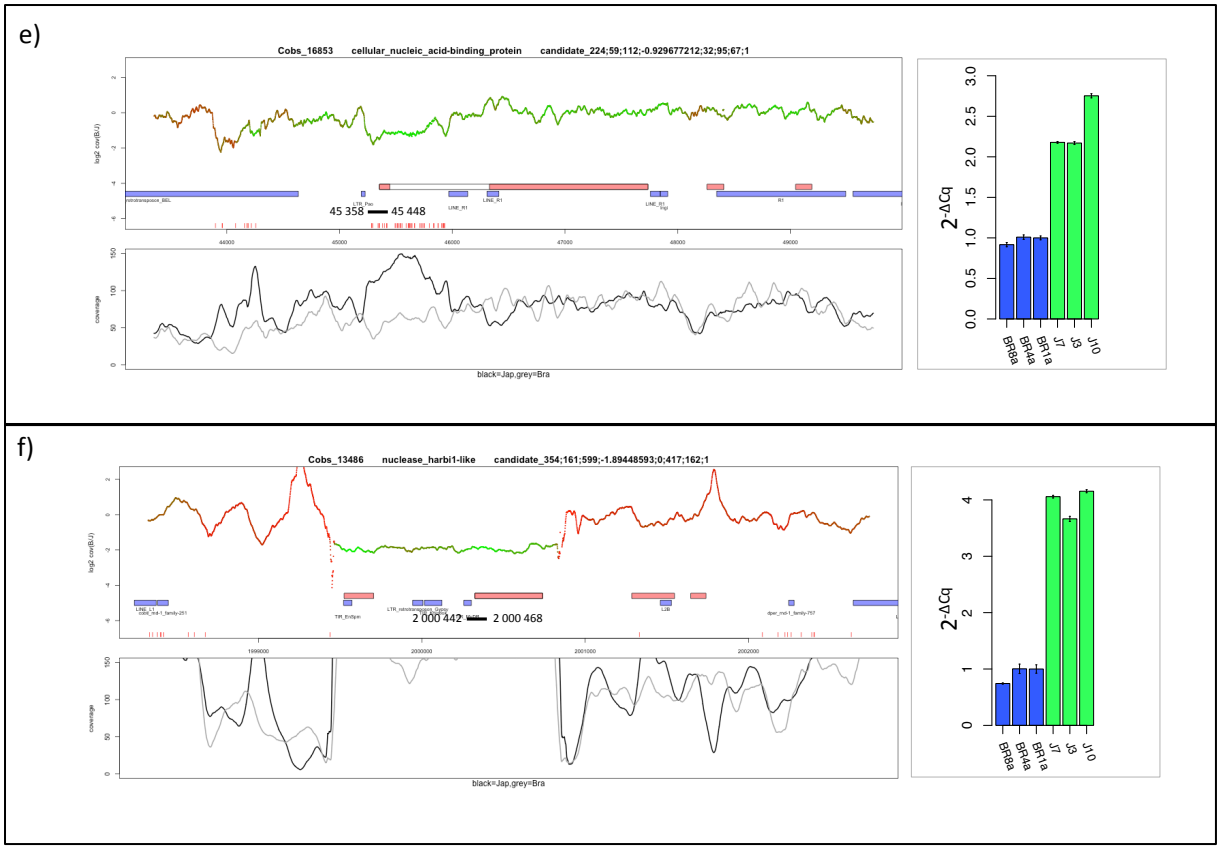


Supplementary Figure S1.5: Length polymorphism in TE superfamilies and simple repeats between LDRs and TE islands. Length polymorphism in TE superfamilies and simple repeats between LDRs and TE islands. Median element length for all analysed superfamilies is higher in TE islands than LDRs, suggesting local differences in TE dynamics.



Supplementary Figure S1.6: MA plot for differential expression of genes between 3rd instar larvae and queens. MA plot for differential expression of genes between 3rd instar larvae and queens. Log ratios of expression of each gene is plotted against the log mean average expression across all samples. Black dots represent genes with significantly different expression. Blue dots show genes that are not significantly different expressed. Genes located in TE islands are plotted in orange. Most TE island genes appear to be more strongly expressed in queens than larvae, while the overall expression of TE island genes is low.


a)
Cobs_13563 hypothetical_protein_EAG_07634 candidate_129;66;16;2.039;0;273;45;1
black=Jap,grey=Bra
2904370
~720 bp
2906338
- 2500 bp
- 2000 bp
- 1500 bp
- 1000 bp
- 750 bp
JP BR
b)
Cobs_01070 major_royal_jelly_protein_1 candidate_039;47;0;9.410;0;426;256;0
black=Jap,grey=Bra
5243579
~2600 bp
5247411
- 4000 bp
- 3000 bp
- 2000 bp
- 1000 bp
- 500 bp
JP BR
c)
Cobs_13806 ---NA--- candidate_368;67;154;-1.197783686;34;364;0;1
5 029 144 5 029 238
black=Jap,grey=Bra
2-ΔCq
BRBa BRHa BR1a J7 J3 J10
d)
Cobs_17872 PREDICTED:_hypothetical_protein_LOC100573698 candidate_310;39;141;-1.849091768;12;527;99;1
2 813 307 2 813 453
black=Jap,grey=Bra
2-ΔCq
BRBa BRHa BR1a J7 J3 J10



Supplementary Figure S1.7: Diagnostic plots and experimental confirmation for two deletion candidates (a, b) and four duplication candidates (c-f). Diagnostic plots were created for each of the 512 candidate loci by plotting the log2 JP/BR coverage ratio (red to green), gene models (red boxes), repetitive elements (blue boxes), heterozygous SNV calls (red ticks), and the absolute coverage (grey = BR; black = JP, lower panel). A) Partial deletion of *Cobs_13563* in the JP lineage. PCR and Sanger sequencing confirmed deletion of ~720 bp in the JP genome. B) Deletion of a *MRJP* in *Cobs_01070* in the JP lineage. PCR and Sanger sequencing confirmed deletion of ~2600 bp in the JP genome. C-F) Duplications in *Cobs_13806* (c), *Cobs_17872* (d), *Cobs_16853* (e), and *Cobs_13486* (f) as confirmed by real-time qPCR. $2^{-\Delta Cq}$ values for BR (blue) and JP samples (green) were normalized against colony BR1a. Primer combinations used: *Cobs_13563*: fw: 5'-CAGTTCGGGATGGCGCTC-3', rv: 5'-CGAAAGACTGGGGCTGCAA-3'; *Cobs_01070*: fw: 5'-TCCCGTCAAACCAATCGCAACTCG-3', rv: 5'-TGGGTTGCATCAGGCCACGTA-3'; *Cobs_13806*: fw: 5'- GCAACGGTGCTCACAGGAGCC-3', rv: 5'-AAAGGCGATGCCCTCCGTTGC-3'; *Cobs_17872*: fw: 5'-TCGTAGACGATTATATAGAGCG-3', rv: 5'-GTAGCAGAAGTAGAAGGCATTGG-3'; *Cobs_13486*: fw: 5'-TCATTGACATCGAATTCGTCATGGCTG-3', rv: 5'-AACGTGTAATGGCTGCTGCTATACTTC-3'; *Cobs_16853*: fw: 5'-GCGACGTCGAGATAAAGGTTTCG-3', rv: 5'-CGTTAATTGGTAGGGTTCGC-3'.

S1.1.3 Supplementary Tables

Supplementary Table S1.1: 22 compounds in cuticle extracts of BR and JP used for statistics

ID	Compound
1	x-Hexadecene
2	Hexadecane (C16)
3	Heptadecane (C17)
4	Octadecane (C18)
5	Nonadecane (C19)
6	Pentacosane (C25)
7	3-methyl pentacosane
8	x-Heptacosene
9	Heptacosane (C27)
10	3-methyl heptacosane
11	x-methyl octacosane
12	x-nonacosene
13	Nonacosane (C29)
14	13-methyl and 15-methyl nonacosane
15	Hentriacontane (C31)
16	13-methyl hentriacontane
17	11-methyl tritriacontane
18	x,x-dimethyl tritriacontane
19	x-methyl pentatriacontane
20	x,x-dimethyl pentatriacontane (prop. 13,23)
21	x-methyl heptatriacontane (prop. 13)
22	x,x-dimethyl heptatriacontane

Supplementary Table S1.2: Data used for the *C. obscurior* draft genome assembly. Genome coverage computed assuming 195 Mb estimated genome size

	No. of reads	Average length (bp)	Genome coverage
220 bp Illumina paired end	209 740 014	100	105x
8 Kb paired end (reads)	1 318 264	189	1.2x
8 Kb paired end (valid pairs)	416 174		
20 Kb paired end (reads)	1 131 046	194	1.1x
20 Kb paired end (valid pairs)	326 815		

Supplementary Table S1.3: Quantitative assembly statistics for Cobs1.4

Scaffolded sequence (bp)	177 892 999
N50 scaffold size (bp)	3 105 814
Total number of scaffolds	1 854
GC content	0.3958
Total no. of annotated genes	17 552
Total no. of gene models with AED<0.5	12 752
Total no. of genes with Interpro domain	9 552
Scaffolded sequence (bp)	177 892 999

Supplementary Table S1.4: Comparison of gene body, exon and intron structure of *C. obscurior* and other analysed ant genomes

	Gene Body				Exon				Intron			
Species	**Median length (bp)**	**Total length (Mbp)**	**Count**	**% of assembly**	**Median length (bp)**	**Total length (Mbp)**	**Count**	**% of assembly**	**Median length (bp)**	**Total length (Mbp)**	**Count**	**% of assembly**
C. floridanus	1790	71.85	17059	29.94%	179	20.75	83401	8.64%	199	51.10	66342	21.29%
P. barbatus	1838	52.51	17152	22.34%	180	20.49	81657	8.72%	170	32.03	64504	13.63%
H. saltator	1273	72.33	18561	24.35%	184	20.41	78601	6.87%	163	51.91	60040	17.48%
L. humile	2068	54.00	16097	25.04%	181	20.48	80622	9.50%	172	33.52	64525	15.55%
S. invicta	1365	44.30	16522	12.55%	182	17.25	66992	4.89%	161	27.06	50470	7.66%
A. echinatior	2783	92.35	17278	30.78%	180	21.00	85030	7.00%	271	71.35	67752	23.78%
A. cephalotes	1932	59.72	18090	18.84%	169	19.53	83900	6.16%	205	40.19	65810	12.68%
C. obscurior	1844	60.35	17552	33.92%	171	22.14	92173	12.45%	139	38.21	74621	21.48%

Supplementary Table S1.5: Relative content (%) of repetitive elements and TEs in the genomes of *C. obscurior* and other Hymenoptera

Type	Repeat	Cobs	Cflo	Lhum	Pbar	Acep	Hsal	Aech	Sinv	Amel	Nvit
LTR	**Gypsy**	**1.643**	0.94	1	1.247	0.725	1.284	0.901	1.216	0.389	4.754
LTR	**Copia**	0.369	0.227	0.247	0.179	0.147	0.348	0.2	0.4	0.131	0.957
LTR	**BEL**	**0.499**	0.244	0.264	0.151	0.158	0.206	0.266	0.485	0.061	0.488
LTR	**DIRS**	**0.095**	0.008	0.072	0.027	0.018	0.065	0.019	0.072	0.003	0.089
LTR	**Ngaro**	**0.018**	0	0	0	0	0	0	0	0	0
LTR	**Pao**	0.033	0.034	0.099	0.024	0.083	0.023	0.206	0.02	0.043	0.167
LTR	**ERV1**	0.081	0.056	0.021	0.061	0.079	0.093	0.09	0.074	0.045	0.055
LTR	**ERV2**	**0.041**	0.027	0.01	0.025	0.019	0.037	0.022	0.02	0.024	0.027
LTR	**ERV3**	0.01	0.007	0.002	0.007	0.006	0.01	0.007	0.006	0.006	0.008
LTR	**ERVK**	0	0	0	0	0	0	0	0	0	0.146
LTR	**ERVL**	0	0.001	0.001	0	0	0.004	0	0	0	0
LTR	**Unclassified**	0.307	0.392	0.152	0.089	0.162	0.202	0.257	0.358	0.142	1.558
LINE	**CR1**	**0.232**	0.091	0.035	0.061	0.059	0.156	0.063	0.121	0.073	1.349
LINE	**L1**	**0.163**	0.071	0.022	0.063	0.057	0.136	0.065	0.057	0.067	0.113
LINE	**L2**	0.034	0.025	0.034	0.035	0.113	0.025	0.196	0.061	0.007	0.093
LINE	**L2A**	**0.002**	0	0	0	0	0.001	0	0	0	0.001
LINE	**L2B**	0.007	0.004	0.002	0.021	0.002	0.009	0.003	0.009	0.001	0.004
LINE	**Jockey**	0.033	0.038	0.022	0.02	0.016	0.04	0.02	0.023	0.01	0.034
LINE	**LOA**	0.046	0.006	0.088	0.017	0.007	0.01	0.007	0.068	0.003	0.303
LINE	**R1**	0.199	0.144	0.255	0.119	0.065	0.667	0.136	0.315	0.02	0.253
LINE	**R2**	**0.032**	0.008	0.014	0.006	0.003	0.023	0.004	0.014	0.007	0.034
LINE	**R4**	0.015	0.008	0.008	0.008	0.007	0.009	0.007	0.007	0.012	0.012
LINE	**RTEX**	**0.007**	0.004	0.003	0.005	0.003	0.003	0.002	0.006	0.001	0.01
LINE	**Penelope**	0.04	0.071	0.068	0.097	0.221	0.035	0.269	0.12	0.01	0.113
LINE	**RTE**	0.149	0.088	0.146	0.087	0.071	0.147	0.083	0.182	0.047	0.243
LINE	**CRE**	**0.005**	0.002	0	0.002	0.001	0.003	0.001	0.001	0.001	0.003
LINE	**NeSL**	**0.033**	0.011	0.005	0.009	0.009	0.01	0.006	0.004	0.009	0.016
LINE	**Rex1**	0.002	0.003	0.003	0.002	0.001	0.009	0.001	0.025	0.008	0.002
LINE	**RandI**	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002
LINE	**Tx1**	**0.019**	0.007	0.001	0.007	0.006	0.014	0.007	0.005	0.006	0.012
LINE	**Crack**	**0.035**	0.006	0.001	0.005	0.007	0.021	0.007	0.006	0.006	0.018
LINE	**Nimb**	0.006	0.004	0.002	0.003	0.002	0.011	0.004	0.008	0.003	0.008
LINE	**Proto1**	0.005	0.001	0	0.002	0.002	0.005	0.001	0.001	0.002	0.004
LINE	**Proto2**	**0.003**	0.001	0	0.001	0.001	0.002	0.001	0.001	0.001	0.002
LINE	**Hero**	0	0	0	0	0	0	0	0	0	0
LINE	**Tad1**	0	0	0	0	0	0	0	0.003	0	0
LINE	**Ingi**	0.001	0.001	0.002	0.001	0	0.002	0.001	0.001	0.001	0.003
LINE	**Outcast**	**0.007**	0.002	0	0.001	0.002	0.004	0.003	0.002	0.004	0.006
LINE	**Daphne**	**0.002**	0.001	0	0	0	0.001	0	0.001	0	0.001
LINE	**Ambal**	0.001	0.001	0	0.001	0.001	0.001	0.001	0	0.001	0
LINE	**Vingi**	0.001	0	0	0.001	0.001	0.001	0	0	0	0.001
LINE	**I**	**0.05**	0.047	0.039	0.023	0.027	0.022	0.025	0.049	0.008	1.136

LINE	**DRE**	0.001	0.001	0.002	0	0	0	0	0	0.001	0.002
LINE	**telomeric**	0.007	0.004	0.012	0.004	0.003	0.013	0.003	0.003	0.002	0.016
LINE	**Unclassified**	**0.015**	0.006	0.006	0.005	0.008	0.013	0.006	0.005	0.006	0.038
SINE	**SINE1_7SL**	0	0	0	0	0	0	0	0	0	0
SINE	**SINE2_trna**	0.012	0.058	0.004	0.009	0.003	0.006	0.004	0.003	0.002	0.014
SINE	**SINE3_5S**	0.001	0.001	0.001	0.002	0.001	0.037	0.001	0.002	0.002	0.002
SINE	**SINE_MIR**	0	0.035	0.064	0.012	0.006	0.001	0.005	0	0.001	0
SINE	**SINE_B4**	0	0	0	0	0	0	0	0	0	0.006
SINE	**SINE_RTE**	0	0	0	0.001	0	0	0	0	0	0
SINE	**SINE_L1**	0	0	0	0	0	0.001	0	0.001	0	0
SINE	**SINE_R1**	0	0	0	0	0	0	0	0	0	0.013
SINE	**Unclassified**	0.023	0.017	0.006	0.027	0.048	0.022	0.05	0.018	0.009	0.024
Unclassified classI		**0.02**	0.008	0.002	0.007	0.01	0.016	0.012	0.01	0.007	0.02
TIR	**hAT**	**0.554**	0.219	0.188	0.251	0.403	0.501	0.447	0.399	0.152	0.392
TIR	**Mariner**	0.346	0.285	0.255	0.561	1.527	2.283	1.454	0.623	0.297	0.279
TIR	**MuDR**	**0.262**	0.072	0.027	0.053	0.06	0.167	0.059	0.05	0.071	0.157
TIR	**EnSpm**	**0.531**	0.256	0.277	0.302	0.236	0.487	0.253	0.324	0.153	0.65
TIR	**piggyBac**	0.032	0.014	0.009	0.006	0.01	0.042	0.01	0.017	0.023	0.018
TIR	**P**	**0.123**	0.049	0.071	0.021	0.025	0.08	0.031	0.063	0.02	0.076
TIR	**Merlin**	0.01	0.003	0.002	0.008	0.013	0.012	0.014	0.006	0.002	0.003
TIR	**Harbinger**	**0.073**	0.028	0.013	0.021	0.019	0.05	0.02	0.021	0.022	0.047
TIR	**Transib**	0.063	0.028	0.037	0.015	0.019	0.174	0.02	0.035	0.015	0.09
TIR	**Novosib**	0.002	0.001	0.001	0.001	0.001	0.009	0.001	0.001	0.001	0.007
TIR	**Mirage**	0.001	0.001	0	0	0.001	0.001	0	0	0	0.001
TIR	**Rehavkus**	**0.041**	0.028	0.02	0.009	0.007	0.024	0.009	0.034	0.005	0.084
TIR	**Kolobok**	0.058	0.031	0.034	0.019	0.023	0.045	0.034	0.078	0.006	0.278
TIR	**ISL2EU**	**0.011**	0.002	0.001	0.001	0.002	0.005	0.001	0.002	0.002	0.008
TIR	**Chapaev**	0.072	0.08	0.031	0.089	0.137	0.253	0.18	0.144	0.02	0.126
TIR	**Crypton**	**0.002**	0.001	0	0	0	0.001	0.001	0.001	0	0.035
TIR	**Sola**	0.105	0.117	0.101	0.034	0.083	0.788	0.099	0.187	0.027	0.212
TIR	**Zator**	0.009	0.002	0	0.002	0.002	0.005	0.002	0.003	0.002	0.006
TIR	**Ginger1**	0.039	0.015	0.004	0.007	0.01	0.035	0.011	0.007	0.009	0.026
TIR	**Ginger2/TDD**	**0.022**	0.004	0.001	0.003	0.003	0.009	0.004	0.004	0.005	0.01
TIR	**Academ**	0.011	0.008	0.026	0.001	0.004	0.008	0.004	0.018	0.002	0.024
TIR	**Other TIR**	0.867	1.001	1.369	1.168	3.764	2.283	4.034	2.808	0.359	0.824
MITE	**MITE**	0	0	0	0	0	0	0	0	0	0
Helitron	**Helitron**	**0.235**	0.099	0.078	0.127	0.08	0.181	0.09	0.071	0.07	2.938
Polinton	**Polinton**	0.325	0.158	0.224	0.822	0.509	0.408	0.543	0.226	0.051	1.452
Unclassified classII		0.005	0.11	0.036	0.012	0.017	0.103	0.032	0.019	0.006	0.042
SSR	**Simple repeat**	0.419	0.734	0.519	1.072	1.205	0.851	1.409	0.794	0.379	0.41
SSR	**Low complexity**	0.861	3.196	3.438	2.935	5.925	2.806	6.629	6.662	1.288	2.225
SSR	**Satellite**	0.01	0.042	0.019	0.007	0.006	0.011	0.006	0.015	0.004	1.747
SSR	**Other**	0.01	0.042	0.019	0.007	0.006	0.011	0.006	0.015	0.004	1.747
SSR	**Unclassified**	8.325	8.394	3.933	5.884	5.795	7.73	4.838	5.338	7.355	4.376
Unclassified		3.257	3.75	3.091	3.32	7.14	7.107	8.254	7.398	0.76	3.996

Supplementary Table S1.6: Enrichment of TE superfamilies in TE islands

Element type	Total bp in TE islands	Total bp in LDRs	Total number in TE islands	Total number in LDRs	FDR (base count)	FDR (element number)
Unclassified	1892769	4847567	9616	41643	>4.53E-155	>4.53E-155
TcMar-Tc1	84041	22152	515	164	>4.53E-155	>4.53E-155
DIRS	125721	30503	330	201	>4.53E-155	4.53E-155
RTE	138967	64262	309	502	>4.53E-155	5.68E-73
Ngaro	22579	5560	115	31	>4.53E-155	1.99E-72
TcMar-Mariner	9797	686	76	9	>4.53E-155	2.05E-56
Maverick	51886	64613	353	865	>4.53E-155	1.14E-49
LOA	18863	3596	67	53	>4.53E-155	4.80E-28
Kolobok-Hydra	6187	1190	20	7	>4.53E-155	4.24E-12
on	2804	1230	19	11	>4.53E-155	1.10E-09
Loa	38653	13148	42	93	>4.53E-155	1.57E-07
R1	121215	152620	286	1357	>4.53E-155	3.05E-07
BEL	553183	285447	560	2984	>4.53E-155	5.48E-07
Academ	5341	13540	52	167	0.033	3.75E-05
Merlin	6610	8835	41	120	>4.53E-155	5.32E-05
R2	20282	31169	37	139	>4.53E-155	0.007

Supplementary Table S1.7: List of duplication/deletion candidate loci and intersection with Cobs1.4 annotated genes

DUPLICATIONS										
Scf	**Start**	**Stop**	**Name;covBR;covJP;log2ratio;JPhet;exon bases;TE bases; island_binary**	**Het SNV**	**In**	**Affected gene**	**Type**	**Gene alias**	**Isl.**	**RNA seq**
scf0022	552001	553001	candidate_002;191;109;0.807;2;781;99;0	2	BR	Cobs_06524	Single exon duplication	upf0468_protein_cg5343-like	NO	**YES**
scf0022	552001	553001	candidate_002;191;109;0.807;2;781;99;0	2	BR	Cobs_06530	Single exon duplication	transcription_initiation_factor_tfiid_subunit_9-like	NO	**YES**
scf0072	9001	10001	candidate_004;121;64;0.922;0;1000;115;0	0	BR	Cobs_15626	Partial exon duplication	pin2-interacting_protein_x1	NO	**YES**
scf0028	384001	385001	candidate_007;65;130;-1.002546172;4;215;102;0	4	JP	Cobs_04037	Whole gene duplication		NO	**YES**
scf0030	1215001	1216001	candidate_007;96;54;0.820;4;373;193;0	4	BR	Cobs_05921	Multiple exon duplication	odorant_receptor_168	NO	**YES**
scf0055	661001	662001	candidate_014;81;268;-1.732135971;1;459;52;0	1	JP	Cobs_10648	Multiple exon duplication	hypothetical_protein_G5I_05212	NO	**YES**
scf0044	556001	557001	candidate_016;441;160;1.467;8;1000;0;0	8	BR	Cobs_02801	Partial exon duplication	Mucin-1	NO	**YES**
scf0010	108001	109001	candidate_037;368;128;1.522;9;273;214;0	9	BR	Cobs_00609	Multiple exon duplication		NO	NO
scf0003	617001	618001	candidate_045;175;94;0.900;0;128;0;1	0	BR	Cobs_14275	Multiple exon duplication		**YES**	NO
scf0003	1714001	1715001	candidate_046;139;78;0.839;6;903;0;0	6	BR	Cobs_14337	Whole gene duplication	85_kda_calcium-independent_phospholipase_a2-like	NO	**YES**
scf0003	1715001	1716001	candidate_047;228;98;1.210;10;1000;0;0	10	BR	Cobs_14337	Whole gene duplication	85_kda_calcium-independent_phospholipase_a2-like	NO	**YES**
scf0003	1716001	1717001	candidate_048;128;58;1.129;8;455;0;0	8	BR	Cobs_14337	Whole gene duplication	85_kda_calcium-independent_phospholipase_a2-like	NO	**YES**
scf0002	1474001	1475001	candidate_059;323;78;2.054;32;973;24;1	32	BR	Cobs_17748	Whole gene duplication		**YES**	**YES**
scf0002	2683001	2684001	candidate_075;99;54;0.887;7;347;346;1	7	BR	Cobs_17834	Multiple exon duplication	cytochrome_p450_4c1	**YES**	**YES**
scf0035	1363001	1364001	candidate_092;52;102;-0.95364953;14;406;343;0	14	JP	Cobs_04670	Multiple exon duplication	coiled-coil_domain-containing_protein_95	NO	**YES**
scf0035	1364001	1365001	candidate_093;66;132;-1.009649218;25;780;0;0	25	JP	Cobs_04689	Multiple exon duplication	sin3_histone_deacetylase_corepressor_complex_component_sds3-like	NO	**YES**
scf0001	4465001	4466001	candidate_094;80;20;2.000;5;313;243;1	5	BR	Cobs_07170	Multiple exon duplication	hypothetical_protein_EAI_00174	**YES**	**YES**
scf0001	4466001	4467001	candidate_095;148;48;1.628;4;628;59;1	4	BR	Cobs_07170	Multiple exon duplication	hypothetical_protein_EAI_00174	**YES**	**YES**
scf0042	2001	3001	candidate_098;119;217;-0.873101452;4;349;329;0	4	JP	Cobs_15501	Multiple exon duplication	nucleoside_diphosphate_kinase_7	NO	**YES**
scf0001	4823001	4824001	candidate_101;371;176;1.075;10;647;7;1	10	BR	Cobs_07201	Multiple exon duplication	hypothetical_protein_SINV_05299	**YES**	**YES**
scf0007	1468001	1469001	candidate_115;216;107;1.021;0;126;0;1	0	BR	Cobs_13418	Multiple exon duplication	odorant_receptor_13a	**YES**	NO
scf0007	1469001	1470001	candidate_116;213;99;1.107;0;366;0;1	0	BR	Cobs_13416	Whole gene duplication		**YES**	NO
scf0037	73001	74001	candidate_123;89;156;-0.812559297;10;129;0;0	10	JP	Cobs_00477	Whole gene duplication		NO	**YES**
scf0037	122001	123001	candidate_127;61;157;-1.372576506;6;452;153;0	6	JP	Cobs_00487	Whole gene duplication	nicotinic_acetylcholine_receptor_subunit_alpha_6_transcript_variant_partial	NO	**YES**
scf0037	122001	123001	candidate_127;61;157;-1.372576506;6;452;153;0	6	JP	Cobs_00483	Whole gene duplication		NO	NO
scf0037	126001	127001	candidate_128;59;150;-1.354662195;5;310;0;0	5	JP	Cobs_00482	Whole gene duplication		NO	**YES**
scf0037	952001	953001	candidate_129;49;101;-1.056522402;8;185;126;0	8	JP	Cobs_00543	Whole gene duplication		NO	NO
scf0037	953001	954001	candidate_130;53;100;-0.900691163;5;366;90;0	5	JP	Cobs_00543	Whole gene duplication		NO	NO
scf0018	2159001	2160001	candidate_143;58;104;-0.846853003;15;743;0;0	15	JP	Cobs_09630	Multiple exon duplication	hypothetical_protein_EAG_01487	NO	**YES**
scf0025	148001	149001	candidate_149;33;89;-1.441702554;20;274;165;0	20	JP	Cobs_06204	Whole gene duplication	hypothetical_protein_SINV_03739	NO	NO
scf0012	754001	755001	candidate_178;67;128;-0.935076915;11;535;291;1	11	JP	Cobs_15866	Multiple exon duplication	fatty_acid_synthase	**YES**	NO
scf0004	4177001	4178001	candidate_215;44;78;-0.840153732;25;533;0;0	25	JP	Cobs_05107	Whole gene duplication	isoform_a	NO	**YES**
scf0004	4177001	4178001	candidate_215;44;78;-0.840153732;25;533;0;0	25	JP	Cobs_05102	Whole gene duplication		NO	NO

scf0004	4178001	4179001	candidate_216;58;107;-0.890949904;31;791;0;0	31	JP	Cobs_05107	Whole gene duplication	isoform_a	NO	**YES**
scf0004	4178001	4179001	candidate_216;58;107;-0.890949904;31;791;0;0	31	JP	Cobs_05102	Whole gene duplication		NO	NO
scf0004	4179001	4180001	candidate_217;46;128;-1.456836393;38;848;78;0	38	JP	Cobs_05107	Whole gene duplication	isoform_a	NO	**YES**
scf0004	4179001	4180001	candidate_217;46;128;-1.456836393;38;848;78;0	38	JP	Cobs_05102	Whole gene duplication		NO	NO
scf0004	4180001	4181001	candidate_218;90;173;-0.952840484;34;1000;0;0	34	JP	Cobs_05107	Whole gene duplication	isoform_a	NO	**YES**
scf0004	4180001	4181001	candidate_218;90;173;-0.952840484;34;1000;0;0	34	JP	Cobs_05102	Whole gene duplication		NO	NO
scf0004	5302001	5303001	candidate_219;39;74;-0.936086412;0;542;200;0	0	JP	Cobs_05246	Whole gene duplication	iron-sulfur_cluster_co-chaperone_protein_mitochondrial-like	NO	**YES**
scf0009	45001	46001	candidate_224;59;112;-0.929677212;32;95;67;1	32	JP	Cobs_16853	Single exon duplication	cellular_nucleic_acid-binding_protein	**YES**	NO
scf0009	289001	290001	candidate_226;50;109;-1.114057336;12;395;50;1	12	JP	Cobs_16890	Whole gene duplication	adenylate_cyclase_type_10	**YES**	NO
scf0009	580001	581001	candidate_229;70;182;-1.371314387;33;729;0;1	33	JP	Cobs_16903	Whole gene duplication	hypothetical_protein_EAG_04423	**YES**	NO
scf0009	4113001	4114001	candidate_238;22;50;-1.194886622;0;185;141;0	0	JP	Cobs_17195	Whole gene duplication		NO	**YES**
scf0009	5349001	5350001	candidate_241;48;92;-0.942831796;10;529;251;1	10	JP	Cobs_17316	Whole gene duplication		**YES**	NO
scf0009	5638001	5639001	candidate_243;53;112;-1.081607453;9;821;410;1	9	JP	Cobs_17356	Whole gene duplication	PREDICTED:_uncharacterized_protein_K02A2.6-like	**YES**	NO
scf0010	155001	156001	candidate_252;43;115;-1.419704346;4;561;463;0	4	JP	Cobs_00607	Multiple exon duplication	hypothetical_protein_EAG_02592	NO	NO
scf0010	5129001	5130001	candidate_259;29;64;-1.157054559;0;153;34;0	0	JP	Cobs_01066	Whole gene duplication		NO	NO
scf0006	5610001	5611001	candidate_267;38;75;-0.971717664;0;106;56;0	0	JP	Cobs_09179	Multiple exon duplication		NO	NO
scf0008	2911001	2912001	candidate_276;46;102;-1.137478081;1;314;57;1	1	JP	Cobs_16467	Whole gene duplication		**YES**	NO
scf0008	3313001	3314001	candidate_278;33;66;-1.002599274;0;458;0;0	0	JP	Cobs_16545	Whole gene duplication	hypothetical_protein_SINV_06963	NO	**YES**
scf0008	3313001	3314001	candidate_278;33;66;-1.002599274;0;458;0;0	0	JP	Cobs_16554	Whole gene duplication		NO	NO
scf0003	85001	86001	candidate_282;47;81;-0.802449243;12;618;0;1	12	JP	Cobs_14245	Partial exon duplication	jerky_protein_homolog-like	**YES**	NO
scf0003	480001	481001	candidate_283;25;48;-0.923793684;0;207;151;1	0	JP	Cobs_14265	Multiple exon duplication	odorant_receptor_13a	**YES**	NO
scf0003	712001	713001	candidate_285;121;269;-1.155246806;8;502;42;1	8	JP	Cobs_14284	Multiple exon duplication	hypothetical_protein_G5I_09119	**YES**	NO
scf0003	943001	944001	candidate_286;35;63;-0.838878536;0;323;168;0	0	JP	Cobs_14297	Multiple exon duplication	transcriptional_regulator_atrx	NO	**YES**
scf0003	3184001	3185001	candidate_292;28;83;-1.546172902;9;365;116;1	9	JP	Cobs_14441	Whole gene duplication	hypothetical_protein_SINV_08869	**YES**	**YES**
scf0002	1440001	1441001	candidate_301;50;153;-1.605684692;0;448;174;1	0	JP	Cobs_17749	Whole gene duplication		**YES**	NO
scf0002	1441001	1442001	candidate_302;12;29;-1.293020781;0;212;113;1	0	JP	Cobs_17749	Whole gene duplication		**YES**	NO
scf0002	1723001	1724001	candidate_305;46;89;-0.937334261;7;177;120;1	7	JP	Cobs_17787	Multiple exon duplication		**YES**	NO
scf0002	2337001	2338001	candidate_307;88;346;-1.980760634;18;433;245;1	18	JP	Cobs_17885	Whole gene duplication	tyrosine_recombinase	**YES**	NO
scf0002	2813001	2814001	candidate_309;41;114;-1.470732135;11;632;163;1	11	JP	Cobs_17872	Whole gene duplication	PREDICTED:_hypothetical_protein_LOC100573698	**YES**	NO
scf0002	2814001	2815001	candidate_310;39;141;-1.849091768;12;527;99;1	12	JP	Cobs_17872	Whole gene duplication	PREDICTED:_hypothetical_protein_LOC100573698	**YES**	NO
scf0002	3224001	3225001	candidate_312;47;130;-1.477483194;4;871;0;1	4	JP	Cobs_17892	Whole gene duplication	gustatory_receptor_28b	**YES**	NO
scf0002	3432001	3433001	candidate_313;56;113;-1.01704099;32;462;0;1	32	JP	Cobs_17925	Multiple exon duplication	hypothetical_protein_EAG_00455	**YES**	NO
scf0002	3432001	3433001	candidate_313;56;113;-1.01704099;32;462;0;1	32	JP	Cobs_17927	Multiple exon duplication	hypothetical_protein_EAG_00088	**YES**	NO
scf0001	2003001	2004001	candidate_322;67;132;-0.96762421;11;791;0;1	11	JP	Cobs_07015	Whole gene duplication	serine_threonine-protein_kinase	**YES**	NO
scf0001	2003001	2004001	candidate_322;67;132;-0.96762421;11;791;0;1	11	JP	Cobs_07016	Whole gene duplication		**YES**	NO
scf0007	1234001	1235001	candidate_350;255;456;-0.839904468;9;279;0;1	9	JP	Cobs_13394	Whole gene duplication		**YES**	NO
scf0007	1463001	1464001	candidate_352;34;60;-0.833221343;0;340;127;1	0	JP	Cobs_13418	Multiple exon duplication	odorant_receptor_13a	**YES**	NO
scf0007	1999001	2000001	candidate_353;118;396;-1.750453075;0;182;109;1	0	JP	Cobs_13483	Multiple exon duplication	uncharacterized_aminotransferase_sso0104	**YES**	**YES**
scf0007	2000001	2001001	candidate_354;161;599;-1.89448593;0;417;162;1	0	JP	Cobs_13483	Multiple exon duplication	uncharacterized_aminotransferase_sso0104	**YES**	**YES**

scf0007	2000001	2001001	candidate_354;161;599;-1.89448593;0;417;162;1	0	JP	Cobs_13486	Whole gene duplication	nuclease_harbi1-like	**YES**	NO
scf0007	4178001	4179001	candidate_361;66;122;-0.899327359;6;709;183;0	6	JP	Cobs_13711	Whole gene duplication	hypothetical_protein_SINV_02031	NO	**YES**
scf0007	5028001	5029001	candidate_367;50;112;-1.161303072;32;351;0;1	32	JP	Cobs_13806	Whole gene duplication		**YES**	NO
scf0007	5029001	5030001	candidate_368;67;154;-1.197783686;34;364;0;1	34	JP	Cobs_13806	Whole gene duplication		**YES**	NO
scf0007	5030001	5031001	candidate_369;49;120;-1.286188148;65;386;356;1	65	JP	Cobs_13802	Whole gene duplication	pol-like_protein	**YES**	NO
scf0007	5092001	5093001	candidate_371;43;91;-1.069968987;18;318;227;1	18	JP	Cobs_13817	Whole gene duplication	hypothetical_protein_EAG_06404	**YES**	NO
scf0007	5093001	5094001	candidate_372;39;97;-1.312569542;16;479;335;1	16	JP	Cobs_13813	Multiple exon duplication	hypothetical_protein_EAG_00248	**YES**	NO
scf0007	5093001	5094001	candidate_372;39;97;-1.312569542;16;479;335;1	16	JP	Cobs_13817	Whole gene duplication	hypothetical_protein_EAG_06404	**YES**	NO
scf0007	5094001	5095001	candidate_373;47;110;-1.212024683;19;997;72;1	19	JP	Cobs_13813	Multiple exon duplication	hypothetical_protein_EAG_00248	**YES**	NO
DELETIONS										
Scf	**Start**	**Stop**	**Name;covBR;covJP;log2ratio;JPhet;exon bases;TE bases;island_binary**	**Gap (kb)**	**In**	**Affected gene**	**Type**	**Gene alias**	**Isl**	**RNA seq**
scf0093	9001	10001	candidate_003;111;58;0.935;0;1000;0;0	0.3	JP	Cobs_14758	Partial exon deletion	g-protein_coupled_receptor_mth2	NO	**YES**
scf0009	261001	262001	candidate_032;59;1;6.659;0;367;0;1	1.8	JP	Cobs_16872	Gene deletion		**YES**	NO
scf0009	466001	467001	candidate_034;77;1;5.689;1;382;195;1	4.6	JP	Cobs_16892	Gene deletion		**YES**	NO
scf0010	88001	89001	candidate_036;124;8;4.020;5;218;104;0	2.4	JP	Cobs_00602	Single exon deletion		NO	NO
scf0010	5244001	5245001	candidate_038;65;0;Inf;0;440;53;0	1.6	JP	Cobs_01070	Gene deletion	major_royal_jelly_protein_1	NO	**YES**
scf0010	5245001	5246001	candidate_039;47;0;9.410;0;426;256;0	1.3	JP	Cobs_01070	Gene deletion	major_royal_jelly_protein_1	NO	**YES**
scf0008	3060001	3061001	candidate_042;44;6;2.880;0;257;247;1	2.7	JP	Cobs_16510	Multiple exon deletion	fatty_acid_synthase	**YES**	NO
scf0003	307001	308001	candidate_044;37;0;10.053;0;324;124;1	1.3	JP	Cobs_14262	Multiple exon deletion		**YES**	NO
scf0003	3279001	3280001	candidate_052;67;22;1.593;2;215;46;1	3.1	JP	Cobs_14454	Gene deletion	hypothetical_protein_SINV_05682	**YES**	NO
scf0003	3288001	3289001	candidate_053;52;6;3.106;0;732;708;1	3.2	JP	Cobs_14460	Gene deletion	integrase_core_domain_protein	**YES**	**YES**
scf0003	3293001	3294001	candidate_054;17;0;5.489;0;346;218;1	4.9	JP	Cobs_14465	Gene deletion	bel12_ag_transposon_polyprotein	**YES**	NO
scf0002	1622001	1623001	candidate_061;59;4;4.011;0;74;70;1	5.9	JP	Cobs_17755	Gene deletion	vitellogenin_receptor	**YES**	NO
scf0002	1623001	1624001	candidate_062;71;6;3.507;1;274;0;1	3.7	JP	Cobs_17755	Gene deletion	vitellogenin_receptor	**YES**	NO
scf0002	1705001	1706001	candidate_063;44;0;7.241;0;287;124;1	4.3	JP	Cobs_17789	Gene deletion	period_circadian_protein	**YES**	**YES**
scf0002	2572001	2573001	candidate_073;48;14;1.840;0;771;225;1	2.4	JP	Cobs_17838	Gene deletion	zinc_finger_mym-type_protein_1-like	**YES**	NO
scf0001	1377001	1378001	candidate_078;77;0;7.584;0;951;0;1	2.0	JP	Cobs_06974	Gene deletion	zinc_knuckle_domain_protein	**YES**	NO
scf0001	1409001	1410001	candidate_083;52;6;3.001;0;180;51;1	0.7	JP	Cobs_06972	Single exon deletion		**YES**	NO
scf0001	1445001	1446001	candidate_085;91;4;4.640;5;120;43;1	3.5	JP	Cobs_06975	Multiple exon deletion	er_degradation-enhancing_alpha-mannosidase-like_3-like	**YES**	**YES**
scf0001	1861001	1862001	candidate_092;56;20;1.470;6;411;254;1	1.8	JP	Cobs_07000	Gene deletion		**YES**	NO
scf0001	4968001	4969001	candidate_106;70;0;12.194;0;521;103;1	3.6	JP	Cobs_07210	Single exon deletion	hypothetical_protein_SINV_09002	**YES**	NO
scf0007	1299001	1300001	candidate_112;31;4;3.044;0;467;100;1	2.5	JP	Cobs_13411	Partial exon deletion	hypothetical_protein_EAG_01487	**YES**	**YES**
scf0007	1423001	1424001	candidate_114;203;5;5.451;3;474;86;1	1.5	JP	Cobs_13414	Gene deletion	hypothetical_protein_SINV_00831	**YES**	NO
scf0007	2576001	2577001	candidate_125;95;0;12.213;0;780;123;1	2.6	JP	Cobs_13524	Gene deletion	tyrosine_partial	**YES**	NO
scf0007	2576001	2577001	candidate_125;95;0;12.213;0;780;123;1	2.3	JP	Cobs_13525	Single exon deletion	hypothetical_protein_EAI_02370	**YES**	NO
scf0007	2905001	2906001	candidate_129;66;16;2.039;0;273;45;1	2.3	JP	Cobs_13563	Single exon deletion	hypothetical_protein_EAG_07634	**YES**	NO
scf0007	5225001	5226001	candidate_130;42;0;11.347;0;616;0;1	0.6	JP	Cobs_13822	Multiple exon deletion	hypothetical_protein_GIP_L7_0050	**YES**	NO
scf0007	5226001	5227001	candidate_131;86;38;1.197;3;242;173;1	3.7	JP	Cobs_13822	Multiple exon deletion	hypothetical_protein_GIP_L7_0050	**YES**	NO
scf0003	2774001	2775001	candidate_288;243;484;-0.995002301;9;325;136;1	5.1	JP	Cobs_14410	Multiple exon deletion	reverse_transcriptase_and_recombinase	**YES**	**YES**

Supplementary Table S1.8: List of *de novo* assembled contigs containing an ORF

Contig ID	Length	Best blastx hit	Length of hit	Description	E value	Bit score	Frame	Query start	Query end	Hit start	Hit end	Positives	Identical
NODE_388976	14592	XP_003689693	1765	PREDICTED: LOW QUALITY PROTEIN: vitellogenin-like [Apis florea]	0	410	2	3416	5455	251	897	54.60%	35.10%
NODE_372516	1358	EGI70062	895	Transmembrane protein C9orf5 [Acromyrmex echinatior]	1.00E-83	283	1	527	1165	1	212	76.40%	65.70%
NODE_478441	1239	EGI65030	898	Sorting nexin-25 [Acromyrmex echinatior]	1.00E-76	204	2	314	634	31	137	96.30%	92.50%
NODE_426727	2425	EFZ16328	620	hypothetical protein SINV_06913 [Solenopsis invicta]	9.00E-70	190	2	540	1451	352	602	52.80%	44.00%
NODE_417388	855	EFZ16328	620	hypothetical protein SINV_06913 [Solenopsis invicta]	3.00E-56	152	0	424	747	123	231	85.30%	74.30%
NODE_396624	7803	EFN68490	1573	Peripheral-type benzodiazepine receptor-associated protein 1 [Camponotus floridanus]	3.00E-30	141	2	209	538	48	157	74.50%	70.00%
NODE_451872	579	EGI66069	1045	Protein toll [Acromyrmex echinatior]	9.00E-28	117	0	286	516	966	1045	78.80%	75.00%
NODE_241469	369	AEV76939	320	NADH dehydrogenase subunit 1 [Camponotus vafer]	9.00E-24	100	0	1	369	37	159	57.70%	43.90%
NODE_401423	643	XP_003706903	1705	PREDICTED: uncharacterized protein LOC100874905 [Megachile rotundata]	4.00E-20	96.7	2	381	536	34	85	98.10%	88.50%
NODE_443138	611	EGI57979	162	hypothetical protein G5I_13957 [Acromyrmex echinatior]	2.00E-15	78.2	0	135	389	55	142	67.80%	60.00%
NODE_352113	136	EFN84361	94	hypothetical protein EAI_03729 [Harpegnathos saltator]	5.00E-15	69.7	1	1	132	38	81	84.10%	68.20%
NODE_440382	6431	EFZ16363	78	hypothetical protein SINV_09927 [Solenopsis invicta]	8.00E-15	79.7	2	5119	5346	1	78	59.00%	57.70%
NODE_467846	151	EFZ11980	128	hypothetical protein SINV_01763 [Solenopsis invicta]	4.00E-12	63.5	1	1	147	23	71	73.50%	55.10%
NODE_593100	151	EFZ11980	128	hypothetical protein SINV_01763 [Solenopsis invicta]	5.00E-12	63.2	1	5	151	23	71	73.50%	55.10%
NODE_459514	226	EGI65596	271	hypothetical protein G5I_05988 [Acromyrmex echinatior]	8.00E-11	63.5	0	1	126	230	271	83.30%	66.70%
NODE_96922	191	EFN83463	257	hypothetical protein EAI_08541 [Harpegnathos saltator]	2.00E-10	62	0	19	189	136	192	66.70%	52.60%
NODE_144210	300	EFZ09717	207	hypothetical protein SINV_15650 [Solenopsis invicta]	7.00E-10	61.2	1	6	263	9	97	57.80%	38.90%

Supplementary Table S1.9: List of GO terms underrepresented in TE islands

GO-ID	Term	Cat	FDR	#Test	#Ref
GO:0005515	protein binding	F	1.74E-19	48	2533
GO:0016021	integral to membrane	C	4.65E-07	23	1159
GO:0003700	sequence-specific DNA binding transcription factor activity	F	9.48E-07	0	357
GO:0005667	transcription factor complex	C	2.38E-05	1	348
GO:0043565	sequence-specific DNA binding	F	4.50E-05	0	284
GO:0044430	cytoskeletal part	C	1.03E-04	1	317
GO:0007010	cytoskeleton organization	P	1.86E-04	0	250
GO:0006928	cellular component movement	P	2.89E-04	1	291
GO:0005524	ATP binding	F	3.47E-04	17	807
GO:1901566	organonitrogen compound biosynthetic process	P	0.001158612	1	260
GO:0071822	protein complex subunit organization	P	0.001619753	0	207
GO:0065008	regulation of biological quality	P	0.001856809	5	397
GO:0030234	enzyme regulator activity	F	0.002235494	1	245
GO:0050790	regulation of catalytic activity	P	0.002261041	1	248
GO:0015630	microtubule cytoskeleton	C	0.002286604	1	249
GO:0009966	regulation of signal transduction	P	0.004465379	3	307
GO:0009888	tissue development	P	0.004526085	3	310
GO:1901137	carbohydrate derivative biosynthetic process	P	0.004539984	0	182
GO:0007264	small GTPase mediated signal transduction	P	0.004660173	0	185
GO:0015031	protein transport	P	0.004823331	2	274
GO:0007267	cell-cell signaling	P	0.006042638	3	299
GO:0040007	growth	P	0.00628819	0	173
GO:0048667	cell morphogenesis involved in neuron differentiation	P	0.00628819	0	174
GO:0007017	microtubule-based process	P	0.00628819	1	227
GO:0048812	neuron projection morphogenesis	P	0.006382808	0	176
GO:0009887	organ morphogenesis	P	0.008118047	3	293
GO:0090407	organophosphate biosynthetic process	P	0.008811062	0	164
GO:0006468	protein phosphorylation	P	0.009516719	4	323
GO:0046907	intracellular transport	P	0.011008803	3	285
GO:0051128	regulation of cellular component organization	P	0.011653655	1	208
GO:0007276	gamete generation	P	0.012249983	1	212
GO:0015077	monovalent inorganic cation transmembrane transporter activity	F	0.012737248	0	160
GO:0004672	protein kinase activity	F	0.012778601	4	315
GO:0065003	macromolecular complex assembly	P	0.013013762	0	162
GO:0044723	single-organism carbohydrate metabolic process	P	0.013013762	0	162
GO:0048523	negative regulation of cellular process	P	0.015124635	6	372
GO:0005694	chromosome	C	0.01594876	1	200
GO:0009790	embryo development	P	0.01594876	2	236
GO:0034613	cellular protein localization	P	0.01594876	2	238
GO:0022402	cell cycle process	P	0.01594876	2	238
GO:0031090	organelle membrane	C	0.015981543	2	239
GO:0006184	GTP catabolic process	P	0.016245078	2	241
GO:0006091	generation of precursor metabolites and energy	P	0.017699389	0	152
GO:0061061	muscle structure development	P	0.017852228	0	153
GO:0012505	endomembrane system	C	0.018462732	0	155
GO:0044459	plasma membrane part	C	0.02196491	1	190
GO:0007444	imaginal disc development	P	0.022309481	2	233
GO:0008289	lipid binding	F	0.022309481	1	194
GO:0042623	ATPase activity, coupled	F	0.022565122	1	195
GO:0048522	positive regulation of cellular process	P	0.023079374	4	302
GO:0015672	monovalent inorganic cation transport	P	0.02523218	0	145
GO:0009791	post-embryonic development	P	0.030802838	2	226
GO:0006357	regulation of transcription from RNA polymerase II promoter	P	0.031210105	1	187
GO:0048610	cellular process involved in reproduction	P	0.031210105	1	187
GO:0042330	taxis	P	0.035500923	0	134
GO:0009069	serine family amino acid metabolic process	P	0.035923367	3	252
GO:0019226	transmission of nerve impulse	P	0.041678654	2	216
GO:2000026	regulation of multicellular organismal development	P	0.042831962	1	174
GO:0048646	anatomical structure formation involved in morphogenesis	P	0.043495522	1	178
GO:0031981	nuclear lumen	C	0.044951027	6	340

Supplementary Table S1.10: List of GO terms overrepresented in TE islands

GO-ID	Term	Cat	FDR	#Test	#Ref
GO:0003964	RNA-directed DNA polymerase activity	F	7.63E-52	53	16
GO:0006278	RNA-dependent DNA replication	P	7.63E-52	53	16
GO:0015074	DNA integration	P	6.74E-50	48	10
GO:0004190	aspartic-type endopeptidase activity	F	6.46E-15	18	9
GO:0004984	**olfactory receptor activity**	**F**	**3.23E-14**	**37**	**106**
GO:0005549	**odorant binding**	**F**	**5.00E-13**	**37**	**119**
GO:0003723	RNA binding	F	9.01E-12	64	386
GO:0050911	**detection of chemical stimulus involved in sensory perception of smell**	**P**	**6.36E-11**	**24**	**53**
GO:0007187	**G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger**	**P**	**2.62E-06**	**24**	**105**
GO:0006313	transposition, DNA-mediated	P	1.85E-05	7	4
GO:0004803	transposase activity	F	1.85E-05	7	4
GO:0004523	ribonuclease H activity	F	5.06E-05	8	9
GO:0019012	virion	C	7.83E-05	7	6
GO:0003968	RNA-directed RNA polymerase activity	F	3.78E-04	5	2
GO:0004482	mRNA (guanine-N7-)-methyltransferase activity	F	8.49E-04	5	3
GO:0005835	**fatty acid synthase complex**	**C**	**0.00712326**	**5**	**7**
GO:0006370	7-methylguanosine mRNA capping	P	0.01060855	5	8
GO:0016297	**acyl-[acyl-carrier-protein] hydrolase activity**	**F**	**0.03092985**	**4**	**6**

Supplementary Table S1.11: GLM of high aggression against intruding workers (intercept = BR x BR)

	Estimate	Std.	t value	Pr(>|t|)	
(Intercept)	0.24324	0.04470	5.442	8.91e-08	***
BR x JP	0.05526	0.06485	0.852	0.394568	
BR x *Waur*	0.37995	0.06435	5.904	7.22e-09	***
JP x BR	-0.24324	0.06301	-3.861	0.000131	***
JP x JP	-0.22991	0.06301	-3.649	0.000296	***
JP x *Waur*	0.28378	0.06322	4.489	9.21e-06	***

Null deviance: 87.265 on 433 degrees of freedom
Residual deviance: 63.287 on 428 degrees of freedom
AIC: 410.03

Supplementary Table S1.12: GLM of high aggression against intruding queens (intercept = BR x BR)

	Estimate	Std.	t value	Pr(>|t|)	
(Intercept)	0.546875	0.054254	10.080	< 2e-16	***
BR x JP	-0.005891	0.077664	-0.076	0.94	
JP x BR	-0.412547	0.075863	-5.438	1.23e-07	***
JP x JP	-0.373542	0.073860	-5.057	7.98e-07	***

Null deviance: 59.663 on 266 degrees of freedom
Residual deviance: 49.545 on 263 degrees of freedom
AIC: 317.99

Supplementary Table S1.13: Quantitative assembly statistics for the raw draft genome assembly

Scaffolded sequence (bp)	182 048 038
N50 scaffold size (bp)	2 570 857
Total number of scaffolds	11 084
N50 contig size (bp)	14 935

Supplementary Table S1.14: Draft genome sequences from other organisms used for gene annotation or comparative studies

Species	URL
Atta cephalotes	http://antgenomes.org/downloads/acep_scaffolds.fasta.zip http://antgenomes.org/downloads/acep_genome.OGS.1.2.gff.zip
Acromyrmex echinatior	http://antgenomes.org/downloads/aech/Aech_v2.0.fa.gz http://antgenomes.org/downloads/aech/Aech_v3.8.gff.gz
Camponotus floridanus	http://antgenomes.org/downloads/cflo_v3.3.fa.zip http://antgenomes.org/downloads/cflo_v3.3.gff.zip
Harpegnathos saltator	http://antgenomes.org/downloads/hsal_v3.3.fa.zip http://antgenomes.org/downloads/hsal_v3.3.gff.zip
Linepithema humile	http://antgenomes.org/downloads/arg_ant_scf4.fasta.zip http://antgenomes.org/downloads/lhum_genome.OGS.1.2.gff.zip
Solenopsis invicta	http://antgenomes.org/downloads/Si_gnF.454scaffolds.fasta.zip http://antgenomes.org/downloads/SI2.2.3.corrected.gff.zip
Pogonomyrmex barbatus	http://antgenomes.org/downloads/pbar_scaffolds_v03.fasta.zip http://antgenomes.org/downloads/pbar_genome.OGS.1.2.gff.zip
Apis mellifera	http://antgenomes.org/downloads/Amel_4.5.AGP.linearScaffold.fa.zip
Nasonia vitripennis	http://antgenomes.org/downloads/Nvit_2.0.linear.fa.zip

Supplementary Table S1.15: Differences in gene evolution across hymenopteran genomes, based on copy number differences within orthologous groups

	Present in all / missing	Single-copy in all / duplicated in
Cardiocondyla obscurior	78	251
Nasonia vitripennis	98	150
Apis mellifera	56	110
Harpegnathos saltator	50	76
Linepithema humile	18	67
Camponotus floridanus	32	67
Pogonomyrmex	14	65
Solenopsis invicta	155	92
Acromyrmex echinatior	22	30
Atta cephalotes	7	87

Supplementary Table S1.16: Reads generated per sample in the RNAseq experiment

Sample	Type	Raw read Count	Reads mapped to genes
QUI01	Imago	18 234 979	3 023 851
QUI05	Imago	24 958 424	2 635 246
QUI14	Imago	22 518 877	4 415 260
QUI16	Imago	17 864 237	3 803 684
QUI17	Imago	17 246 589	2 674 107
QUI51	Imago	23 405 068	3 810 613
QUI54	Imago	18 675 375	2 966 886
QUL65	Larva	20 406 972	4 572 339
QUL66	Larva	27 288 687	7 386 863
QUL68	Larva	29 225 761	3 192 075
QUL69	Larva	23 043 865	5 214 123
QUL73	Larva	23 071 947	5 296 047
QUL72	Larva	22 855 327	9 096 679
QUL71	Larva	26 443 413	5 567 634

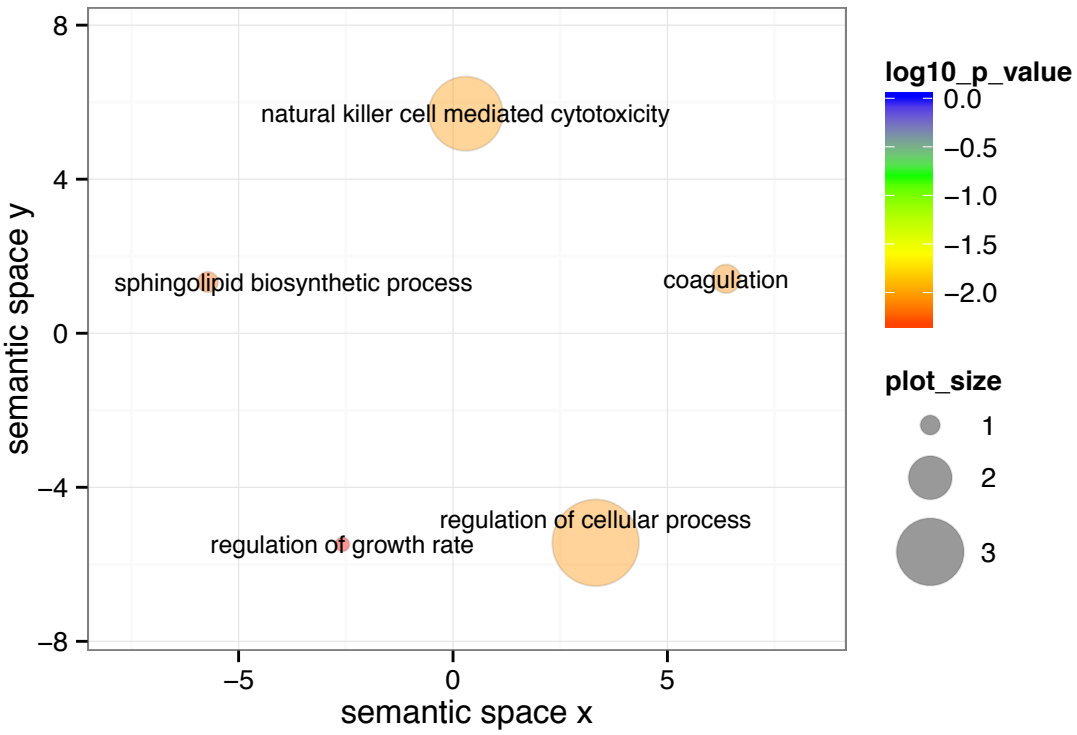
S1.2 Supplementary Material - Chapter 3

Sphingolipids, transcription factors, and conserved toolkit genes: Developmental plasticity in the ant *Cardiocondyla obscurior*

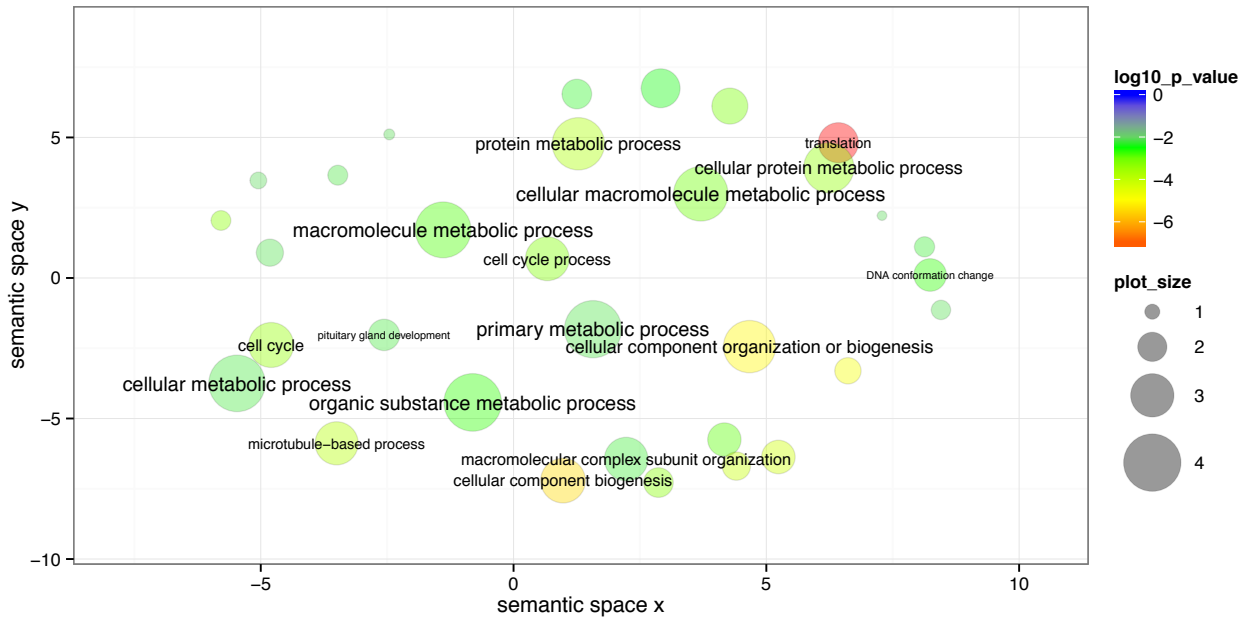
S1.2.1 Supplementary Data

Gene Ontology enrichment analysis

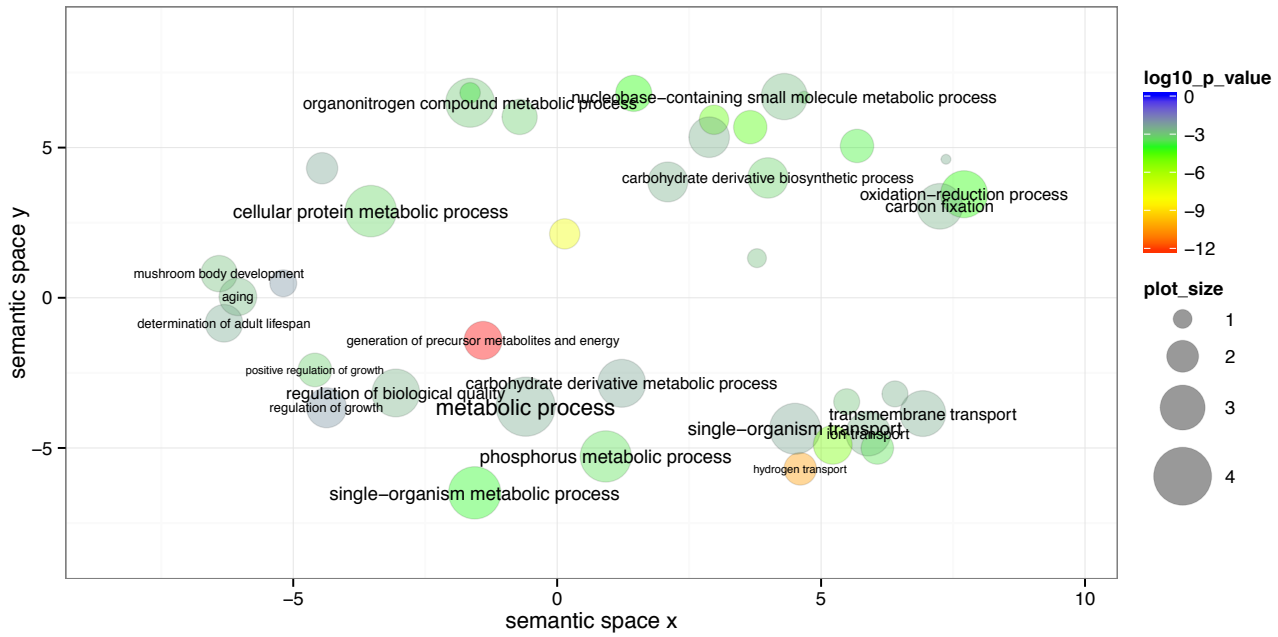
Gene Ontology (GO) functional annotation terms for all *Cardiocondyla obscurior* protein-coding genes have been published (Schrader et al. 2014) and can be downloaded from antgenomes.org. We used topGO (topgo.bioinf.mpi-inf.mpg.de) to test for enrichment of GO terms (in the domain Biological Process) in the caste-biased gene sets using Fisher's exact test and the Parent-Child-algorithm (Grossmann et al. 2007) and considered GO terms to be significantly enriched at $p<0.01$. We did not correct for multiple testing as the GO enrichment analysis was preliminarily performed for hypothesis generation. Visualization and summarization of the results was done with REVIGO (Supek et al. 2011). We retrieved five (WM), 39 (WO), 32 (QU) and 93 (EM) significantly enriched GO terms for the different gene sets (Supplementary Figures S2.1-S2.4).



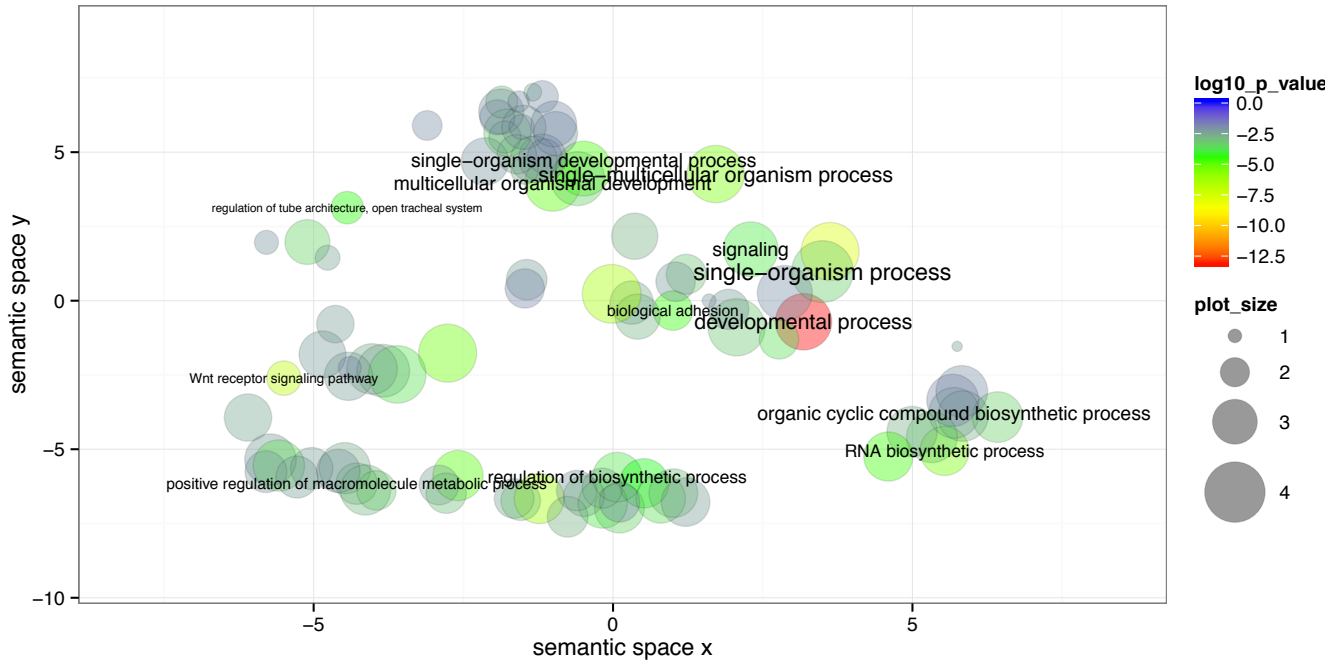
Supplementary Figure S2.1: GO enrichment analysis on the WM-biased gene set. The scatter plot shows GO terms in a two-dimensional semantic space after reducing redundancy with REVIGO. Circle sizes are relative to the frequency of GO terms in the data.



Supplementary Figure S2.2: GO enrichment analysis on the QU-biased gene set. The scatter plot shows GO terms in a two-dimensional semantic space after reducing redundancy with REVIGO. Circle sizes are relative to the frequency of GO terms in the data.



Supplementary Figure S2.3: GO enrichment analysis on the WO-biased gene set. The scatter plot shows GO terms in a two-dimensional semantic space after reducing redundancy with REVIGO. Circle sizes are relative to the frequency of GO terms in the data.



Supplementary Figure S2.4: GO enrichment analysis on the EM-biased gene set. The scatter plot shows GO terms in a two-dimensional semantic space after reducing redundancy with REVIGO. Circle sizes are relative to the frequency of GO terms in the data.

S1.2.2 Supplementary Tables

Supplementary Tables S2.1 to S2.4 accompanying Chapter 3 are available upon request and online at mbe.oxfordjournals.org/content/suppl/2015/02/27/msv039.DC1/supplementary_data.pdf .

The Supplementary Tables S2.1 to S2.4 provide detailed information about each gene in the four compiled gene sets (QU, WO, EM and WM). For each gene, the average expression levels in each of the four phenotypes and FDR values for pairwise comparisons are given in the tables.

Supplementary Table S2.5: Expression patterns of sphingolipid metabolic genes. The binary flag Orthology is set to 1, if orthology was confirmed in reciprocal BLASTs. The columns EM, QU, WM, and WO give the mean expression as assessed by limma. The last six columns give FDR values for pairwise comparisons, also produced with limma.

FB gene symbol	FB gene ID	FB gene name	*C.obs* ID	Orthology	EM	QU	WM	WO	EM vs WM	EM vs WO	EM vs QU	QU vs WO	QU vs WM	WO vs WM
bsk	FBgn0000229	*basket*	*Cobs_05382*	1	4.31	4.26	4.46	3.98	4.38E-01	7.05E-02	8.15E-01	1.38E-01	2.40E-01	1.85E-02
bwa	FBgn0045064	*brain washing*	*Cobs_11439*	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
bwa	FBgn0045064	*brain washing*	*Cobs_16616*	0	2.05	2.83	1.41	1.95	4.06E-01	9.99E-01	1.54E-01	1.67E-01	1.59E-02	4.62E-01
CDase	FBgn0039774	*Ceramidase*	*Cobs_12432*	1	4.35	5.80	4.64	4.93	6.90E-01	2.97E-01	7.43E-03	9.67E-02	1.86E-02	6.52E-01
CDase	FBgn0039774	*Ceramidase*	*Cobs_12439*	0	5.00	6.44	4.97	5.38	9.48E-01	4.92E-01	6.49E-03	4.21E-02	3.87E-03	4.92E-01
hep	FBgn0010303	*hemipterous*	*Cobs_07321*	1	5.75	5.95	5.53	5.67	1.11E-01	5.72E-01	1.50E-01	3.46E-02	1.75E-03	3.65E-01
ifc	FBgn0001941	*infertile crescent*	*Cobs_16634*	1	6.39	7.12	6.93	6.87	1.72E-01	1.92E-01	4.42E-02	5.17E-01	6.08E-01	9.23E-01
lace	FBgn0002524	*lace*	*Cobs_18257*	1	7.60	8.06	7.76	6.85	6.87E-01	2.47E-02	1.80E-01	7.74E-04	3.81E-01	1.71E-02
schlank	FBgn0040918	*schlank*	*Cobs_05570*	1	6.48	6.63	7.09	6.05	1.42E-02	6.24E-02	5.53E-01	1.29E-02	3.68E-02	3.36E-04
slpr	FBgn0030018	*slipper*	*Cobs_17593*	1	7.44	7.56	7.37	7.25	6.75E-01	1.40E-01	3.91E-01	1.62E-02	1.43E-01	4.24E-01
Sply	FBgn0010591	*Sphingosine-1-phosphate lyase*	*Cobs_00908*	1	6.54	7.43	7.56	7.73	5.36E-03	6.02E-04	7.64E-03	3.97E-01	7.31E-01	7.11E-01
egh	FBgn0001404	*egghead*	*Cobs_06699*	1	4.28	4.09	4.56	4.59	2.53E-01	1.63E-01	5.34E-01	3.81E-02	4.71E-02	9.13E-01
Sk2	FBgn0052484	*Sphingosine kinase 2*	*Cobs_17596*	1	5.41	6.77	5.73	4.98	2.56E-01	1.09E-01	6.87E-06	5.59E-07	1.27E-04	1.27E-02

S1.3 Supplementary Material - Chapter 5

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

S1.3.1 Supplementary Methods

Phylogenomic reconstruction

For phylogenetic placement of *Westeberhardia,* we used translated CDS sequences of *Westeberhardia* and followed the approach of (Husník et al. 2010). We constructed ortholog clusters using the proteomes from different Gammaproteobacteria 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 and Standley 2013). Alignments were refined using Gblocks v0.91b (Talavera and Castresana 2007) (Supplementary file protein_sequences.fasta available upon request). 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 ≤ 0.3, and all effective sizes of all summary variables were higher than 100.

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 5.1).

We used DNA material from a previous study (Oettler et al. 2010), and extracted DNA from additional samples using a chloroform-based method (Sambrook and Russell 2001). We performed PCR on the *nrdB1* (*ribonucleoside-diphosphate reductase 1 subunit beta*) gene of *Westeberhardia* (*WEOB_403*) (*nrdB1*for: 5'-GGAAGGAGTCCTAATGTTGCG-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 Cycler 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 *nrdB* gene to assess inter- and intrapopulational variation

To screen for intra- and interpopulational 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 *nrdB*.

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 $MgCl_2$, 3.3 µl H_2O 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_2O 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 sec; 60 °C 30 sec) followed by a melting curve with 0.5 °C temperature reduction every 5 seconds 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 *nrdB* 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 °C until DNA extraction.

Real-time quantitative PCR to quantify *Westeberhardia* titres

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 5.3b). All samples were sampled directly from stock colonies. Larvae and prepupae 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* titre on morph and age of an individual (Figure 5.3e, 5.3f).

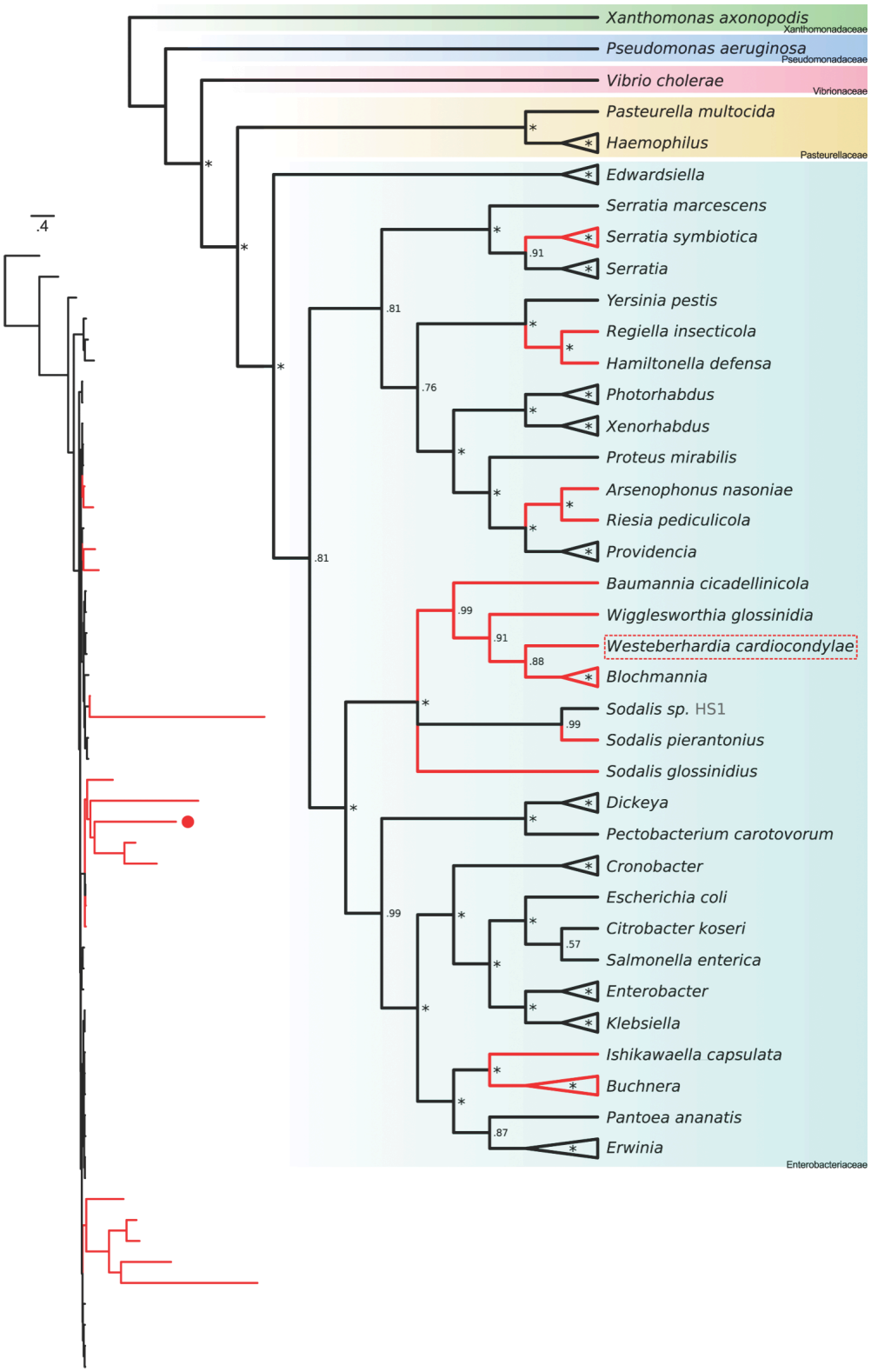
DNA of single ants was extracted using a chloroform-based method (Sambrook and Russell 2001), quantified fluorometically with Qubit® 2.0 Fluorometer (Life technologies) and diluted to 2.5 ng/µl. For each sample, we amplified a 204 bp fragment of the *nrdB* (*ribonucleoside-diphosphate reductase 1 subunit 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'-TGCCAACACCGTTCTGTCTG-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 housekeeping 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_2O up to 10 µl was added. Hard-Shell® 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™ Real-Time PCR Detection System (BioRad) with the following protocol: 95 °C 4 min; 41 x (95 °C 10 sec; 60 °C 30 sec) followed by a melting curve with 0.5 °C temperature reduction every 5 seconds from 95 °C to 65 °C. Analyses of C_q values were performed with a modified protocol of the $2^{-\Delta\Delta Cq}$ protocol (Livak and 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 ΔC_q value. To compare between all samples, an artificial calibrator with a C_q value of 40 was used to normalize all samples. The ΔCq value for the artificial calibrator was determined by subtracting the mean of all geometrical means (mean=21.38; SD=0.24) from C_q(calibrator) = 40, yielding ΔC_q(calibrator) = 18.72. For each target gene $\Delta\Delta C_q$ was calculated by subtracting ΔC_q(target) - ΔC_q(calibrator). Statistical tests were carried out with the ln transformed $2^{-\Delta\Delta Cq}$ values in R (version 3.0.2). 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. For non-parametric data we used Kruskal-Wallis tests or pairwise Mann-Whitney U-tests with post-hoc Benjamini-Hochberg correction. The boxplots (Fig. 3B, E, F) show ln transformed $2^{-\Delta\Delta Cq}$ 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'-ATCAGTTTCGAACGCCATTC-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_2O, air-dried, and finally covered with VectaShield® (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.

S1.3.2 Supplementary Figures



Supplementary Figure S3.1: Phylogenetic tree based on the phylogenomic analysis of *Westeberhardia* and other Enterobacteriaceae.

S1.3.3 Supplementary Tables

Supplementary Table S3.1: Pairwise T-tests using post-hoc Benjamini-Hochberg correction for intrapopulational comparisons in the BR Una (2012) population.

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

Supplementary Table S3.2: Pairwise Mann-Whitney U-tests using post-hoc Benjamini-Hochberg correction for morph comparisons of BR (Una 2012) individuals.

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

Supplementary Table S3.3: Pairwise t-tests using post-hoc Benjamini-Hochberg correction for BR (Una 2012) workers.

	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

Supplementary Table S3.4: Pairwise Mann-Whitney U-tests using post-hoc Benjamini-Hochberg correction for BR (Una 2012) queens.

	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	0.0028	-	-
V29	0.0082	0.0315	0.0188	0.0329	0.2029	-
Q48	0.0047	0.0065	0.0028	0.0017	0.0363	0.0047

S1.4 Supplementary References

Alasaad S, Rossi L, Maione S, Sartore S, Soriguer RC, Pérez JM, Rasero R, Zhu X-Q, Soglia D. **2008**. HotSHOT Plus ThermalSHOCK, a new and efficient technique for preparation of PCR-quality mite genomic DNA. *Parasitology Research* **103**: 1455–1457.

Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. **1990**. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and Environmental Microbiology* **56**: 1919–1925.

Anders S, Pyl PT, Huber W. **2015**. HTSeq - a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Benson G. **1999**. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* **27**: 573–580.

Blüthgen N, Brand K, Cajavec B, Swat M, Herzel H, Beule D. **2005**. Biological profiling of gene groups utilizing Gene Ontology. *Genome informatics. International Conference on Genome Informatics* **16**: 106–115.

Carlson DA, Bernier UR, Sutton BD. **1998**. Elution patterns from capillary GC for methyl-branched alkanes. *Journal of Chemical Ecology* **24**: 1845–1865.

Chen F, Mackey AJ, Vermunt JK, Roos DS. **2007**. Assessing performance of orthology detection strategies applied to eukaryotic genomes. *PLoS ONE* **2**: e383.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. **2012**. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; *iso-2*; *iso-3*. *Fly* **6**: 80–92.

Conesa A, Götz S. **2008**. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International journal of plant genomics* **2008**: 1–12.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. **2011**. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**: 491–498.

Edgar RC, Myers EW. **2005**. PILER: identification and classification of genomic repeats. *Bioinformatics* **21**: i152–i158.

Glastad KM, Hunt BG, Goodisman MAD. **2012**. Evidence of a conserved functional role for DNA methylation in termites. *Insect Molecular Biology* **22**: 143–154.

Grossmann S, Bauer S, Robinson PN, Vingron M. **2007**. Improved detection of overrepresentation of Gene-Ontology annotations with parent child analysis. *Bioinformatics* **23**: 3024–3031.

Holt C, Yandell M. **2010**. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics* **12**: 491–491.

Husník F, Chrudimský T, Hypša V. **2010**. Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches. *BMC biology* **9**: 87–104.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. **2005**. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* **110**: 462–467.

Katoh K, Standley DM. **2013**. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular biology and evolution* **30**: 772–780.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. **2013**. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**: R36.

Kohany O, Gentles AJ, Hankus L, Jurka J. **2006**. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC bioinformatics* **7**: 474.

Korf I. **2004**. Gene finding in novel genomes. *BMC bioinformatics* **5**: 59.

Langmead B, Salzberg SL. **2012**. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**: 357–359.

Lartillot N, Lepage T, Blanquart S. **2009**. PhyloBayes 3: a Bayesian software package for phylogenetic

reconstruction and molecular dating. *Bioinformatics* **25**: 2286–2288.

Li H. **2011**. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993.

Li H, Durbin R. **2010**. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589–595.

Libbrecht R, Oxley PR, Kronauer DJC, Keller L. **2013**. Ant genomics sheds light on the molecular regulation of social organization. *Genome Biology* **14**: 212.

Livak KJ, Schmittgen TD. **2001**. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* **25**: 402–408.

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

Lunter G, Goodson M. **2011**. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome research* **21**: 936–939.

Martin M. **2011**. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**: pp. 10–12.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. **2010**. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**: 1297–1303.

Nielsen R, Paul JS, Albrechtsen A, Song YS. **2011**. Genotype and SNP calling from next-generation sequencing data. *Nature Publishing Group* **12**: 443–451.

Oettler J, Suefuji M, Heinze J. **2010**. The evolution of alternative reproductive tactics in male *Cardiocondyla* ants. *Evolution* **64**: 3310–3317.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Oksanen MJ. **2013**. Package "vegan." *Community ecology package, version 2.9*.

Parra G, Bradnam K, Korf I. **2007**. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**: 1061–1067.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. **2005**. InterProScan: protein domains identifier. *Nucleic acids research* **33**: W116–W120.

Sambrook J, Russell DW. **2001**. *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.

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

Schrempf A, Heinze J. **2006**. Proximate mechanisms of male morph determination in the ant *Cardiocondyla obscurior*. *Evolution & development* **8**: 266–272.

Simola DF, Wissler L, Donahue G, Waterhouse RM, Helmkampf M, Roux J, Nygaard S, Glastad KM, Hagen DE, Viljakainen L, Reese JT, Hunt BG, Graur D, Elhaik E, Kriventseva EV, Wen J, Parker BJ, Cash E, Privman E, Childers CP, Muñoz-Torres MC, Boomsma JJ, Bornberg-Bauer E, Currie CR, Elsik CG, Suen G, Goodisman MAD, Keller L, Liebig J, Rawls A, Reinberg D, Smith CD, Smith CR, Tsutsui N, Wurm Y, Zdobnov EM, Berger SL, Gadau J. **2013**. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome research* **23**: 1235–1247.

Smith CD, Edgar RC, Yandell MD, Smith DR, Celniker SE, Myers EW, Karpen GH. **2007**. Improved repeat identification and masking in Dipterans. *Gene* **389**: 1–9.

Stanke M, Schöffmann O, Morgenstern B, Waack S. **2006**. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC bioinformatics* **7**: 62.

Supek F, Bošnjak M, Škunca N, Šmuc T. **2011**. REVIGO summarizes and visualizes long lists of Gene Ontology terms (C Gibas, Ed.). *PLoS ONE* **6**: e21800.

Talavera G, Castresana J. **2007**. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Systematic Biology* **56**: 564–577.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. **2008**. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome research* **18**: 1979–1990.

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. **3**: RESEARCH0034.

Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV. **2012**. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic acids research* **41**: D358–D365.

Zerbino DR, Birney E. **2008**. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* **18**: 821–829.