Cryptic Species Divergence and the Evolution of Cuticular Hydrocarbon Profiles in Parabiotic Ants

Dissertation

Zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Juliane Hartke

geb. am 20.08.1989

In Münster, Deutschland

Mainz, Februar 2020

Dekan: *Information removed for privacy purposes*

1. Berichterstatterin: *Information removed for privacy purposes*

2. Berichterstatterin: *Information removed for privacy purposes*

Tag der mündlichen Prüfung: 27. April 2020

TABLE OF CONTENTS

SUMMARY

The success of social insects is largely entwined with their elaborate communication system. Long-chained hydrocarbon molecules on the surface of insects make up specific profiles that function in species and nest-mate recognition, task allocation and mate-choice. Cuticular hydrocarbons (CHCs) thus facilitate the functioning of the sophisticated lifestyle of social insects, but aside from that also provide a barrier to pathogens and water loss. Due to this dual role in communication and ecological adaptation, CHCs are a strongly selected trait and are postulated to be drivers of rapid species divergence. This makes CHCs a particularly interesting trait to study, yet because of their dual role elucidating their genomic basis proved difficult in the past.

Here, I examine the genomic basis of cuticular hydrocarbon diversity and evolution in mutualistic ants and investigate the interplay between CHC profiles and local environmental parameters. I focus on the two species *Crematogaster levior* and *Camponotus femoratus* that inhabit a shared nest, a so-called ant-garden, in large parts of the Amazonian rainforest. Both species were found to show striking diversity in their cuticular hydrocarbon profiles that were postulated to be a sign for cryptic species. This species complex thus provides an ideal model to study the evolution of CHC profiles either within highly diverse species or sister species.

In the **first chapter** of this thesis, I elucidate species status of both *Cr. levior* and *Ca. femoratus* and show that both species split into morphologically nearly indistinguishable cryptic sister species (*Cr. levior* A and B, and *Ca. femoratus* PAT and PS) with exceptionally distinct CHC profiles. For this, morphometric measurements, DNA barcoding and the analysis of CHC profiles were combined in an integrative approach and provide three lines of conclusive evidence for the existence of cryptic species. I furthermore examine the distribution of the single species across the sampling range in French Guiana and find that in both *Crematogaster* species, the distribution is sympatric and independent of climate, in contrast to the two *Camponotus* species that show discrete distributions dependent on temperature and precipitation.

The **second chapter** focuses on a detailed characterization of the cuticular hydrocarbon profiles of all four species. We conclude that mutualisms might lead to longer chain lengths that are, due to their physical properties, not suited for communication and thus may represent a case of chemical insignificance to facilitate acceptance of the mutualistic partner. We furthermore compared CHC profiles between the species and found signals of the parabiotic partner within CHC profiles, possibly pointing to a form of chemical mimicry as an adaptation to the mutualistic lifestyle. We also found that climate had a significant influence on the CHC profiles of all species, underlining the importance of cuticular hydrocarbons in ecological adaptation.

As a basis for genome-wide analyses and to enable the identification of genes implicated in synthesis and detection of cuticular hydrocarbons, I sequenced the genome of one of the parabiotic ants, *Cr. levior* A. As described in **chapter three**, I used a combination of three different sequencing techniques and short and long reads to gain a highly accurate and contiguous depiction of the genome. Annotation of the genome showed an unusually high number of desaturases and a high number of elongases in this species, that might underly the increased chain lengths and specific adaptation to the mutualistic lifestyle. Comparisons to 40 other Hymenopteran species revealed a general expansion of both of those gene families within the order of Formicidae and an especially massive expansion of desaturases within *Cr. levior* A, which may be reflective of their complex mutualistic lifestyle.

The genome was used as a basis for genome-wide population genomic analyses in both *Cr. levior* species, as detailed in **chapter four**. By correlating population-specific allele frequencies to climate and chemical parameters, I identified candidate genes implicated in climate adaptation as well as the synthesis and detection of CHCs. Noticeably, different odorant receptors are implicated in climate adaptation, a correlation that has not been shown before. Additionally, odorant receptors were also correlated to population-specific CHC profiles, underlining their suggested importance in the elaborate communication system of social insects. Signatures of selection showed no overlap between the cryptic species despite shared selection pressures, confirming the patterns of haphazard evolutionary mechanisms found in other studies.

In conclusion, this thesis provides a closer look at a complex system of mutualistic ants. I present proof that both mutualistic partners encompass two cryptic species. I point to possible mechanisms that drove speciation in both cases, which most likely differed between the genera. Detailed analyses of signatures of selection among cryptic *Cr. levior* species show little overlap and solidify claims that evolution acts largely unpredictably. I identify candidate genes underlying communication and climate adaptation and thus provide a solid basis for future studies of the evolutionary trajectories of cuticular hydrocarbons and the intricate interplay of CHC synthesis, detection and their role in climate adaptation.

General Introduction

Juliane Hartke

"Seen in the light of evolution, biology is, perhaps, intellectually the most satisfying and inspiring science. Without that light it becomes a pile of sundry facts -- some of them interesting or curious but making no meaningful picture as a whole."

-- *Theodosius Dobzhansky*

1.1 SPECIES DIVERGENCE

The publication of Charles Darwin's "On the Origin of Species" was the birth of evolutionary biology by proposing natural selection and resulting species divergence as explanations for the observed biodiversity and repeated patterns of homology (Darwin 1859). Species divergence, or divergent evolution, is a process during which differences between closely related populations accumulate, that may in the end lead to speciation and reproductive isolation. There are different types of speciation, that can be separated into the two broader categories 'ecological speciation' and 'non-ecological speciation' (Sobel *et al.* 2009). The fist category encompasses scenarios during which organisms adapt to differential environmental selection pressures or to different ecological niches and was already described by Darwin (1859) together with sexual selection. Since then, many examples for this type of speciation have been described (Schluter 2001, 2009; Rundle & Nosil 2005; Shafer & Wolf 2013). In the past decades, also non-ecological types of speciation have been postulated and substantiated with examples from natural systems (Rieseberg & Willis 2007; Svensson 2012). Non-ecological speciation can be separated into mutation-order, genetic drift and polyploid speciation (Sobel *et al.* 2009). The latter has often been found in plants, in which duplications of the genome lead to reproductive isolation, since polyploids are often unable to reproduce with diploid individuals (Wood *et al.* 2009). Details on ecological and non-ecological speciation that are relevant to this thesis will be discussed in more detail below in 1.1.1 and 1.1.2.

In general, ecological as well as non-ecological speciation can occur between populations that are separated by geographical barriers (*allopatry*), between populations that occur in adjacent habitats (*parapatry*), or between populations that share the same habitat (*sympatry*) (Schluter 2001; Mallet *et al.* 2009). The main difference between these modes of speciation is the potential amount of *gene flow* between diverging populations (Svensson 2012). Gene flow is the transfer of genetic variation from one population to another and thus increases variation within populations. It is thought to constrain species divergence, as it leads to a homogenization of the gene pool between populations (Petit & Excoffier 2009). Physical geographical barriers such as mountain ranges, oceans etc., but also reproductive barriers restrict gene flow. Gene flow can sometimes also facilitate adaptation by introgression of favourable alleles. In modern humans for example, adaptation to the hypoxic environment of the Himalayas was enabled by haplotype introgression from the Denisovan genome that still leads to a lower haemoglobin percentage in today's Tibetan population (Huerta-Sánchez *et al.* 2014). Adaptive introgression between two species of *Anopheles* mosquitoes has recently been shown to facilitate insecticide resistance (Norris *et al.* 2015), and in *Heliconius* butterflies, hybridization between three species leads to the exchange of mimicry colour patterns (Dasmahapatra *et al.* 2012). At the same time, gene flow can increase the risk

of maladaptation when individuals immigrate from different environments and introduce alleles that are detrimental to the local populations' fitness (Nuismer *et al.* 2012).

It has often been speculated whether speciation in sympatry without barriers to prevent gene flow is possible. One of the proposed drivers of speciation in sympatry is *sexual selection* (e.g. Lande 1981; West-Eberhard 1984; Higashi *et al.* 1999). This mode of selection has already been presented by Darwin (1859) and is increasingly given attention as a possible mechanism to prevent gene flow during speciation without any geographical barriers (Higashi *et al.* 1999; Kirkpatrick & Ravigne 2012). The theory of sexual selection describes that mating success is dependent on morphological or behavioural traits that will be chosen over one another by members of the opposite sex (Clutton-Brock 2009). These traits can for example be antlers in deer (Gould 1974), bright colours or conspicuous ornaments in some birds (Lande 1981), or loud mating calls in e.g. frogs (Rand 1985). Often, those traits are costly in terms of energy and detrimental to the individual's survival ability, but still grant a higher level of fitness due to increased mating success (Lande 1981). Sexual selection thus has the potential for rapid disruptive selection acting on morphological traits and subsequently causing prezygotic reproductive barriers (Higashi *et al.* 1999).

1.1.1 Ecological speciation

Darwin suggested contrasting selection between environments as one of the drivers of speciation (Darwin 1859). Most of the examples for this mode of species divergence include a scenario in which a part of a population is confronted with a new habitat that poses new selection pressures, e.g. after migration events. Given enough time and limited gene flow between the newly established and the original population, they will accumulate differences to an extent where they become reproductively isolated, and thus two species are formed. Ecological speciation by niche differentiation is a broadly accepted concept and has been described in several instances (Schluter 2001, 2009; Rundle & Nosil 2005; Shafer & Wolf 2013). Adaptive radiation in Darwin's finches is one of the prime examples. This group consists of various species that show remarkable diversity in beak morphology, matching their specific food niches on separate islands (Grant & Grant 2003; Podos & Nowicki 2004). Another example can be observed in threespine sticklebacks, that evolved freshwater ecotypes from their ancestral marine state (Taylor & McPhail 2000). The speciation event between polar and brown bear is also likely due to ecological selection. It is postulated that a brown bear population inhabited the arctic regions during an extended glacial minimum. When that population became separated by growing ice sheets, climatic and environmental change drove adaptations to the new environment that in the end led to speciation between the brown bear and the new polar bear population (Liu *et al.* 2014).

Adaptation due to ecological selection pressures can be influenced by a multitude of variables from an organism's habitat. Altitude for example, has led to adaptation in haemoglobin percentages in many montane species, including humans, that allows them to survive in hypoxic conditions (Weber 2007; Huerta-Sánchez *et al*. 2014). Water availability, but especially the absence of water is another factor that promotes adaptation. Desert inhabiting species show a multitude of behavioural and morphological characteristics intended to keep water loss to a minimum (Randall 1993; Wu *et al.* 2014). Where water is available, salt concentration may still play a role in adaptations. Mangroves grow in the brackish water of intertidal zones. To cope with the high salinity level some species developed ultra-filtration systems within their roots, others actively secrete excess salt through special glands on their leaves (Liang *et al*. 2008). These adaptations allowed mangrove species to grow in habitats that are uninhabitable for other species, and therefore minimize competition.

An important selection pressure is the *ecological interaction* with other species within a community (Schluter 2001). These interactions can be classified into different categories depending on their either neutral, negative or positive outcome for the involved players (Abrams 1987). Interactions that are either negative or positive for both participants can be classified as competition and mutualism respectively (Abrams 1987). Predation classically involves a positive outcome for one party and a negative outcome for the other, similar to parasitism (Chamberlain *et al.* 2014). Such close interactions between species can lead to coevolution, a process during which evolutionary changes are being prompted reciprocally between two or more species (Hoeksema & Bruna 2000; Guimarães *et al*. 2011). This can for example be seen in host-parasite relationships that exert selection pressures on both parties due to continuous conflict. To avoid negative consequences from parasites, hosts should develop defence mechanism, which in turn will pose selective pressure on the parasite to overcome the hosts defences. This kind of interplay is described in the Red Queen hypothesis, that assumes a faster evolutionary speed in parasite-host relationships due to the enhanced selective pressures (Van Valen 1973). Also in mutualists, species are closely interacting and may co-evolve traits positively affecting the interaction (Guimarães *et al*. 2011). Similar to parasitism, this close evolution is corroborated on the genomic level with a high mutation rate (Rubin & Moreau 2016), as the persistent and often close interactions between mutualists lead to additional selection pressures on individuals.

Another consequence of such close interactions can be co-speciation, a process by which a speciation event in one species (host or mutualistic partner) can initiate a speciation event in the other species. According to *Fahrenholz's* rule (Fahrenholz 1913), co-speciation can be frequently observed in host-parasite and mutualistic interactions, reflected in the similarity of species' phylogenies. This has for example been observed in an ant-plant mutualism

between Asian tropical *Crematogaster* ants and *Macaranga* plants (Itino *et al*. 2001), and also between south American *Acacia* plants and *Pseudomyrmex* ants (Janzen 1966), where both phylogenies closely mirror each other. Opposing to *Fahrenholz's* rule, Janzen rather attributes this pattern to ecological fitting rather than co-speciation, postulating that closely related species will share traits due to a common evolutionary history that allows them to be associated with similar species (hosts or mutualistic partner) (Janzen 1979, 1985). This theory also accounts for frequently observed host switches. Nonetheless, co-speciation is still a plausible scenario under close parasitic or mutualistic associations, especially when these associations are obligate and partner switches are not possible.

In addition to interspecific interactions, also within-species interactions can pose selection pressures and in the end lead to population divergence and speciation. In cichlid fish for example, competition for resources led to speciation and to the development of different ecomorphs in Lake Tanganyika (Winkelmann *et al.* 2014). Density dependent diversification has also been shown in *Drosophila* (Bolnick 2001) and the North American wood warbler (Rabosky & Lovette 2008).

1.1.2 Non-ecological speciation

While in the case of ecological adaptation, the direction of selection is shaped by external factors, another mode of species divergence can be ascribed to stochastic processes, during which the frequency of different mutations will fluctuate independent of ecological factors and even under the same selection pressures. These processes include genetic drift and mutation-order (Mani & Clarke 1990). *Genetic drift* describes that the mode of inheritance of a certain allele from one generation to the next is random and consequently, by stochastic processes alone, the frequencies of alleles in a population will vary over time (Charlesworth 2009). In the absence of selective pressures, also neutral alleles can drive to fixation just by chance (Mallet *et al*. 2009). The *mutation-order* principle also postulates that reproductive isolation can be facilitated by the random occurrence of alleles that will then rise to fixation (Schluter 2009). It is distinctly different from genetic drift, in that the underlying principle here is that when regarding two populations, even under the same selection pressures, the occurrence of the same mutations in the same temporal order are highly unlikely, which will then lead to evolutionary constraints, e.g. epistasis (= *effect of genetic background on the phenotypical effect of a gene*) (Schluter 2009).

When neglecting selection, the time to fixation of an allele by purely stochastic processes highly depends on the effective population size (N_e). In small populations alleles can rise to fixation much quicker within a few generations than is possible in populations with a large N_e (Charlesworth 2009). Of course, selection, mutation-order and genetic drift are not mutually exclusive and will most certainly go hand in hand, with each process influencing

the other ones (Sobel *et al.* 2009). Due to this fact it is hard to determine the underlying causes for species divergence in natural systems.

1.1.3 Sources of genomic variation and prerequisites to speciation

Adaptation of any kind is always founded on random mutations within the genome, in the form of either standing genetic variation or new mutations (Barrett & Schluter 2008; Lai *et al.* 2019). Within a given population of multicellular organisms no individual is genetically identical to any other individual due to mistakes during DNA replication or repair that lead to point mutations, insertions or deletions, which increases variation within a population (Goriely 2016). In addition to mutations, also genetic recombination during meiosis is a source of genetic variation (Modesti & Kanaar 2001), as is gene flow from other populations (Barrett & Schluter 2008). In general, the chances of a population of being able to adapt to changing conditions increases with the number of available alleles (= *standing genetic variation*) within the population (Nuismer *et al.* 2012). Meaning, a population with a high level of genetic variation will be more likely to already harbour a useful mutation to e.g. better access resources, deal with rising temperatures and so forth. This intuitively makes sense. A population that is perfectly adapted to e.g. current temperature and possesses a very low level of standing genetic variation, will almost certainly go extinct if temperature levels rise above the population's thermal tolerance.

Most phenotypical traits are not based on a single gene, but on multiple genes at once (polygenic trait) (Barghi *et al*. 2019). There are, however, several prominent exceptions to this. Sickle cell disease in humans, for example is caused by a single point mutation that changes the hydrophilic glutamic acid to the hydrophobic amino acid valine and thereby transforms the shape of red blood cells to the typical name-giving sickle shape (Ingram 1957). In heterozygous individuals, this confers some resistance to malaria that is especially beneficial in infants (Aidoo *et al*. 2002). Another prominent point mutation is responsible for the unusual epidemic Ebola outbreak in West Africa from 2013-2016 by facilitating adaptation to the human host and promoting virulence (Bedford & Malik 2016). And in *Anopheles gambiae*, a single point mutation provides the mosquito with resistance to the insecticide DDT (Ranson *et al*. 2000). Aside from mutations, also transposable elements can lead to adaptive change. Such is the case in peppered moths, where the switch from light coloured variant to dark morphs was enabled by the insertion of a transposable element into the gene *cortex* (van't Hof *et al.* 2016).

1.1.4 Parallel evolution

Phenotypical traits will often develop in parallel between species that inhabit similar habitats or that are subject to the same selection pressures (MacPherson & Nuismer 2017). When the species that show similar traits are only distantly related, this is called

convergence (e.g. wings of bats and birds), while in closely related species this phenomenon is termed parallel evolution (Elmer & Meyer 2011). Examples for parallel evolution in the literature are primarily found in threespine sticklebacks that evolved similar phenotypes as adaptation to freshwater habitats (Colosimo *et al*. 2005). The likelihood of parallel evolution is hard to determine and largely depends on the level of hierarchy of the traits. It decreases from the phenotypic level to the lowest order of hierarchy, a change in the actual gene or even a specific change in nucleotides (Bailey *et al*. 2015). Thus, witnessing the same genetic patterns between species seems rather unlikely. However, several factors can increase the chance of parallel evolution on the genomic level. A close phylogenetic relationship, and as follows, a similar genomic background is one of the most important factors (Bailey *et al*. 2015), as the genomes of more distantly related organisms will have accumulated more mutations and might be subject to evolutionary constraints (Kaeuffer *et al.* 2011). Furthermore, the closer the species are to a trait optimum, the higher the chance for parallel changes, as the number of possible beneficial mutations decreases with increasing fitness (Bailey *et al*. 2015). The strength of natural selection also seems to play a role, as strong selective pressures will increase the need for beneficial mutations (Orr 2005; MacPherson & Nuismer 2017). In general, following these suggestions from current literature, the highest likelihood for parallel genomic changes should be found in closely related species, that are subject to exactly the same strong selective pressures, but that are already close to their trait optimum.

So far, studies on this topic have shown mixed results. While some found clear evidence for repeatable changes in many specific SNPs (Haenel *et al*. 2019), or in at least some SNPs (Chaturvedi *et al*. 2018), other studies found only a small number of parallel adaptive changes on a functional level (Dennis *et al.* 2015; Pfenninger *et al*. 2015), and a study on ants found no signs for parallel evolution regardless of hierarchical level (Feldmeyer *et al.* 2017). Reasons for the latter could either be evolutionary constraints that differ between species, or the opposite: genomic redundancy, where many changes in different nucleotides, genes or metabolic pathways are possible that in the end lead to the same adaptive value (Barghi *et al*. 2019).

1.2 LOCAL AND CLIMATE ADAPTATION

Species do not occur just at a single point in space, but instead have a certain distribution range that can show variation in selection pressures dependent on e.g. altitude and latitude (Klepsatel *et al.* 2014). As a result, species often will encounter differential selection pressures across their distribution range, to which local populations should adapt to increase fitness. Abiotic factors, such as water and light availability, temperature and soil

can be highly variable with altitude and latitude (Halbritter *et al.* 2013). Light availability for example has been shown to influence foraging behaviour in fire salamanders (Manenti *et al*. 2013). Also, biotic factors, such as food availability, competitors and predators can vary across distribution ranges and can prompt local adaptation (Sexton *et al*. 2009; Lewis *et al*. 2017). In aposematic frogs, for instance, colouration and activity patterns vary with predator density (Kang *et al.* 2017). Variance based on predatory pressure was also found in guppies, where females mature younger and mate more frequently in low-predation habitats (Torres Dowdall *et al.* 2012).

The most important factors to which populations have to adapt, are climate parameters such as temperature, precipitation, wind speed and air pressure, and the daily and seasonal changes thereof (Klepsatel *et al.* 2014). Many morphological changes that evolved as a response to climate factors are obviously identifiable, such as thicker fur, bigger body sizes and thick layers of blubber in arctic regions (Blix 2016), or huge ears in several desert-inhabiting species (Maloiy *et al*. 1982). Aside from directly visible morphological changes, animals will also adapt internal processes, i.e. metabolism, as a response to external climatic conditions. In insects for example, heat shock proteins facilitate acclimation to heat and cold (Rinehart *et al.* 2007). As the climate impacts many other factors, such as resource availability, within an individuals' environment (Walther *et al.* 2002), it is often hard to disentangle whether any adaptation is in fact a response to a change in climate or a response to another factor that is influenced by climate.

Climate parameters are particularly prone to vary across a species distribution range. Changes can occur with shifting latitude or altitude, which in particular influences temperature and precipitation (Halbritter *et al.* 2013). In theory, populations from either end of an extended distribution range that are shaped by different climates, should show different trait optima specific to local conditions. This has indeed been found in several species such as threespine and ninespine sticklebacks that show contrasting adaptations in response to temperature differences between marine and freshwater habitats (Bruneaux *et al.* 2014; Gibbons *et al*. 2017). Climatic differences across the distribution range also influence the American pika that shows adaptive genomic divergence dependant on temperature and precipitation (Henry & Russello 2013).

1.2.1 Insects: Adaptation and Acclimation

In general, organisms can react to changing conditions either by acclimation or adaptation. Acclimation refers to a short-term and reversible physiological response that is induced by an organism's environment and can involve morphological, behavioural or biochemical changes (Chaffee & Roberts 1971). Adaptation on the other hand describes a long-term, heritable evolutionary response that often occurs gradually over many generations (Reeve &

Sherman 1993). In contrast to endotherms that can regulate their body temperature, ectotherms, such as insects, are particularly vulnerable to changes in ambient temperature (Deutsch *et al.* 2008) that will directly influence several important functions, for example locomotion or metabolism (Lehmann 1999; Seebacher 2009). A recent study on the non-biting midge *Chironomus riparius* for example reflects this by showing that more than 1% of the annotated genes are implicated in climate adaptation, with many of them functioning in pathways such as cell division, transcription and translation (Waldvogel *et al.* 2018). Also in *Drosophila melanogaster*, several studies show the profound influence of climate clines on population divergence: populations along Australia's East coast and America's West coast show a genomic inversion polymorphism (Kolaczkowski *et al.* 2011; Bergland *et al*. 2016), which has been linked to clinal variation in body size and thus might indicate latitude-dependent climate adaptation (Rane *et al.* 2015).

1.3 CUTICULAR HYDROCARBONS

Adaptation and acclimation to ambient temperature in insects, is achieved on multiple levels. Insects can modify their metabolism, physiology and behaviour (Enriquez & Colinet 2019) as a response to changes in temperature. Because of their high surface-to-volume-ratio, insects are particularly vulnerable to water loss that mainly occurs through the cuticle (Menzel *et al*. 2018). Modifications to the cuticle are thus one of the responses to desiccation stress (Hadley 1981). The upper cuticular layer, the epicuticle, of virtually all terrestrial arthropods is made up of a thin layer of cuticular hydrocarbons (CHCs), hydrophobic molecules, that, by tight aggregation, prevent desiccation (Blomquist & Bagnères 2010). The adaptive value of CHCs in response to climate has been shown in several insect species and the observed modifications in CHC profiles were often parallel between different species: Two *Drosophila* species showed the same shift from longer to shorter chained hydrocarbons with increasing latitude, i.e. a shift towards cooler temperatures (Frentiu & Chenoweth 2010). The same patterns were found in the comparison of two *Myrmica* ant species (Sprenger *et al*. 2018) and two *Temnothorax* ant species (Menzel *et al.* 2018) in response to cooler temperatures under laboratory conditions. Studies such as the one from Frentiu and Chenoweth (2010) show that CHC profiles not only differ between species, but that substantial differences can also be found between individuals within the same species. These intraspecific differences can mostly be explained by climate selection pressures, that vary across the distribution range (Rajpurohit *et al.* 2017).

While the studies mentioned so far show that constant climatic conditions pose selection pressures on CHC profiles in insects, individuals also have to deal with fluctuating conditions over the span of a single day. CHCs not only enable adaptation to different

climates, but also allow for short-term acclimation by shifting the relative quantities of single substances and thereby changing the level of viscosity and permeability of the waxy layer (Wagner *et al*. 2001; Stinziano *et al*. 2015; Menzel *et al.* 2018; Sprenger *et al.* 2018). These processes have been shown to occur over relatively short time spans (Hoffmann 1990; Stinziano *et al*. 2015) and thereby allow individuals to, for example, tackle the differences between night and day or summer and winter.

1.3.1 Cuticular hydrocarbons and their role in recognition

Aside from climate adaptation, CHCs also evolved secondary functions. As highly elaborate communication agents, they enable the distinction between species, colonies and castes (Blomquist & Bagnères 2010) and facilitate mate-choice (Otte *et al*. 2018), nest-mate discrimination (Buchwald & Breed 2005) and task allocation (Greene & Gordon 2003), and thus are at the basis of the complex social structure displayed by social insects. Individuals can change their cuticular hydrocarbon profile as a direct response to social stimuli, e.g. after losing a fight (Thomas & Simmons 2011), or in courtship (Petfield *et al*. 2005). This double role of CHCs makes it even more complicated to pinpoint the causes of alterations in the profiles. Whether a shift in substance abundance was caused by environmental factors, such as climate, or by social factors is hard to determine. Several studies therefore see CHCs as a so-called "magic trait", i.e. a trait that is both affecting ecological adaptation and mate-choice (Smadja & Butlin 2009). A change in such a trait as an adaptive response to ecological factors can cause sexual selection and ultimately drive speciation (Chung & Carroll 2015).

1.3.2 Genetic basis of cuticular hydrocarbons

To facilitate the response to changing conditions within a single lifespan or even within a single day, the genetic architecture underlying cuticular hydrocarbons needs to provide a certain level of flexibility. This points to a rather complex genetic basis with the involvement of multiple gene families that allow for minute changes at several points of the pathway (Menzel *et al.* 2017b). And indeed, several different gene families have been shown to influence the CHC profile in insects. The main genes that are implicated to be involved in the synthesis of CHCs are elongases, that catalyse carbon chain extension (Blomquist & Bagnères 2010), and desaturases that implement double bonds (Dallerac *et al*. 2000). Aside from these two families, other genes seem to play a role as well. Cytochrome P450 genes are involved in decarbonylation steps (Qiu *et al*. 2012; Kefi *et al*. 2019), Fatty acyl-CoA reductases are likely involved in the core functions of fatty acid synthesis and in the synthesis of CHCs (Finet *et al.* 2019), and long-chain-fatty-acyl-CoA synthetases activate the oxidation of complex fatty acids (Alves-Bezerra *et al*. 2016). It is however difficult to

determine which genes are specific to CHC synthesis, as members of the same gene families are also involved in fatty acid biosynthesis (Dembeck *et al.* 2015).

1.4 STUDY SYSTEM

Crematogaster levior and *Camponotus femoratus* are one of the few examples where two ant species mutualistically share a common nest (Davidson 1988). This specific form of lifestyle, in which two species inhabit the same nest and use the same foraging trails but keep their brood separate is termed parabiosis (Forel 1898), which is derived from the Greek words *para* = next to each other, and *bios* = living. The term was coined to distinguish it from *symbiosis*, in which organisms of different species are living together. Such parabiotic associations are rare and mostly restricted to ants of the genera *Camponotus* and *Crematogaster* (Menzel & Blüthgen 2010).

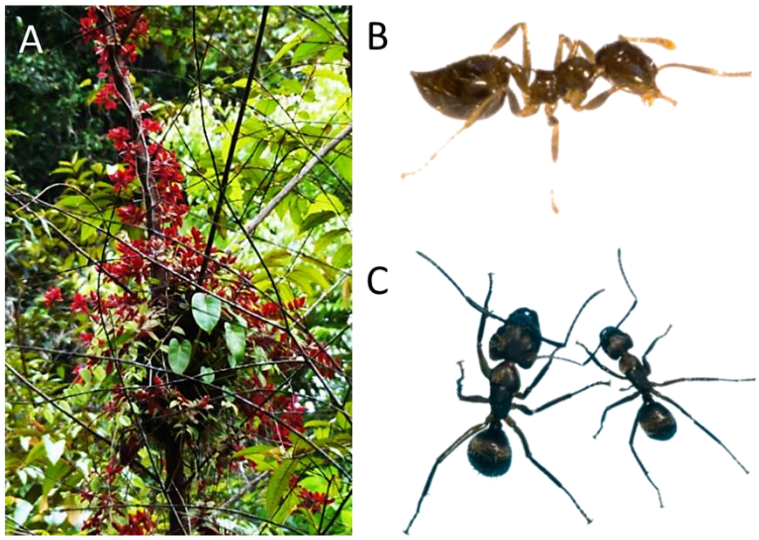


Figure Introduction 1: **(A)** Ant garden; **(B)** Worker of *Crematogaster levior*, **(C)** Workers of different sizes of *Camponotus femoratus*.

Cr. levior and *Ca. femoratus* live in so-called ant-gardens that can be found frequently in trees throughout the neotropics (Figure Introduction 1). So far, ant-gardens have been reported from across the Amazonian rainforest, but also in Asian rainforests mutualisms within ant-gardens exist between species of the same genera (Menzel & Blüthgen 2010). The carton part of the nest consists of soil, animal faeces and plant materials (Davidson 1988). The term ant-garden refers to the fact that *Ca. femoratus* continuously brings seeds of epiphytes (e.g. *Peperomia*, *Philodendron*, *Codonanthe*) into the nest that germinate and provide stability and humidity (Yu 1994). In turn, the ants provide the epiphytes with soil, fertilizer and protection against herbivores (Vantaux *et al*. 2007).

It is often speculated whether both ant species actually profit from this unusual form of symbiosis (Davidson 1988; Vantaux *et al.* 2007), but so far no obvious parasitic behaviour has been reported and the consensus within the literature seems to be that *Cr. levior* profits from the protective abilities and the provision and maintenance of the ant garden by the much larger *Ca. femoratus*, while *Ca. femoratus* follows the pheromone trails from *Cr. levior* to food sources (Orivel & Leroy 2010). In fact, a recent study on trophic niche and dietary preference on the species complex showed, that *Cr. levior* was often the first to discover resources but was displaced from baits by *Ca. femoratus* (Sprenger *et al*. in prep). The two species live polydomously, with multiple nests belonging to the same colony, as well as polygynously with multiple queens per nest (Davidson 1988; Céréghino *et al*. 2010).

Apparently parabioses provide the players with an advantage over other insect species in the same habitat, since garden ants are often the most abundant species in the canopy (Davidson 1988; Orivel & Leroy 2010). They commonly chose disturbed parts of the forest for their nests (Orivel & Leroy 2010). That is in tree fall gaps, along rivers, or near roads. One mode of colony founding seems to be by 'budding' from an already existing ant garden nearby (Davidson 1988). Hence, it is not unusual to find aggregates of ant-gardens of different sizes on the same tree or neighbouring trees that all belong to the same colony and that are connected by trails. However, it is not entirely clear how colonies are founded in previously unestablished habitats independently of already existing nests. The most likely scenario is that the nest is built by *Ca. femoratus*, and *Cr. levior* follows (Vantaux *et al*. 2007).

1.4.1 Chemical traits of *Cr. levior* and *Ca. femoratus*

For ants it is highly unusual to allow other species into the same nest and even members of the same species will be rejected if they belong to another colony. Such recognition of nest-mates in social insects is facilitated by CHCs (Buchwald & Breed 2005). It has therefore been studied whether *Cr. levior* and *Ca. femoratus* possess common chemical traits that would allow them to recognize their parabiotic partner (Orivel *et al.* 1997; Menzel *et al.* 2014). However, only very few chemicals are shared between the species (Orivel *et al.* 1997; Emery & Tsutsui 2013), which on the one hand points to a mechanism by which individuals learn to recognize the chemical characteristics of their nesting-partner, and which on the other hand shows how limited the actual physical contact between species is. If they would actively engage in grooming or trophallaxis, chemical compounds would be transferred from one species to the other (Orivel *et al*. 1997).

Later studies found unusual diversity in the CHC profiles in each of the two species (Emery & Tsutsui 2013). In fact, both species perfectly split into two distinct chemical types that have been postulated to represent cryptic species (Menzel *et al*. 2017).

1.5 AIMS OF THIS THESIS AND RESEARCH QUESTIONS

The aim of my dissertation was to analyse the genomic underpinnings of cuticular hydrocarbons and their role in recognition, ecological adaptation and suspected species divergence in mutualistic ants.

Fluctuating ambient temperature is one of the most important selection pressures for ectothermic organisms such as ants, to which individuals must often react within short time spans. Temperature affects many internal processes from metabolism to locomotion and poses particular adaptive pressure on organisms with a high surface-to-volume-ratio that are prone to desiccation stress. Acclimation, but also long-term adaptation, is facilitated by cuticular hydrocarbons, of which the composition can change plastically. Because CHCs have evolved secondary functions in communication and mate-choice, they are often proposed as so-called magic traits, that are involved in ecological adaptation but that also facilitate non-random mating and are thus believed to drive speciation (Smadja & Butlin 2009; Chung *et al.* 2014; Chung & Carroll 2015). *Crematogaster levior* and *Camponotus femoratus* are ideal model organisms to establish studies on the dual role of cuticular hydrocarbons, as they occur across a temperature-precipitation gradient in French Guiana and both possess highly variable CHC profiles. This thesis was furthermore aimed at analysing the population structure of both *Ca. femoratus* and *Cr. levior* in more detail and untangling the interplay between environmental factors, genomic differentiation and cuticular hydrocarbon diversity. My objectives thus were the following:

(**i**) Is the highly divergent CHC profile within both *Cr. levior* and *Ca. femoratus* reflective of cryptic species divergence or do CHC profiles reflect local adaptation to climate parameters?

We tackled this question in *chapter 1* by a large-scale sampling approach across French Guiana and subsequent barcoding of the mitochondrial COI region of both ant partners. This region is particularly suited to clarify species identity, because mitochondrial DNA normally does not undergo recombination, and therefore sequences are structurally rather conserved. Second, the COI region accumulates more mutation between species as other mitochondrial genes and because of that can more reliably detect speciation events. For *Ca. femoratus* also nuclear markers were sequenced to gain a better resolution of population structure. Primers were targeted at conserved exonic sequences close to variable introns that generally accumulate more mutations and should thus allow to resolve even recently diverged species. We furthermore analysed the morphology and the cuticular hydrocarbon profiles of the ants and in the end have three lines of evidence for cryptic speciation in both

Cr. levior and *Ca. femoratus*. It was of utmost importance to clarify this first, as the existence of cryptic species would influence all of the following analyses.

(**ii**) How are (cryptic) species distributed across French Guiana and is the observed distribution influenced by climate or other environmental parameters?

Cr. levior and *Ca. femoratus* are indeed both comprised of two previously unidentified cryptic species (question (**i**)). This of course poses new questions, such as whether the distribution across the sampling range is the same for both cryptic species pairs, or if they show patterns of distribution that reflect the adaptation to specific environmental or climatic parameters. I was furthermore interested in the population structure, i.e. the relationship between populations and species. A next step was therefore to correlate genetic divergence found in the COI locus to the distribution of the different species and to environmental parameters. During sampling, we noted several nest-characteristics, such as the genera of epiphytic plants, height of the nest, canopy coverage above the nest and elevation above sea level. Furthermore, for every sampling point, climate parameters were correlated with occurrence patterns of *Cr. levior* and *Ca. femoratus* to find potential niche specialization between cryptic species that might provide an explanation for species divergence (*chapter 1*).

(**iii**) What characterizes the CHC profiles of the different species and do they show fine-scale geographical differences that depend on climate?

As CHCs function both as social signals and as an ecological adaptation to desiccation stress, many factors in an organism's surrounding can pose selection pressures on the specific CHC profile. The CHC profiles of each of the four species were analysed in detail and the differences between them were characterized. Subsequently it was tested whether CHC profiles correlated with climate or mutualistic partner (*chapter 2*).

(**iv**) What are the potential drivers of species divergence within the mutualistic complex and do we find patterns of co-speciation?

I was curious about the mechanisms that could have driven the divergence of species within *Cr. levior* and *Ca. femoratus.* I use the results obtained for questions (**ii**) and (**iii**) as a basis for discussing potential scenarios that might have shaped the evolutionary history of the species. Since close interspecific relationships such as parasite-host associations or mutualisms are supposed to be drivers for evolution due to enhanced selection pressures,

also in this system, a speciation event in one species could have led to a speciation event in the mutualistic partner. To this end, I checked whether the different cryptic species showed any preference towards species identity of their mutualistic partner. If co-speciation occurred in this association, this should be visible in distinct partner preferences. If no such patterns would be found, this would rather point to another mechanism that drove speciation, such as divergent adaptation events in both species. I speculate on these questions on the basis of *chapters 1*, *2*, *3* and *4*.

(**v**) What is the genomic basis of cuticular hydrocarbon variation and are there genes that are implicated in both local climate adaptation and hydrocarbon synthesis or perception?

The main objective of this thesis was to analyse the dual role of cuticular hydrocarbons as recognition cues and in ecological adaptation and furthermore gain insight into their genomic basis and evolution. Since there was no reference genome available for this species, I sequenced and annotated the genome for one of the cryptic species, *Cr. levior* A, by combining different sequencing techniques and thus obtaining a high-quality reference genome (*chapter 3*). I then used this genome as a basis for a so-called Pool-Sequencing (Pool-Seq) approach on five populations of *Cr. levior* A and B respectively (*chapter* 4). Pool-Seq has the advantage of being a cost-effective method to sequence the variation within a population not just for a restricted number of sites, but genome-wide. Consequently, this approach allowed me to analyse the population structure in more detail and to compare the results to the previous analyses that were conducted with only one genetic marker (*chapter 1*). Furthermore, I correlated different population-specific parameters with genome-wide patterns of differentiation and thus identify candidate loci and genes that are implicated in the genomic underpinnings of certain traits, i.e. cuticular hydrocarbons and local climate. French Guiana is spanned by a temperature-precipitation gradient with a dry and hot west and a cooler and wetter east. This means that across our sampling range populations are subject to different climates and should, to enhance fitness, show differential adaptations to their local climatic selection pressures.

(**vi**) Are there patterns of parallel genomic evolution in both *Cr. levior* species due to parallel selection pressures?

Both cryptic *Cr. levior* species occur in the same regions across French Guiana and should be subject to exactly the same climatic and local environmental selection pressures. In general, this should lead to adaptations that improve fitness in these environments in

both species. The question thus is whether selection leaves the same traces across the genomes of the two sister species. I tested this by comparing the results from genome-wide association studies between *Cr. levior* A and *Cr. levior* B and checked whether the same loci, genes or pathways were selected as a response to the same environment (*chapter 4*). In general, if I would not find patterns of parallel genomic evolution, this points to either evolutionary constraints or genomic redundancy (see 1.1.4).

CHAPTER 1

Cuticular hydrocarbons as potential mediators of cryptic species divergence in a mutualistic ant association

Juliane Hartke, Philipp P. Sprenger, Jacqueline Sahm, Helena Winterberg, Jérôme Orivel, Hannes Baur, Till Beuerle, Thomas Schmitt, Barbara Feldmeyer, Florian Menzel

Published as:

Hartke J, Sprenger PP, Sahm J, Winterberg H, Orivel J, Baur H, Beuerle T, Schmitt T, Feldmeyer B, Menzel F (2019) Cuticular hydrocarbons as potential mediators of cryptic species divergence in a mutualistic ant association. Ecology and Evolution 9:9160–9176.

ABSTRACT

Upon advances in sequencing techniques, more and more morphologically identical organisms are identified as cryptic species. Often, mutualistic interactions are proposed as drivers of diversification. Species of the Neotropical parabiotic ant association between *Crematogaster levior* and *Camponotus femoratus* are known for highly diverse cuticular hydrocarbon (CHC) profiles, which in insects serve as desiccation barrier but also as communication cues. In the present study, we investigated the association of the ants' CHC profiles with genotypes and morphological traits and discovered cryptic species pairs in both genera. To assess putative niche differentiation between the cryptic species, we conducted an environmental association study that included various climate variables, canopy cover, and mutualistic plant species. Although mostly sympatric, the two *Camponotus* species seem to prefer different climate niches. However, in the two *Crematogaster* species, we could not detect any differences in niche preference. The strong differentiation in the CHC profiles may thus suggest either a possible role during speciation itself by inducing assortative mating, or by reinforcing sexual selection after the speciation event. We did not detect any further niche differences in the environmental parameters tested. Thus, it remains open how the cryptic species avoid competitive exclusion, with scope for further investigations.

Keywords: speciation, population structure, niche differentiation, environmental association, sexual selection, integrative taxonomy

INTRODUCTION

Diversity on earth is reflected in the ongoing discovery of a large number of species every year. Among animals, insects are especially species-rich and, out of an estimated 5 million species, only about 1 million have been described (Stork 2018). Finding new species can be challenging due to remote and undiscovered habitats or a high morphological similarity to closely related species. The latter, so-called cryptic species, are defined as distinct, but morphologically similar species (Bickford *et al*. 2007). They are often identified coincidentally based on genetic data, chemical profiles or behaviour. The lack of morphological differentiation between cryptic species can be due to recent divergence and too little time for distinct morphological features to evolve (Grundt *et al.* 2006; Gustafsson *et al.* 2014), or by selection on morphological stasis (Bickford *et al*. 2007; Struck *et al*. 2018). It has also been postulated that taxa, which communicate mating signals via non-visual cues (e.g. chemicals, vibrations, sounds) are more likely to harbour cryptic species, as morphological differentiation in these taxa is less important than e.g. in some birds, which use visual signals as mating displays (Andersson 1982; Hudson & Price 2014).

Given that cryptic species are morphologically alike and often closely related, one would expect them to be ecologically very similar and to exhibit only slight niche differentiation (Violle *et al.* 2011). Already very subtle ecological divergence in traits like thermal niche or food preferences, as well as spatio-temporal heterogeneity (e.g. different availability of resources) could allow such species to share the same habitat and avoid competitive exclusion (Gause 1932; Hardin 1960; Scriven *et al*. 2016). In ants for example, cryptic species can occur sympatrically, if they inhabit distinct niches, e.g. by specializing on different symbiotic fungi (Schultz *et al*. 2002). Next to the question how cryptic species coexist, it is also often unclear how species boundaries can be maintained between closely related species sharing the same habitat. One proposed mechanism is the expression of phenotypic traits that lead to assortative mating, and thus reduce gene flow (Dieckmann & Doebeli 1999). In this context, phenotypic traits might favour speciation even in sympatry if they are shaped by ecological selection pressures and at the same time induce assortative mating (so-called 'magic traits'), such as colour patterns or smell (Servedio *et al*. 2011; Nosil 2012; Thibert-Plante & Gavrilets 2013).

Species interactions can promote and speed up the emergence of novel phenotypic traits, lead to coevolution and diversification (Hoeksema & Bruna 2000; Guimarães *et al*. 2011; Thompson *et al*. 2013). For mutualisms, adaptive dynamics models predict that if in a population of a mutualistic species certain groups of one species become more attractive and are thus chosen as partners more often, evolutionary branching should occur (i.e. the split into two distinct phenotypic clusters; Doebeli & Dieckmann 2000). This dimorphism in one mutualistic partner can lead to disruptive selection in the other partner and therefore

to a co-speciation event (Doebeli & Dieckmann 2000). Although strict co-speciation seems rather rare (de Vienne *et al*. 2013), in mutualisms it was described repeatedly e.g. between arthropods and their endosymbionts (Degnan *et al.* 2004; Hosokawa *et al.* 2006; Bolaños *et al*. 2019), in specialized ant-plant mutualisms (Chomicki *et al.* 2015), and fig-pollinating wasps and figs (Jousselin *et al.* 2008; Cruaud *et al.* 2012). Alternatively, species diversification in mutualisms can also be facilitated by partner switches like in pollination mutualisms (Janz *et al*. 2001; Kawakita *et al.* 2004) or ant-plant associations (Quek *et al.* 2004).

A remarkable example of mutualism are parabioses, which are defined as interactions between two different ant species sharing a nest with separate brood chambers (Orivel *et al.* 1997; Menzel *et al*. 2008b). Here, we investigate the Neotropical ant species *Crematogaster levior* and *Camponotus femoratus* that live parabiotically in so-called ant gardens and both profit from abilities of their partners (Davidson 1988; Vantaux *et al.* 2007). Although the two species share a common nest and show interspecific tolerance, they keep their own species-specific cuticular hydrocarbon (CHC) profiles (Emery & Tsutsui 2013). Previous studies revealed two substantially different chemical phenotypes (or chemotypes) in both *Cr. levior* and *Ca. femoratus*, that otherwise were morphologically and ecologically indistinguishable (Emery & Tsutsui 2013; Menzel *et al*. 2014). CHCs cover the cuticle of basically all terrestrial arthropods. They are the main component of the waxy epicuticular layer, whose primary role is to prevent desiccation (Blomquist & Bagnères 2010). However, CHCs secondarily evolved several important roles in chemical communication like mediating recognition of mating partners (Thomas & Simmons 2008), and (in social insects) of nestmates and castes (van Zweden & d'Ettorre 2010). A CHC profile usually consists of structurally different groups of hydrocarbons, namely straight-chained n-alkanes, mono- or poly-methyl-branched alkanes and mono- or poly-unsaturated alkenes, in different combinations (Blomquist 2010). As CHC profiles are usually species-specific, but similar even between distant populations (Martin *et al.* 2008), high diversity is unusual within a single species.

In this study we elucidate the species status of the different chemotypes of both, *Cr. levior* and *Ca. femoratus*, by multiple lines of evidence within the framework of integrative taxonomy (Heethoff *et al*. 2011; Steiner *et al.* 2018). We compared cuticular hydrocarbons, secondary metabolites, morphological traits and genotypes between different colonies, and find clear evidence for two cryptic species in each of the two genera. Next, we asked whether these cryptic species differ ecologically, and conducted an environmental association study including local climate, mutualistic partners, ant garden plants and canopy cover. Finally, we tested for partner preferences among the mutualistic species.

MATERIALS AND METHODS

Sampling

We collected parabiotic ants of the species *Crematogaster levior* and *Camponotus femoratus* along an east-west gradient in French Guiana from August to October 2016. The east-west transect in French Guiana coincides with a climatic gradient (i.e. higher precipitation and lower temperatures in the east of the country and vice versa). We only collected ants foraging outside the nests, thereby leaving the colonies intact. To make sure we sampled different colonies of these polydomous species, we only collected ants from ant gardens which were at least 20 meters apart from each other. In total, we collected 333 colonies from 13 different locations (Table 1.1). If we could not reach the garden itself, we looked for shared trails or extrafloral nectaries attended by both species. In some of these cases (n = 20), we were not able to obtain individuals of *Ca. femoratus*. For each colony collected, we took a GPS point using a Garmin eTrex H personal navigator (Garmin Europe Ltd., Southampton, UK), noted plant genera present on the ant gardens (*Philodendron*, *Aechmea*, *Codonanthe*, *Peperomia and Anthurium*) and took a vertical photo of the canopy with a Nikon Coolpix W100 (Nikon GmbH, Düsseldorf, Germany). Samples for genetic and morphological analyses were stored in 99% ethanol.

Table 1.1: Sampling sites with details on sampled and analysed colonies. Numbers (#) of sampling sites refer to numbers on the map in Figure 1.1.

Site	*Code*	*#*	*Latitude*	*Longitude*	*Eleva-tion (m)*	*Number of colonies*	*Genetically analysed samples (Cr | Ca)*	*Chemi-cally analysed samples (Cr | Ca)*
Apatou	AP	1	5.200783	-54.312017	28	16	16 | 16	16 | 16
Saint-Laurent	SL	2	5.463902	-53.997322	63	36	33 | 29	36 | 32
Angouleme	AN	3	5.409200	-53.650933	64	1	01 | 01	01 | 01
Sinnamary	SI	4	5.352035	-53.077604	45	20	20 | 20	19 | 20
Petit Saut	PS	5	5.061213	-52.988772	93	21	19 | 17	18 | 18
Paracou	PAR	6	5.265905	-52.933605	41	53	50 | 47	50 | 49
Les Nouragues	LN	7	4.039650	-52.673933	63	74	72 | 60	72 | 61
Kourou	KO	8	5.083106	-52.643022	23	12	12 | 10	11 | 11
Montsinery	MT	9	4.866000	-52.538483	26	4	04 | 04	04 | 04
Cacao	CA	10	4.557416	-52.463067	71	22	21 | 19	21 | 20
Cayenne	CAY	11	4.793831	-52.317594	20	6	06 | 05	06 | 05
Regina	RE	12	4.181286	-52.131963	82	16	16 | 13	16 | 14
Patawa	PAT	13	4.546067	-52.130483	282	52	52 | 48	52 | 51

Chemical analyses

To analyse the CHC profiles, we immersed 10 freeze-killed *Cr. levior* or 5 *Ca. femoratus* workers per colony for 10 minutes in hexane. In *Cr. levior*, the cuticle contained polar secondary metabolites next to CHCs. These two substance groups were separated by fractionation using SiOH columns (Chromabond, 1mL/100mg, Macherey-Nagel, Düren, Germany). CHC fractions were eluted with hexane; the polar compounds with dichloromethane. The samples of polar compounds were dried under a gentle nitrogen stream and re-dissolved in approximately 50 µl hexane for analysis.

Cuticular hydrocarbons were analysed using gas chromatography-mass spectrometry (GC-MS). The gas chromatograph (7890A, Agilent Technologies, Santa Clara, CA, USA) was equipped with a Zebron Inferno ZB5-MS capillary column (length 30 m, Ø 0.25 mm, 0.25 µm coating, Phenomenex, Aschaffenburg, Germany) and Helium was used as carrier gas with a flow rate of 1.2 mL per minute. The mass spectrometer (5975C, Agilent Technologies) was used with electron ionization (EI) at 70 eV.

For the *Cr. levior* CHC extracts, 4 µl were injected into the GC at 40°C using a PTV (Programmed Temperature Vaporizing) method and this temperature was held constant for 2 minutes. Thereafter, the oven heated up with 60°C per minute to 200°C and above this temperature with 4°C per minute to 320°C which were kept for 10 minutes. The PTV method allows a higher injection volume, which was needed because of the presumably lower quantity of the much smaller *Crematogaster* ants. In *Ca. femoratus* 2 µl of extract was injected at 60°C using the splitless method. The oven heated up with 60°C per minute to 200°C and then with 4°C per minute to 320°C which again were kept constant for 10 minutes. The same temperature program as for *Camponotus* CHCs was used to analyse the polar compounds of *Cr. levior*. The resulting chromatograms were integrated manually using *MSD ChemStation* (E.02.02.1431, Agilent Technologies).

CHCs were identified using Kovats indices and diagnostic ions (Carlson *et al.* 1998). We excluded all substances which were not hydrocarbons as well as substances which had proportions less than 0.1% on average or were present in less than 20% of the samples (of the respective chemotype). Because the number of double bonds sometimes differed between colonies, we still included substances with multiple double bonds even if they occurred in less than 20% of the samples if other alkenes of the same chain length were present in other samples.

The polar substances produced by *Cr. levior* were likewise analysed via GC-MS as described above. They were aligned based on their mass spectra using a custom database. To investigate the molecular formula of the polar substances, highly concentrated samples of the *Cr. levior* A and B (100 individuals per sample) were analysed using GC-EI-HRMS (= gas

chromatography coupled with high resolution mass spectrometry). The setup we used allows the generation of accurate masses to establish molecular formulae of molecular and fragment ions at Δm <3.0 mmu. For GC-EI-HRMS we used an Agilent 6890 gas chromatograph equipped with an analytical column (30 m × 0.25 mm i.d., film thickness 0.25 µm; ZB-1MS, Phenomenex, Aschaffenburg, Germany), helium as carrier gas (1.0 mL/min; constant flow mode) and a temperature program of 100 °C (3 min)–10 °C/min–320 °C (10 min). Injection volume was 1 µL in splitless mode. The gas chromatograph (GC) was coupled directly to a JMS-T100GC time-of-flight (TOF) mass spectrometer (GCAccuTOF, JEOL, Tokyo, Japan) in electron ionization (EI) mode at 70 eV. The source and transfer line temperatures were set at 200 and 310 °C, respectively. The detector voltage was set at 2050 V. The acquisition mass range was set from *m/z* 41 to *m/z* 650 with a spectrum recording interval of 0.4 s. The system was tuned with perfluorokerosene to achieve a resolution of 6000 (full width at half maximum) at *m/z* 292.9824. JEOL MassCenterTM workstation software was used for data acquisition and data evaluation.

Statistical analyses – chemical data

In total we analysed 322 different *Cr. levior* and 302 *Ca. femoratus* colonies. The colonies were assigned to the CHC chemotypes described previously (Menzel *et al*. 2014) based on NMDS ordinations (Supplement Fig. S1.1).

To check for major differences in the CHC composition, we pooled substances according to their substance class (n-alkanes, mono-, di- and tri-methyl alkanes, mono-unsaturated alkenes, alkadienes, alkatrienes and methyl-branched alkenes). We tested whether their abundances (dependent variables) differed between the two chemotypes of either genus (fixed factor) using PERMANOVAs (command *adonis*, R package *vegan* (Oksanen *et al*. 2016)). If a certain substance class was absent from several samples, we added minute normally distributed random numbers (mean: 10-8 ± 10-8) to the respective class for all samples, as PERMANOVA cannot manage samples with zero distance. This was only the case for alkadienes and methyl-branched alkenes in *Crematogaster*.

To quantitate the separation of the chemotypes, we adapted the concept of haplotype networks to CHC profiles. As compositional data is continuous, we categorized the profiles based on a principal component analysis (PCA). This method has the advantage that one can quantitate the separation between CHC profiles and display information of multiple PC axes (i.e. more than two dimensions) at the same time and provide a clear visualization of the degree of variation between and within groups. To this end, we firstly performed a PCA based on our CHC data after centered log-ratio (clr) transformation (Aitchison 1982; Brückner & Heethoff 2017). Subsequently, we assigned a number of possible categories to each PC axis based on their eigenvalues (i.e. the number of categories per PC axis equalled

its eigenvalue divided by 5 to obtain a "handable" number of axes and distances between samples), and was rounded to two if the eigenvalue was between 10 and 5. PC axes with eigenvalues < 5 were not considered. In our case most of the CHC variation was explained by the first PCs, which is why we only used the first three PCs for the network of *Crematogaster* (explained variance: 58.75%) and the first two PCs for the network of *Camponotus* (explained variance: 73.75%; all other PCs having eigenvalues < 5).

Then, the PC loadings for each sample were transformed into distinct categories by dividing the distance of a certain PC loading to the minimum by the whole range of the PC loadings and rounding this value to integer numbers. As a result, we obtained a sequence of categories for each sample, with the length of the character sequence being the number of PC axes used. We used the R package *pegas* (Paradis 2010) with the haplotype command to calculate different clusters (chemical types) based on the character sequences. Subsequently, we calculated the (integer) Euclidean distances between samples for each PC axis and summed them up. Networks were then constructed using *haploNet* (package *pegas*).

To find out if *Cr. levior* populations can be differentiated by their polar metabolites, we visualized ordinations based on Bray-Curtis distance matrices. Additionally, we performed random Forest analyses using the *randomForest* package (Liaw & Wiener 2002) to check if we could assign the samples to the CHC chemotype based on their polar substances. All statistics were conducted using R version 3.5.0 (R Core Team 2018).

Morphological measurements

After classification based on the CHC profiles, we measured 30-40 individuals per cryptic species of both genera from independent colonies that were randomly distributed over the different sampling locations (total N = 160). As *Ca. femoratus* workers are dimorphic, we took only minors (the smaller caste) for our analyses. All measurements were taken blindly in a random order (per genus) using a Keyence VHX-2000 digital microscope (Keyence International (Belgium) NV/SA, Urdorf, Switzerland). Thirty specimens of *Cr*. and *Ca*. were photographed and measured twice to assess reliability (= 1 - measurement error, see Bartlett & Frost 2008). In the further analysis, we took the mean of both measurements for those specimens. Variables with reliability < 85% were omitted from the analyses (Supplement Table S1.1; Fig. S1.2). For calculating reliability, we used the Intraclass Correlation Coefficient with the function *ICCest* as provided by the R-package ICC (see also Wolak *et al*. 2012).

We measured 23 characters for *Crematogaster* and 20 characters for *Camponotus* (based on Seifert 2008; Csösz *et al*. 2014; and additional criteria). For *Crematogaster*, all measurements were taken under 200-fold magnification, while for *Camponotus* three different magnifications were used due to their larger body size. We used 100-fold to measure the

mesosoma, 150-fold for head, legs and antennae and 200-fold magnification for all other characters of *Camponotus*. Measurements were taken using ImageJ (version 1.50e, National Institutes of Health, USA) and the *straight* measure tool. We used an in-house ImageJ script to convert pixels into µm for each measurement.

We used multivariate ratio analysis (MRA) to analyse our body measurements. MRA comprises a set of tools for analysing size and shape separately in a multivariate framework (see e.g., Baur & Leuenberger 2011; Baur *et al*. 2014; Gebiola *et al.* 2017 for a detailed description of the application). One of these tools is the shape PCA, which in contrast to a conventional PCA, allows to compare body shape irrespective of *isometric* body size. The effect of allometric variation (e.g., allometric scaling, see Baur & Leuenberger 2011; Klingenberg 2016) may then be explored by plotting the first two shape PCs against isometric size. First, we ran a shape PCA for each genus separately. Next, the PCA ratio spectrum, another method of the MRA toolkit, allowed the interpretation of individual shape PCs in terms of ratios. Finally, isometric size was calculated as the geometric mean of all measurements per individual. For calculating the shape PCA, isometric size and the PCA ratio spectra we used a slightly modified version of the R-script published by Baur *et al.* (2014). Plots were generated using *ggplot2* (Wickham 2016).

To statistically test for morphological separation of the cryptic species, we calculated MANOVAs with the first two shape PCs as dependent variables and the species identity as well as sampling location as fixed factors. We used the first two PC axes since they explained 48% and 56.6% of the variance in *Crematogaster* and *Camponotus*, respectively (the cryptic species did not differ in PC3). To compare the isometric size between each species within a genus, we calculated Welch two sample t-tests. Calculation of these statistics was done with the basic functions *MANOVA* and *t.test* provided by R.

COI barcoding

To test for genetic separation, one individual of *Cr. levior* and *Ca. femoratus* of every sampled colony was barcoded at the mitochondrial COI locus. DNA was extracted following the HotSHOT protocol (see Montero-Pau *et al.* 2008). For DNA extraction, two legs of each individual of *Cr. levior* and one leg for *Ca. femoratus* respectively were used and DNA fragments of the COI locus (primers: LCO1490, HCO2198) were amplified using the following PCR cycling protocol: 5 minutes of denaturation at 95° C, followed by 35 cycles of 30 seconds of denaturation at 95° C, 60 seconds annealing at 48° C and 90 seconds extension at 72° C. This was followed by a final extension step at 72° C for 10 minutes. For detailed PCR and sequencing reaction mix see Supplement Table S1.2. Thermocycler conditions for the sequencing reaction were: 1 minute of denaturation at 95° C, followed by 30 cycles of 10 seconds denaturation at 96° C, 10 seconds of annealing at 50° C and 2

minutes extension at 60° C. This was followed by 10 minutes of final extension at 72° C. Resulting DNA fragments were sequenced on an ABI PRISM 3700 (Thermo Fisher Scientific, Waltham, MA, USA). Sequences were trimmed and aligned in GENEIOUS v. 10.1.3 using the ClustalW (Thompson *et al.* 1994) plugin. All sequences were manually checked and curated if necessary. The final alignment had a length of 449 bases.

COI - parsimony networks, phylogeny, and population genetic parameters

Haplotype networks were created for *Cr. levior* and *Ca. femoratus* using the TCS algorithm in PopART v. 1.7 (Leigh & Bryant 2015). In addition, Bayesian phylogenies were created using MrBayes v. 3.2 (Huelsenbeck & Ronquist 2001) upon identification of the best substitution model (HKY+G for *Crematogaster* and *Camponotus*) with MEGA7 (Kumar *et al*. 2018). Phylogenetic analyses for both species ran for 13,500,000 generations for *Cr. levior* and 9,020,500 for *Ca. femoratus* respectively with a burn-in of 25%; trees were sampled every 500 generations. Resulting trees were visualized in Archaeopteryx v. 0.992 beta (Han & Zmasek 2009). Based on networks and phylogenies, *Cr. levior* and *Ca. femoratus* were both separated into two distinct clusters each corresponding to the previously identified chemotypes. Thus, for the following analyses, we treated them as four separate cryptic species and call them *Cr. levior* A and B, as well as *Ca. femoratus* PAT and PS.

To investigate allele frequency differences between the different sampling sites, pairwise FST values were calculated between all population pairs separately for each of the two cryptic species pairs of *Cr. levior* and *Ca. femoratus*, using Arlequin v. 3.5. (Excoffier & Lischer 2010). In addition, Tajima's D (Tajima 1989) was calculated as a measure for potential selection.

Nuclear markers for *Camponotus*

Based on the small number of SNPs that separate the two cryptic species of *Ca. femoratus* at the COI locus, we sequenced four additional nuclear loci to obtain more details on the genetic population structure. For *Cr. levior* we plan to use a Pool-Seq approach in a future study to obtain this information on a genome wide basis. In the following, we sequenced one individual per colony from locations with at least three PAT and three PS colonies (max. 12 colonies). In total 14 unannotated Exon-primed intron-crossing (EPIC) primers (Supplement Table S1.3; Ströher *et al*. 2013) were tested. Four primer pairs (ant.1FR, ant.389FR, ant.1087FR, ant.1401FR) that amplified and showed variability were sequenced and further analysed. The PCR master mix was the same as for COI barcoding, except for 0.1µl of each primer instead of 0.2µl. Thermocycler conditions were: 5 minutes of denaturation at 95° C followed by 35 cycles of 1 minute of denaturation at 92° C for primer pair 1087 and 40 cycles for the remaining primer pairs respectively, 1 minute of annealing at

59° C and 2 minutes extension at 70° C. This was followed by 6 minutes of final extension at 72° C. For details on the sequencing reaction see above in the COI section. Forward and reverse sequences were assembled and manually curated. Alignment lengths differed between all loci (ant.1: 137 bp, ant.389: 239 bp, ant.1087: 379 bp, ant.1401 399 bp = 1154 bp in total), and so did the number of sequence polymorphisms (ant.1: 4 SNPs, ant.389: 4 SNPs, ant.1087: 5 SNPs, ant.1401: 5 SNPs = 18 SNPs in total).

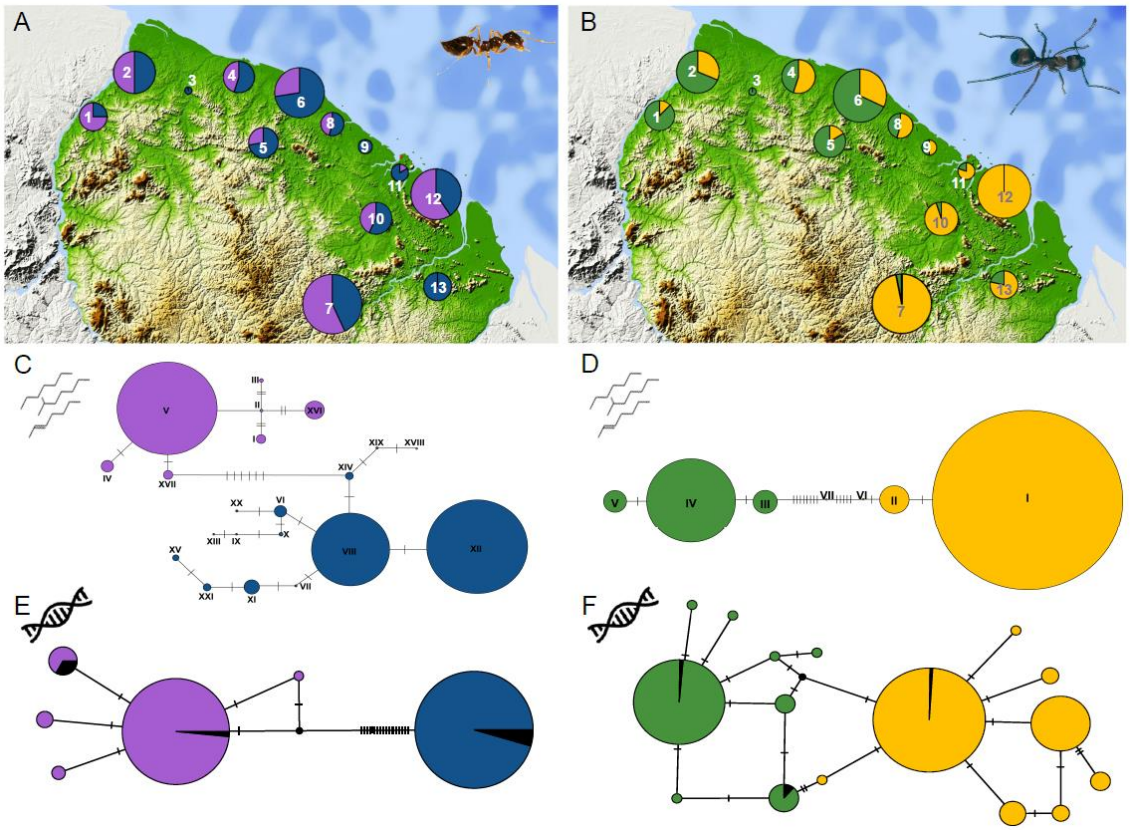


Figure 1.1: Chemotype and haplotype distribution across French Guiana and their differentiation. (**A**) Distribution of the cryptic *Cr. levior* species (*Cr levior* A: blue; *Cr. levior* B: purple). The size of the circles reflects the number of sampled colonies. (**B**) Distribution of cryptic *Ca. femoratus* species (*Ca femoratus* PAT: yellow; *Ca. femoratus* PS: green). Numbers in (**A**) and (**B**) refer to sampling locations in Table 1. (**C**) and (**D**) Chemical networks of *Cr. levior* and *Ca. femoratus*, using the same color code. (**E**) and (**F**) Haplotype networks (based on COI) using the same color code. Black coloration represents colonies without CHC information. Circles represent chemical types or haplotypes, respectively, and hatch marks indicate the number of character changes between them. Circle sizes reflect the number of colonies per chemical type or haplotype with singletons depicted slightly larger than according to their proportion. Pictures of *Cr. levior* (**A**) and *Ca. femoratus* (**B**) (© B. Feldmeyer).

Camponotus nuclear markers – parsimony networks and phylogeny

As for COI, we calculated the TCS networks with PopART (Leigh & Bryant 2015). We furthermore used BEAST v. 2.5 (Bouckaert *et al.* 2014) to calculate a phylogeny based on all four nuclear markers and the previously obtained COI sequences, comprising all individuals for which each locus was successfully sequenced (n = 93). Each locus was tested for the best

substitution model in MEGA7 (Kumar *et al.* 2018). BEAUTi, implemented within the BEAST package, was used to set up specifications for BEAST using StarBEAST2. Based on the Akaike's Information Criterion (AIC), we chose JC69 as best substitution model for nuclear marker ant.1FR and HKY for all others. For all markers a relaxed log normal clock model was used. Remaining parameters were set to default. BEAST was started with a chain length of 100,000,000, sampling trees every 1000 generations. The resulting trees were summarized in TreeAnnotator (included in BEAST) with a burn-in of 20% that was previously established in TRACER v. 1.6 (Rambaut *et al.* 2018). The resulting tree was visualized in Archeaopteryx v. 0.992beta (Han & Zmasek 2009). In addition, we used STRUCTURE 2.3.4 on the same dataset. The admixture model was used for calculations with a Burnin Period of 10,000 and a number of MCMC repetitions of 1,000,000 for a set number of two populations (k = 2).

Ecological and environmental association

Based on chemical and genetic information we could unambiguously assign each colony to *Cr. levior* A or B, or *Ca. femoratus* PAT or PS. First, we tested for non-random associations between the two cryptic *Crematogaster* and *Camponotus* species using a χ2-test. Second, we obtained climate data from CHELSA bioclim variables (Karger *et al*. 2017), consisting of composed climate data for the years 1979-2013 for the GPS location of every sampled colony. We performed a PCA with all 19 climate variables to reduce the number of variables. Most variance was explained by the first PC axis (76.47%) and was characterized by an inverse relationship of precipitation and temperature variables (i.e. higher precipitation correlates with colder temperatures). A high factor loading coincided with high annual precipitation (mean: 3137.08 mm; minimum: 1979 mm; maximum: 4873 mm) and a low annual mean temperature (mean: 25.6°C; minimum: 24.4°C; maximum: 26.3°C). Third, the presence/absence of plant genera on the ant nest was coded as a binomial variable (1 = present, 0 = absent). Canopy cover was estimated in ImageJ: All pictures taken from the canopy above each ant nest were converted to black-and-white using the *Make binary* command; covered areas were measured using the *Histogram* function. The obtained data was transformed to relative proportions.

For each colony, we created binomial variables of the species for *Crematogaster* (A vs. B) and *Camponotus* (PAT vs. PS). These were used as dependent variables in two binomial generalized linear mixed models with logit link function. As explanatory variables, we used the loading of PC1 from the climate PCA described above, the percentage of canopy covered, the identity of the parabiotic partner and a binomial variable for the presence of each plant genus on the ant gardens. We allowed interactions for each of these variables with the climate PC1, because canopy cover or species distributions might be influenced by the climate. Both models were reduced in a stepwise manner until the AIC was lowest.

Statistical analyses – comparing data sets

To analyse associations between chemical profiles, genetic distance and geographical distance, we performed Mantel tests based on Pearson correlation with 9999 permutations. As measure for chemical distance (CHCs and polar substances separately), we used Bray-Curtis dissimilarities (command *vegdist*, package *vegan*, Oksanen *et al.* 2016). For each of the two haplotype pairs, Tamura-Nei (Tamura & Nei 1993) pairwise genetic distances were calculated with MEGA7 based on the COI sequences. Geographical distances were measured as Euclidean distances between the GPS coordinates. All tests were done using R v. 3.5.0.

RESULTS

CHC differences between cryptic species

As described earlier (Emery & Tsutsui 2013; Menzel *et al*. 2014), we found two clearly distinct chemotypes in both *Crematogaster levior* and *Camponotus femoratus*.

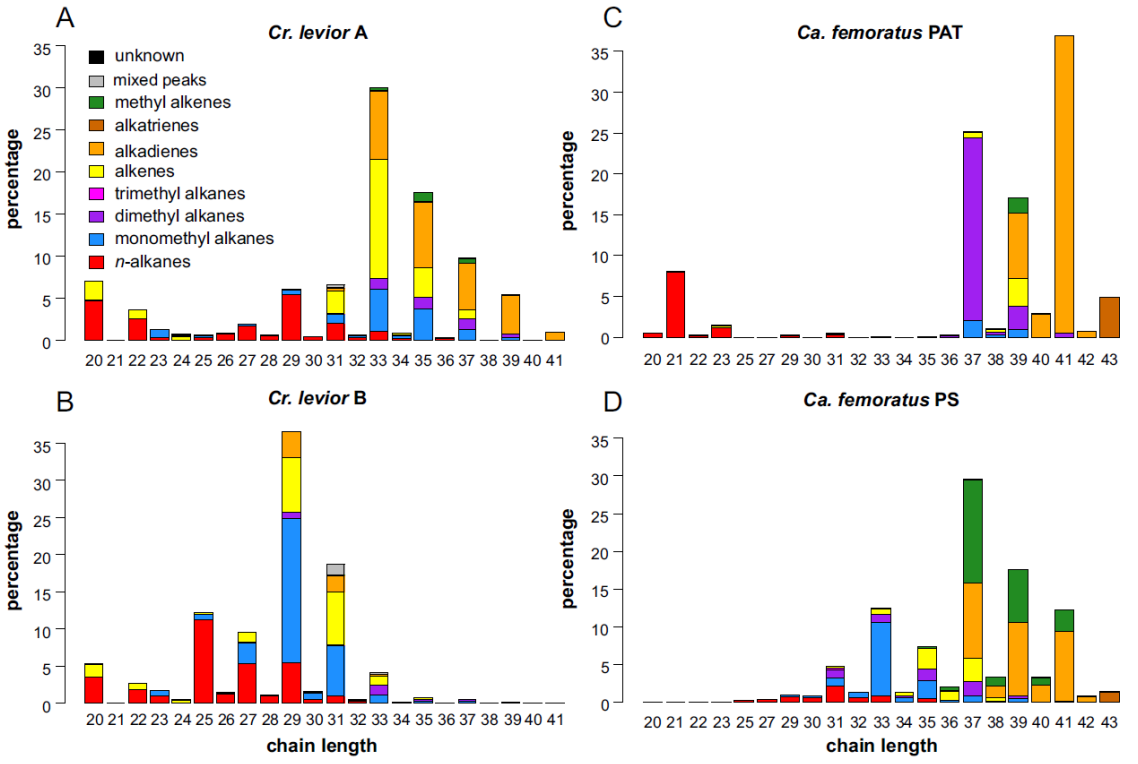


Figure 1.2: Differences between CHC profiles of the cryptic *Crematogaster levior* and Camponotus *femoratus* species. Plots show the mean distribution of different substance classes per chain length for all colonies of the respective species.

For *Crematogaster* (Fig. 1.1C), the chemical networks yielded two large clusters in *Cr. levior* A (cluster XII and VII) and one large cluster in *Cr. levior* B (cluster V). The profiles of *Cr. levior* A seemed more variable as we found 13 different chemical types (with two

singletons), compared to only 8 in *Cr. levior* B (with one singleton). *Crematogaster* A and B were clearly separated in the network. However, there was one exception, with the colony forming the singleton type XVIII showing characteristics of both chemotypes. In the network it was closer connected to *Cr. levior* A, but clearly clustered with chemotype B in an NMDS ordination (Supplement Fig. S1.1). This colony had the same COI-haplotype as other B colonies.

The profile of *Cr. levior* A (n = 174) was dominated by several alkadienes of odd chain length ranging from C29 to C41 (total abundance: 27.44 ± 6.24%; Supplement Fig. S1.3A). In contrast, the main peak in *Cr. levior* B (n = 148) was a mixture of 13- and 15-methyl nonacosane (17.91 ± 7.73 %; Supplement Fig. S1.3B). The CHCs of both cryptic *Crematogaster* species were vastly different with substances most common in A (substances >5% abundance: 30.24 ± 10.50%) being rare in B (6.36 ± 1.99%) and vice versa for substances most common in B (in B: 40.23 ± 9.51%; in A: 8.62 ± 3.37%). In comparison, the profile of *Cr. levior* A had more alkadienes (PERMANOVA: pseudo-F_1 = 137.98, p = 0.001), alkenes (pseudo-F_1 = 73.09, p = 0.001), di-methyl alkanes (pseudo-F_1 = 57.33, p = 0.001) and methyl-branched alkenes (pseudo-F_1 = 155.24, p = 0.001; Fig. 1.2A), while *Cr. levior* B had much higher proportions of mono-methyl alkanes (pseudo-F_1 = 637.39, p = 0.001) and *n*-alkanes (pseudo-F_1 = 191.56, p = 0.001; Fig. 1.2B).

The two cryptic *Ca. femoratus* species were obviously distinct without any exceptions. *Ca. femoratus* PAT colonies were mostly assigned to a single cluster (cluster I) and few colonies to a second one (cluster II). In comparison, PS colonies were distributed among three chemical types (cluster III, IV and V; Fig. 1.1D).

In *Ca. femoratus* PAT (n = 195) the CHC profile was dominated by 13,23-dimethyl heptatriacontane (22.47 ± 10.21%), as well as several different C41 alkadienes (13.24 ± 4.42% and 11.61 ± 3.42% for the two most abundant ones; Supplement Fig. S1.3C). In *Ca. femoratus* PS (n = 107), the most abundant substance was a 13-methyl heptatriacontene (13.49 ± 4.01%) followed by 13- and 15-methyl tritriacontane (9.60 ± 2.75%; Supplement

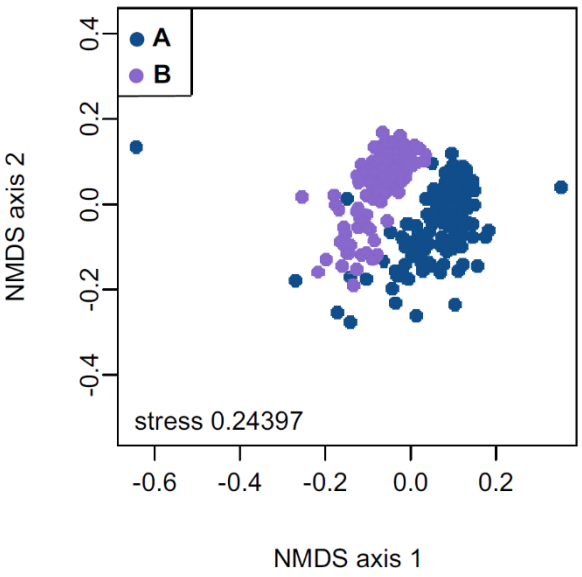


Figure 1.3: Differences in polar secondary metabolites of *Cr. levior*. NMDS ordination of the polar secondary metabolites produced by *Cr. levior*. Each dot represents the polar compound profile of one colony of *Cr. levior*.

Fig. S1.3D). The profiles of the cryptic *Camponotus* species differed strongly with the most common CHCs of *Ca. femoratus* PAT (substances >5% abundance: 62.75 ± 7.24%) being less common in PS (9.67 ± 2.60%) and the other way around although less pronounced (in PS: 50.55 ± 8.23%; in PAT: 19.19 ± 5.19%). The PAT colonies had higher proportions of dimethyl alkanes (PERMANOVA: pseudo-F_1 = 629.70, p = 0.001), alkadienes (pseudo-F_1 = 202.82, p = 0.001) and *n*-alkanes (pseudo-F_1 = 16.87, p = 0.001, Fig. 1.2C), while the PS ones had more mono-methyl alkanes (pseudo-F_1 = 1205.50, p = 0.001), methyl-branched alkenes (pseudo-F_1 = 1013.00, p = 0.001) and alkenes (pseudo-F_1 = 105.53, p = 0.001; Fig 1.2D).

Differentiation by polar metabolites

In 254 out of 322 *Cr. levior* colonies, we found a total of 60 different polar compounds on the cuticle. In the remaining extracts, polar substances were either not detected or had too low concentrations for reliable quantification. Similar to the CHCs, the colonies could be differentiated into two different clusters (Fig. 1.3; Supplement Fig. S1.4A, D). CHC chemotypes could be correctly identified based on polar chemistry using a random Forest algorithm which had a 1.18% OOB estimate of error rate. All 138 samples from A and 113 of 116 samples of B (error rate of 0.026%) were classified correctly.

The most common substances in *Cr. levior* A had abundances of 8.55 ± 8.12% (Retention time 24.10, Supplement Fig. S1.4B), 8.74 ± 6.70% (RT 24.62) and 19.19 ± 10.32% (RT 26.24; Supplement Fig. S1.4C), respectively, but lower abundances in B (3.45 ± 3.11%; 0.86 ± 1.07%; 3.82 ± 2.58%). In *Cr. levior* B, most abundant substances had proportions of 17.26 ± 10.81% (RT 20.20, Supplement Fig. S1.4E), 13.55 ± 4.40% (RT 20.30, Supplement Fig. S1.4F) and 5.45 ± 6.35% (RT 20.90), which were only 1.14 ± 1.13%, 3.35 ± 1.58% and 0.44 ± 0.67% (respectively) in A (all retention times given refer to the Zebron Inferno ZB5-MS capillary column). Using HR-MS the sum formulae of the major polar substances were derived as $C_{24}H_{36}O_4$ (polar substance at retention time 20.20, Supplement Fig. S1.4E), $C_{24}H_{38}O_4$ (RT 20.30, Supplement Fig. S1.4F), $C_{24}H_{36}O_4$ (RT 20.90), $C_{26}H_{38}O_4$ (RT 24.10, Supplement Fig. S1.4B), $C_{26}H_{40}O_4$ (RT 24.62) and $C_{28}H_{44}O_4$ (RT 26.24, Supplement Fig. S1.4C). The results showed a series of closely related compounds characterized by C24 to C28 carbon atoms containing four oxygen atoms, differing in the number of double bonds or rings from 6 to 8. In most cases there was a pair of compounds showing the same number of carbons only differing in the number of double bonds/rings. This pair-wise difference is also reflected in two series of fragment ions of m/z 237, 224, 209 and m/z 235, 222, 207 respectively, indicating an additional double bond isomer. However, to gain more insight into the underlying structures, higher quantities at higher purities are needed for NMR analysis.

Morphology

In shape, the two cryptic species of *Cr. levior* were largely overlapping. Nevertheless, the shape significantly differed between them (MANOVA based on shape-PCA: $F_1 = 18.07$, $p < 0.001$) but not between sampling locations ($F_{11} = 0.79$, $p = 0.73$). *Cr. levior* A and B differed in shape PC1 ($F_1 = 30.37$, $p < 0.001$; Fig. 1.4A) but only insignificantly in shape PC2 ($F_1 = 3.18$, $p = 0.079$; Fig. 1.4B). Shape PC1 was best described by the ratio between spine length and eye width (Fig. 1.4A), while shape PC2 was largely explained by the maximal distance between the spines (Fig. 1.4B). Moreover, *Cr. levior* B was larger than A (Welch t-test: $t_{74.94} = -3.61$, $p < 0.001$; Fig. 1.4A, B).

The morphological traits of *Ca. femoratus* largely overlapped between cryptic species as well, despite significant differences (MANOVA: $F_1 = 16.67$, $p < 0.001$). Again, we found no effect of sampling location ($F_{11} = 0.43$, $p = 0.16$). While we detected differences in body shape (shape PC1: $F_1 = 17.08$, $p = 0.001$, Fig. 1.3C; shape PC2: $F_1 = 10.04$, $p = 0.003$, Fig. 1.4D), the

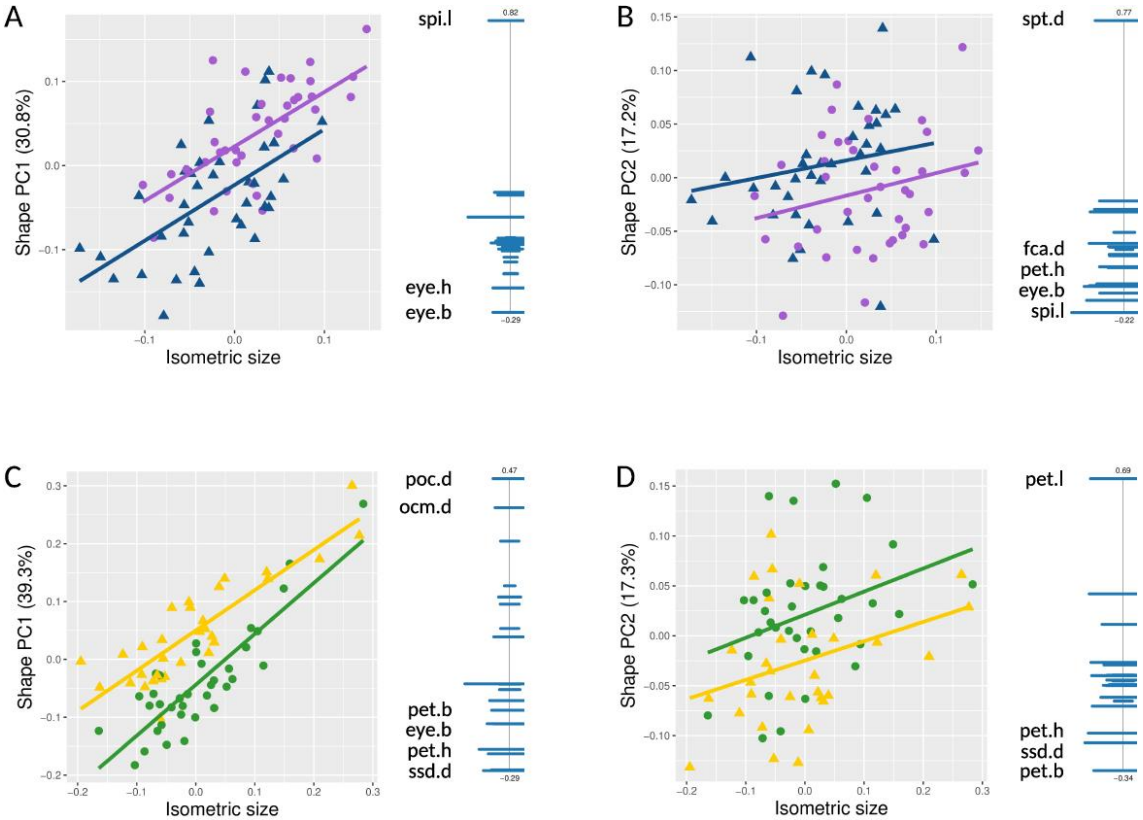


Figure 1.4: Morphological differentiation of the cryptic species of both ant genera. (**A-D**) Scatter plots depicting morphological differences of *Cr. levior* (**A,B**) and *Ca. femoratus* (**C,D**) and PCA ratio spectra. We plotted the first and second axis of a shape PCA (**A,C** and **B,D**, respectively) against isometric size. Each dot represents one individual of independent colonies. Symbols and colours correspond to cryptic species as follows: *Cr. levior*: blue triangle = A, purple dot = B; *Ca. femoratus*: yellow triangle = PAT, green dot = PS. To the right of the scatterplots, the ratio spectrum of the shape PC is shown. Up to four of the most relevant variables for calculating body ratios are indicated the ends of the spectra using the variable codes (see Supplement Table S1.1). Bars indicate the 68% confidence intervals based on 1000 bootstrap replicates (bars trimmed on right hand side due to the arrangement of figures).

cryptic species did not differ in isometric size (t_{57} = -0.41, p = 0.68; Fig. 1.4C, D). While the first shape PC was characterized by multiple traits on different body parts (Fig. 1.4C), shape PC2 was mainly explained by the ratio between petiole length to petiole width (Fig. 1.4D).

Genotyping results and population structure

COI - parsimony networks and phylogeny

The TCS networks of the COI sequences show two distinct genotype clusters for both *Cr. levior* (Fig. 1.1E) and *Ca. femoratus* (Fig. 1.1F) with a 1:1 association of genotype to chemotype. The *Cr. levior* group that corresponds to A consisted of a single haplotype only. *Cr. levior* B showed more genetic variation with five haplotypes. The separation between both species was based on 16 SNPs (single nucleotide polymorphisms), indicating divergent clades. In *Ca. femoratus*, the resulting haplotype networks were more diverse. Both *Ca. femoratus* PS and PAT consisted of eight distinct groups. Here, the cryptic species were separated by two SNPs.

The phylogenies showed a similar pattern. In *Cr. levior* the separation between cryptic species was strongly supported with a posterior probability of 1 (Supplement Fig. S1.6). In *Ca. femoratus* the separation was not as clear, based solely on COI with a posterior probability of 0.61 and two subgroups per cryptic species (Supplement Fig. S1.5).

COI – Population genetic structure

As measure for population differentiation, we calculated pairwise F_{ST} values separately for all four cryptic species, between all sampled sites. In *Cr. levior* A, results are not shown due to a lack of population differentiation (F_{ST} = 0 in all population comparisons). For *Cr. levior* B (Table 1.2), only few populations were genetically different with significant differentiation found between Kourou & Les Nouragues (F_{ST} = 0.308, p = 0.036), Kourou & Saint-Laurent (F_{ST} = 0.531, p = 0.045) and Saint-Laurent & Les Nouragues (F_{ST} = 0.127, p = 0.045). In *Ca. femoratus* we found greater differentiation between populations compared to *Cr. levior*, with six occurrences of fixed differences (F_{ST} = 1). In 42% of all pairwise comparisons, populations were significantly different in PS (Table 1.3), and 29% of all comparisons in PAT yielded significant differences (Table 1.4). We furthermore tested for potential selection using Tajima's D statistic (Supplement Table S1.4). Results for *Cr. levior* A are again not shown due to a lack of genetic differences. In *Cr. levior* B, Tajima's D was not significant in any population. In *Ca. femoratus* PS, Tajima's D was significantly smaller than zero in the Saint-Laurent population (TD = -1.513, p = 0.033) only. In *Ca. femoratus* PAT, Tajima's D was significantly smaller than zero in the populations of Paracou (TD = -2.072, p = 0.003), Les Nouragues (TD = -2.107, p = 0.002) and Saint-Laurent (TD = -1.486, p = 0.04).

Table 1.2: Population pairwise F_{ST} between 10 populations of *Cr. levior* B, based on the COI locus. Bold characters indicate statistical significance ($p < 0.05$) based on a permutation test.

	AP	*PAR*	*PS*	*LN*	*PAT*	*CAY*	*CA*	*KO*	*SI*	*SL*
AP	-									
PAR	-0.006	-								
PS	0.000	-0.130	-							
LN	0.108	0.112	0.010	-						
PAT	-0.012	-0.006	-0.117	0.070	-					
CAY	0.000	-0.096	0.000	0.037	-0.085	-				
CA	0.000	-0.031	0.000	0.088	-0.030	0.000	-			
KO	0.462	0.203	0.195	**0.308**	-0.009	0.250	0.392	-		
SI	0.034	0.007	-0.116	0.104	-0.025	-0.078	0.000	0.253	-	
SL	0.000	0.017	0.000	**0.127**	0.003	0.000	0.000	**0.532**	0.068	-

Camponotus nuclear markers – parsimony networks and phylogeny

As for COI sequences, we constructed TCS parsimony networks based on four additional nuclear markers (Fig. 1.6A-D) (we sequenced additional nuclear loci for *Camponotus* only, since a population genomic study is on the way for *Crematogaster*). In contrast to the network based on COI mitochondrial sequences, the networks of nuclear markers showed less clear separation of cryptic species (Fig. 1.6A-D). In contrast, a phylogenetic tree based on all five sequenced markers (Fig. 1.6E) clearly separated *Ca. femoratus* PAT and PS into two clades. Also the STRUCTURE analysis showed that all individuals could be assigned to one of the two chemotypes (Supplement Fig. S1.7).

Table 1.3: Population pairwise F_{ST} between 9 populations of *Ca. femoratus* PS, based on the COI locus. Bold characters indicate statistical significance ($p < 0.05$) based on a permutation test.

	AP	*PAR*	*PS*	*LN*	*RE*	*MT*	*KO*	*SI*	*SL*
AP	-								
PAR	-0.037	-							
PS	0.156	**0.092**	-						
LN	**1.000**	0.778	**0.796**	-					
RE	**1.000**	**0.787**	**0.811**	0.000	-				
MT	1.000	**0.778**	**0.796**	0.000	0.000	-			
KO	0.189	0.001	-0.135	0.817	**0.847**	0.817	-		
SI	0.000	-0.054	0.120	**1.000**	**1.000**	**1.000**	0.126	-	
SL	0.014	0.009	0.016	**0.892**	**0.897**	**0.892**	-0.079	-0.008	-

Partner preference and environmental association of cryptic species

There was no indication for a preferred association between either cryptic *Cr. levior* and *Ca. femoratus* species (Pearson's χ^2-test: $\chi^2_1 = 1.76$, $p = 0.18$). *Cr. levior* A nested with *Ca. femoratus* PAT in 100 and with PS in 65 cases, while *Cr. levior* B cohabited 96 times with *Camponotus* PAT and 44 times with PS.

Table 1.4: Population pairwise F_{ST} between 12 populations of *Ca. femoratus* PAT, based on the COI locus. Bold characters indicate statistical significance ($p < 0.05$) based on a permutation test.

	AP	*PAR*	*PS*	*LN*	*RE*	*PAT*	*CAY*	*MT*	*CA*	*KO*	*SI*	*SL*
AP	-											
PAR	**0.716**	-										
PS	0.500	-0.084	-									
LN	-0.153	**0.763**	**0.637**	-								
RE	0.248	**0.205**	-0.200	**0.437**	-							
PAT	-0.032	**0.545**	**0.273**	**0.065**	**0.130**	-						
CAY	0.250	0.173	-0.333	**0.451**	-0.209	**0.076**	-					
MT	0.000	0.694	0.368	-0.277	0.164	-0.133	0.111	-				
CA	-0.034	**0.521**	0.202	**0.080**	0.074	-0.034	0.010	-0.144	-			
KO	0.000	**0.732**	0.579	-0.099	0.296	0.013	0.333	0.000	0.017	-		
SI	0.516	0.001	-0.273	**0.648**	-0.020	**0.362**	-0.108	0.464	0.307	**0.551**	-	
SL	-0.167	**0.618**	0.325	-0.051	0.141	-0.063	0.101	-0.313	-0.083	-0.098	**0.400**	-

The distribution of cryptic *Crematogaster* species was independent of PC1, i.e. precipitation and temperature (binomial GLM: $N = 292$, $\chi^2_1 = 1.12$, $p = 0.29$), indicating sympatric occurrence of the cryptic species which is also visible when looking at their distribution across the complete sampling range (Fig. 1.1A). Neither canopy cover, nor the presence of any plant influenced the probability of species membership (A vs. B) in *Crematogaster* (all $p > 0.2$). However, species identity was influenced by an interaction of climate and *Camponotus* partner ($\chi^2_1 = 5.97$, $p = 0.015$). *Ca. femoratus* PS was less common in areas with high annual precipitation and lower annual mean temperature (i.e. the eastern part of French Guiana), while *Ca. femoratus* PAT was present across the whole sampling area (binomial GLM: $N = 279$, climate PC1: $\chi^2_1 = 111.91$, $p < 0.001$; Fig. 1.1B). None of the other factors tested influenced the probability of the species' presence (all $p > 0.15$). However, there was a weak interaction between climate PC1 and *Crematogaster* partner ($\chi^2_1 = 5.06$, $p = 0.025$) indicating slightly differing partner availability depending on climate.

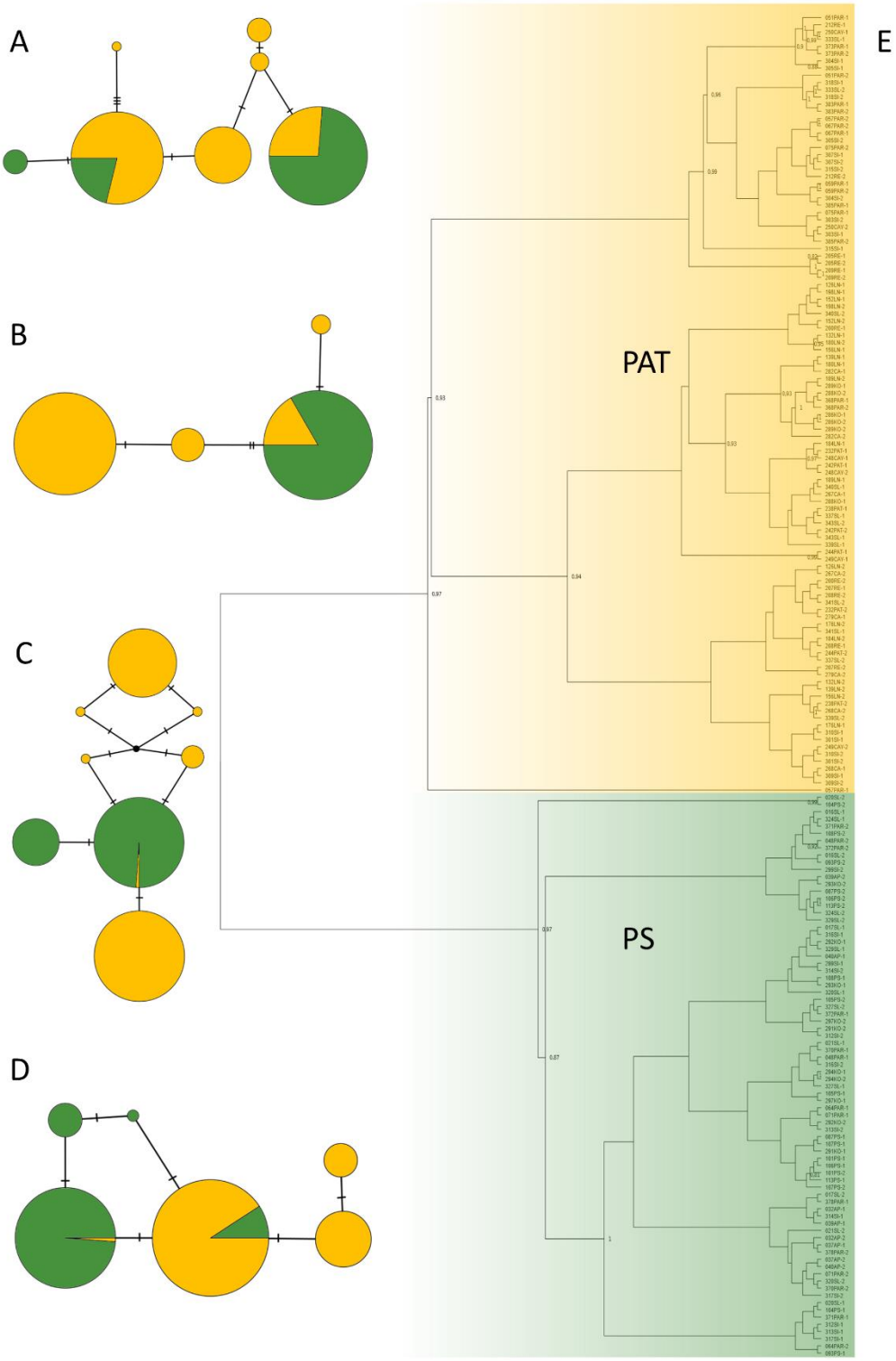


Figure 1.5: Genetic differentiation of cryptic *Ca. femoratus* species. (**A-D**) TCS Haplotype networks of four nuclear markers of *Ca. femoratus*. (**A**) ant.1401FR, (**B**) ant.1FR, (**C**) ant.1087FR, (**D**) ant.389FR. Green color indicates *Ca. femoratus* PS and yellow colour indicates *Ca. femoratus* PAT respectively. Haplotypes are shown as circles, with size depending on the number of included colonies. Number of SNPs (single nucleotide polymorphisms) between the haplotypes are shown as hatch marks. (**E**) Phylogenetic tree based on all four nuclear markers and mitochondrial COI for *Ca. femoratus*. Posterior probabilities >0.8 are displayed. Yellow colour corresponds to chemotype *Ca. femoratus* PAT, green indicates *Ca. femoratus* PS. Only individuals with all five loci sequenced were included (N = 93).

Connecting chemical profiles, genetic background and geographic distance

The CHC distances in *Cr. levior* A slightly increased with geographic distance (Mantel test: r = 0.066, p = 0.011). However, this was not true for *Cr. levior* B (r = 0.044, p = 0.084). In *Camponotus*, CHC distances increased with geographical distances for PS (r = 0.182, p < 0.001), but not for PAT (r = 0.040, p = 0.15). The Bray-Curtis dissimilarities of CHCs and polar compounds of *Crematogaster* were highly correlated (N = 253, r = 0.42, p < 0.001), further indicating that the polar differentiation exactly matches the CHC differentiation. However, within each cryptic species, CHC distance and distance in polar compounds were not correlated (*Cr. levior* A: r = 0.04, p = 0.15; *Cr. levior* B: r = 0.02, p = 0.34).

Mantel tests between pairwise Tamura-Nei distances and geographic distances revealed no isolation-by-distance pattern for *C. levior* B (r = -0.065, p = 0.968), but for *Ca. femoratus* PS (r = 0.38, p < 0.001) and - albeit only weakly - *Ca. femoratus* PAT (r = 0.09, p = 0.038). *Cr. levior* A consisted of only one haplotype without any variation at the COI locus, which is why this analysis was not possible here.

In *Cr. levior* B, colonies that were genetically more distant also had more dissimilar CHC profiles (r = 0.15, p = 0.021). However, such an association was neither detectable within *Ca. femoratus* PAT (r = 0.05, p = 0.15) nor within *Ca. femoratus* PS (r = 0.03, p = 0.29).

DISCUSSION

This study investigated the parabiotic ant species *Cr. levior* and *Ca. femoratus* whose shared nests (so called ant-gardens) are abundant in the Neotropics (Davidson 1988). Both previously identified species occur in two distinct CHC chemotypes, which are morphologically highly similar. We show that within *Cr. levior* and within *Ca. femoratus*, these chemotypes form two distinct units that can be classified as cryptic species. This is supported by multiple lines of evidence, all of which show conclusive results. First, the cuticular hydrocarbon analysis shows that both formerly classified species split into two clearly distinguishable chemotypes across our sampling range without intermediate profiles. For *Cr. levior*, we additionally show a clear separation in polar metabolites. Secondly, we morphometrically analysed the different species. Although there is a large overlap in traits between groups, we found slight but significant differences in body shape between the two cryptic *Camponotus* and between the two cryptic *Crematogaster* species. Moreover, *Cr. levior* B is slightly larger than *Cr. levior* A. Lastly, we barcoded all sampled colonies and found a 1:1 association between the previously assigned CHC chemotypes and newly assigned genotypes. Phylogenies based on COI perfectly split *Cr. levior* into two clusters. The same holds true for *Ca. femoratus* based on COI and four additional nuclear markers, where again

two distinct clusters are found. These results support our initial hypothesis that apparent CHC diversity is in fact a sign of distinct genetic lineages, i.e. cryptic species (in the sense of De Queiroz 2007). In the following sections, we first discuss the distribution and ecological niches of the cryptic species, then their population structures and possible scenarios explaining those, and lastly, the putative role of the vastly different cuticular hydrocarbon profiles during or after the speciation process.

Previous studies that looked at the distribution of cryptic species mostly found evidence for the competitive exclusion principle (García-Robledo *et al*. 2015; Leavitt *et al*. 2015; Vodă *et al*. 2015b). In fig wasps for example, morphologically similar species are less likely to occur in sympatry than morphologically dissimilar sister species (Darwell & Cook 2017). Interestingly, in our case, the two *Crematogaster* and *Camponotus* sister species co-occur across the whole sampling range with only one case of niche differentiation within the factors tested here. *Camponotus femoratus* PS is more common in the drier, western half of the country, while PAT was more frequently found in the wetter and slightly cooler east of the country. The high proportions of alkadienes in the CHC profile of *Ca. femoratus* PAT are in line with this climatic difference. This corroborates other studies in which alkadienes were found to be present more frequently and in higher percentages (only in interaction with cooler temperature) in multiple different species from high precipitation areas (van Wilgenburg *et al*. 2011; Menzel *et al*. 2017a). In contrast, the two *Crematogaster* species occur in similar frequencies across the whole sampling range with no obvious signs for niche differentiation in the parameters we tested. However, other ecological parameters, such as dietary differences or niche partitioning concerning the time of foraging activity or mating flights may still be of importance. Alternatively, *Cr. levior* A and B may represent ecologically neutral species (Hubbell 2001; Adler *et al*. 2007; Bell 2017). In this scenario diverse communities of functionally equivalent species coexist due to neutral dynamics (Hubbell 2005). We furthermore found no preferential association of either *Crematogaster* species for any of the two *Camponotus* species or vice versa, rendering co-speciation a more unlikely scenario. The lack in preference may not be too surprising given the distribution of the species. While the two *Crematogaster* species occur in similar frequencies throughout the sampling range, the two *Camponotus* species show the above mentioned east-west gradient. The choice of the mutualistic partner might therefore be a question of availability rather than preference.

Population structure, as well as haplotype diversity differed strongly between species. It was most extreme, with only a single haplotype and no population differentiation in *Cr. levior* A between all 12 sampled populations. We found five different haplotypes in *Cr. levior* B and eight in both *Ca. femoratus* species. In *Cr. levior* B population structure was very weak and there was no sign for isolation by distance. This result is surprising insofar, as other studies

on the genus *Crematogaster* usually show strong geographical or ecological structure (Türke *et al*. 2010; Boyle *et al*. 2018). In *Ca. femoratus* PS and PAT respectively, the COI locus, as well as two nuclear markers showed clear signs for isolation by distance. Tajima's D analysis furthermore showed signs for sudden population expansions in several of the observed populations of *Ca. femoratus* PS and PAT. Genetic differences between the two *Camponotus* species were generally low and only a small part of the nuclear markers we tested were variable between species. Furthermore, the previously assigned CHC chemotypes did not perfectly match the haplotypes of any of the nuclear loci, which may be due to incomplete lineage sorting, a possible sign of recent speciation between *Ca. femoratus* PS and *Ca. femoratus* PAT.

The lack of any population differentiation in *Cr. levior* A, with only a single COI haplotype in all sampled populations, could be explained by two different scenarios. The first is a strong bottleneck event coupled with a recent population expansion. A second explanation could be a selective sweep in haplotype A together with a population expansion. In insects, this is often found in the context of an infection with the endosymbiont *Wolbachia* that can manipulate its hosts reproduction (through e.g. mate-discrimination, cytoplasmic incompatibilities; Hoffmann *et al*. 1986; Schuler *et al*. 2016). However, the same signatures can be found after the spread of a beneficial mutation within a population, that will lead to reduced heterozygosity around the selected locus (Schlenke & Begun 2004). While we found only weak genetic differences between the cryptic *Camponotus* species, chemical differences were pronounced. Also *Crematogaster* showed unusually high interspecific differences in their chemical profile, which has previously been discussed as a mechanism to reinforce species divergence (Menzel *et al*. 2017b). The overlap in CHC composition between the two species of each genus was low, with peaks that were abundant in one species being low or absent in the other (see Results). This means that the CHC profiles differ much more than one would expect between sister species sharing similar abiotic and biotic niches (Menzel *et al.* 2017b). Especially compared to other traits, e.g. morphology or behaviour, chemical trait differences seem to be higher and less phylogenetically conserved (Blomberg *et al.* 2003; Kamilar & Cooper 2013). Chemical distance and genetic distance were correlated in *Cr. levior* B – but not in A, or any of the cryptic *Ca. femoratus* species. Interestingly, in *Cr. levior* A, in which we only found a single COI haplotype, the chemical diversity was very large compared to the uniformity we observed in the COI locus. Taken together, this in our opinion suggests that the CHC divergence may have played a role in species divergence – either during or after speciation. The main role of cuticular hydrocarbons is to serve as desiccation barrier but, especially in social insects, additionally play a role in communication and as mating cues (Thomas & Simmons 2008). They therefore have been discussed as possible 'magic traits', i.e. traits that affect both ecological adaptation and mate signalling (Smadja & Butlin 2009; Chung & Carroll 2015), which can

be mediated by a single gene only (Chung *et al.* 2014). Changes in such traits will often lead to assortative mating and ultimately to speciation (Chung & Carroll 2015). In *Timema* stick insects, speciation events were generally associated with a divergence in CHC profiles, however, it remained unclear whether speciation followed CHC divergence or if CHC profiles diverged due to selection during the evolution of reproductive isolation (Schwander *et al*. 2013). The same holds true for both cryptic species pairs in *Crematogaster* and *Camponotus.* The surprisingly high chemical divergence, combined with low genetic diversity (at least in *Camponotus*), might be indicative for a role of CHCs in species divergence. But it remains to be elucidated whether CHCs played a role in the speciation event itself by inducing assortative mating, by reinforcing sexual selection after the speciation event or by niche partitioning, i.e. adaptation to a yet unknown factor.

CONCLUSION

We could conclusively show that both *Crematogaster levior* and *Camponotus femoratus* split into two morphologically nearly indistinguishable cryptic species. It remains unclear how speciation took place in the two genera, but the strong separation in cuticular hydrocarbon profiles suggests that they are involved in mediating species divergence. Since *Crematogaster levior* and *Camponotus femoratus* are only found in mutualistic associations, we were rather surprised to find no partner preferences as indication for co-speciation in this mutualistic complex. Moreover, the highly different population structures between and within genera point to a rather loose relationship among the mutualists, whereas similar population structures would be expected if there was a strict partner specialization. Future studies should investigate partner choice and recognition, the evolution of the distinct chemotypes, the phylogeography of the species, as well as genome wide patterns of selection to shed further light on this highly interesting association and its players. This will help to deepen our knowledge on the effect of mutualistic interactions on species divergence.

AUTHOR CONTRIBUTIONS

TS, BF, and FM: conceived the study. JH, PPS, JO, BF, and FM: collected the specimens and field data. PPS, JS, TB, TS, and FM: did the chemical analyses and respective data analyses. JS and HB: did the morphological measurements and corresponding statistical analysis. JH, HW, and BF: performed sequencing and genetic analyses. JH, PPS, BF, and FM: wrote the first version of the manuscript. HB and TB: added to the methods and results sections. All authors contributed to writing this version and approved the submission.

ACKNOWLEDGEMENTS

We thank Philippe Cerdan and Aurelie Dourdain for research per- missions in the Hydreco Lab Petit Saut and the Paracou Research Station, respectively. Similarly, we thank Patrick Châtelet, Philippe Gaucher, and Dorothée Deslignes for permission to sample in the Les Nouragues Reserve. Further on, we thank Heike Stypa for supporting us in preparing the chemical samples. We thank Aidin Niamir for his helpful advice regarding climate data analysis. Financial support for this study was provided by the German Science Foundation (DFG) as a grant to Barbara Feldmeyer (FE 1333/7-1), Thomas Schmitt (SCHM 2645/7-1), and Florian Menzel (ME 3842/5-1) and a grant managed by the French Agence Nationale de la Recherche (CEBA, ref. ANR-10- LABX-25-01) to Jérôme Orivel. The publication of this article was funded by the Open Access Fund of the Leibniz Association. Finally, we thank Markus Pfenninger and two anonymous reviewers for providing helpful comments on an earlier version of this manuscript.

SUPPLEMENTARY MATERIAL

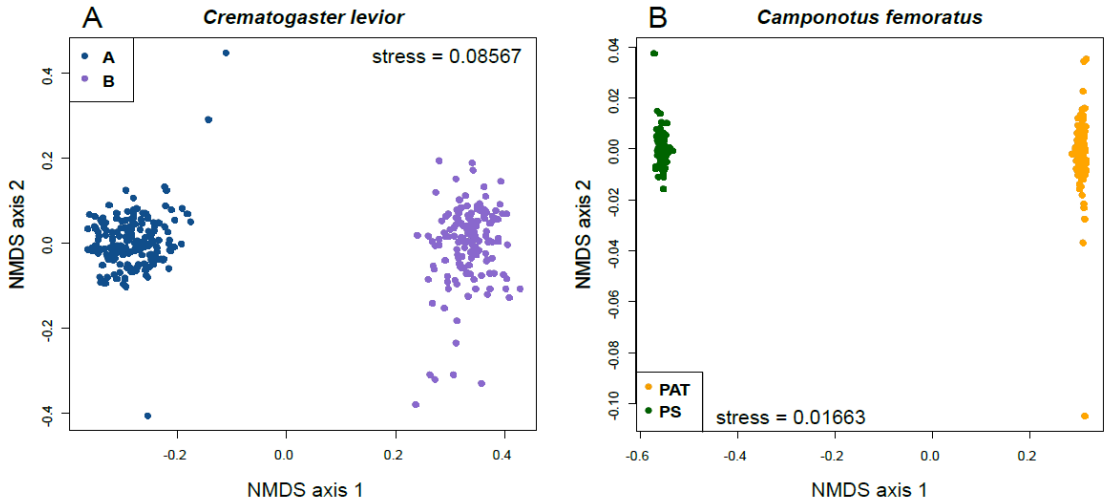


Figure S1.1: Non-metric multidimensional scaling (NMDS) ordinations of the CHC profiles of the cryptic *Cr. levior* and *Ca. femoratus* species. (**A**) Shows an NMDS ordination of *Cr. levior* CHC profiles (*Cr. levior* A: blue; *Cr. levior* B: purple). Each dot represents one colony. (**B**) Shows a similar ordination for *Ca. femoratus* CHCs (*Ca. femoratus* PAT: yellow; *Ca. femoratus* PS: green).

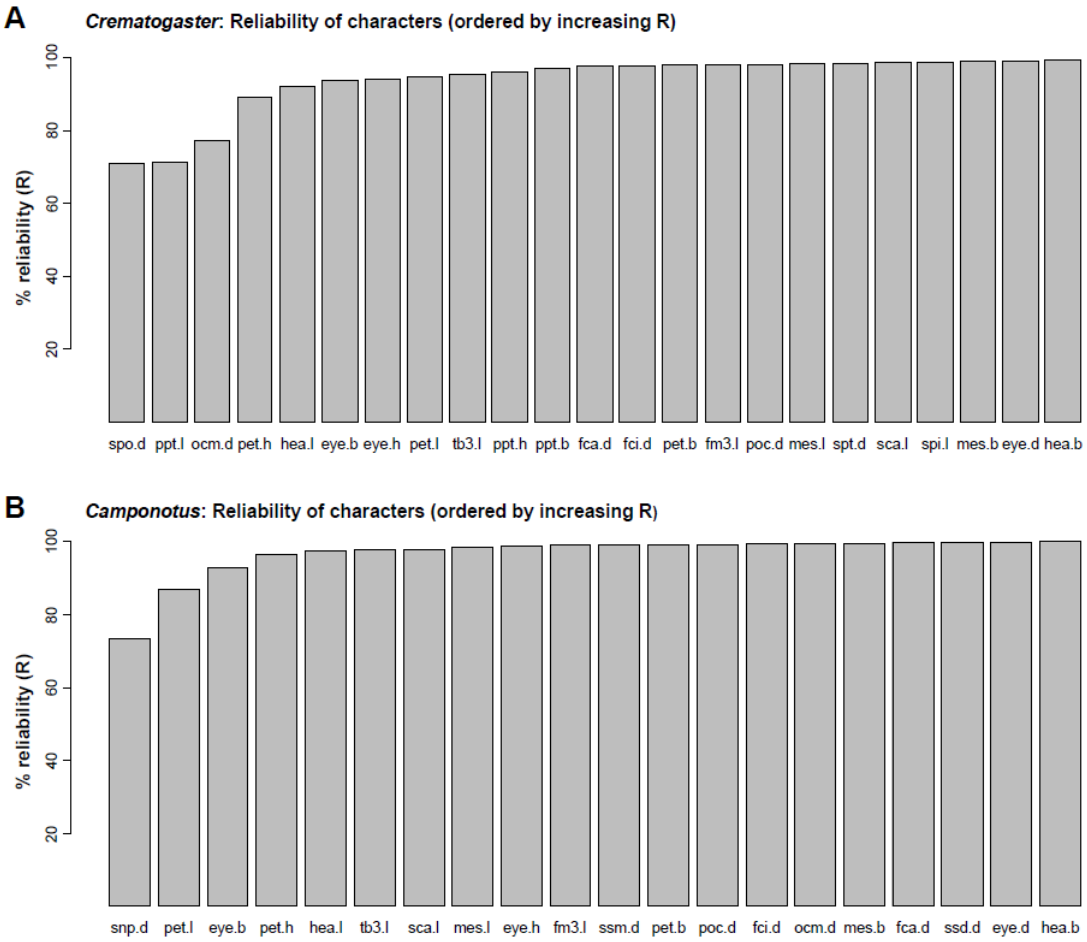


Figure S1.2: Reliability of morphological characters. The barplot show the reliability (R) of characters measured in *Crematogaster* (A) and *Camponotus* (B). Characters with a reliability < 85% were not included into the multivariate ratio analysis.

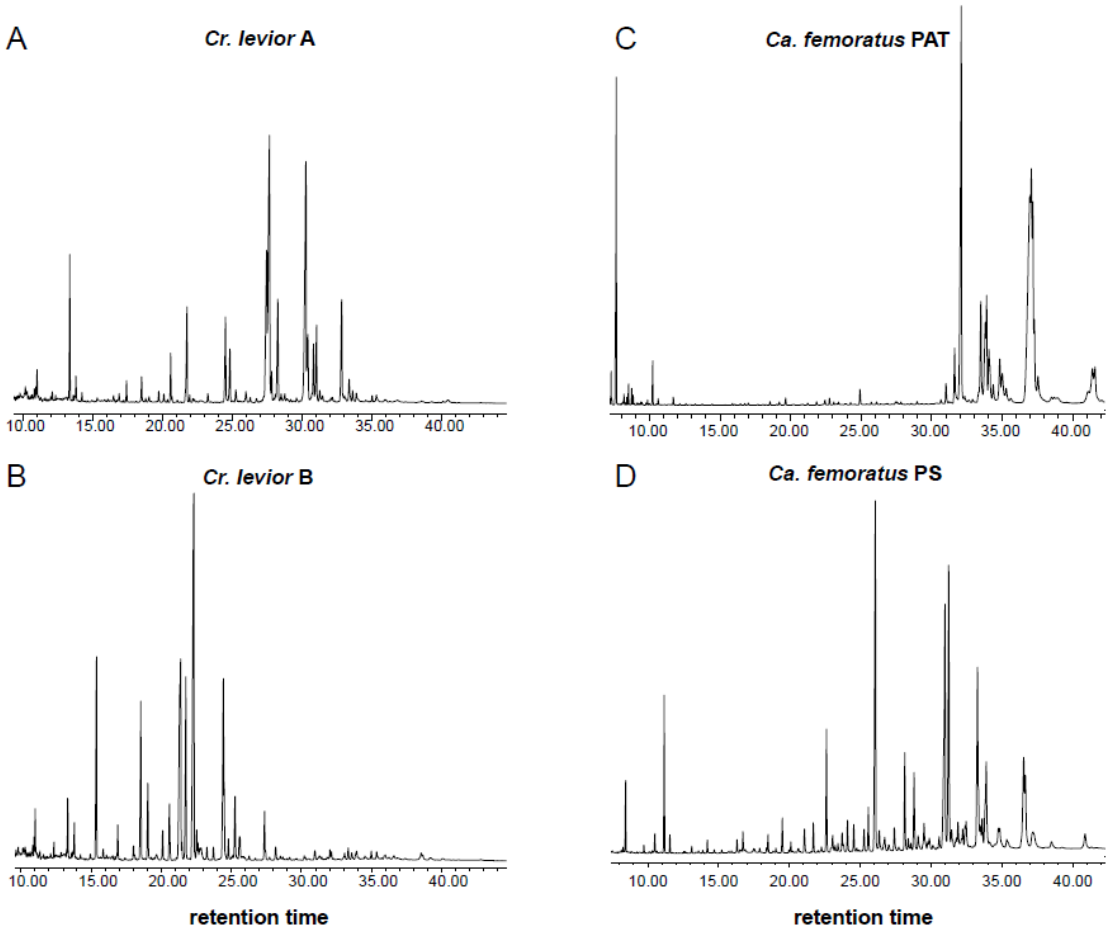


Figure S1.3: Representative chromatograms of the cuticular hydrocarbon profiles of the cryptic species. (**A+B**) Show chromatograms of the CHC profiles of *Cr. levior* A and B. (**C+D**) Show chromatograms of the CHC profiles of *Ca. femoratus* PAT and PS.

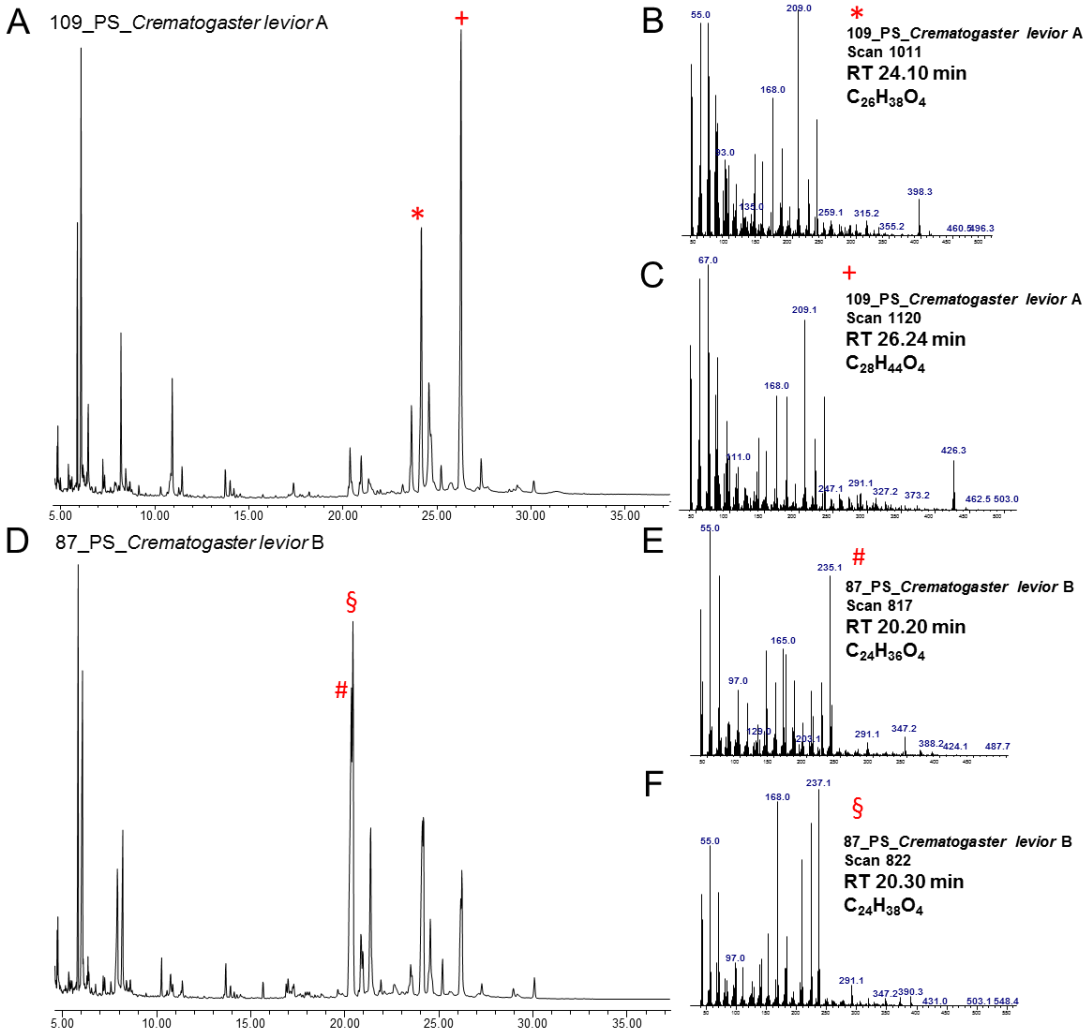


Figure S1.4: Representative chromatograms of the polar substances of *Cr. levior* A and B and mass spectra of the most abundant substances. (**A-C**) Show a representative chromatogram of the polar substances of *Cr. levior* A (**A**), and the mass spectra of the two most common substances in this species (**B+C**). (**D-F**) Show a representative chromatogram of the polar substances of *Cr. levior* B (**D**), and the mass spectra of the two most common substances in this species (**E+F**).

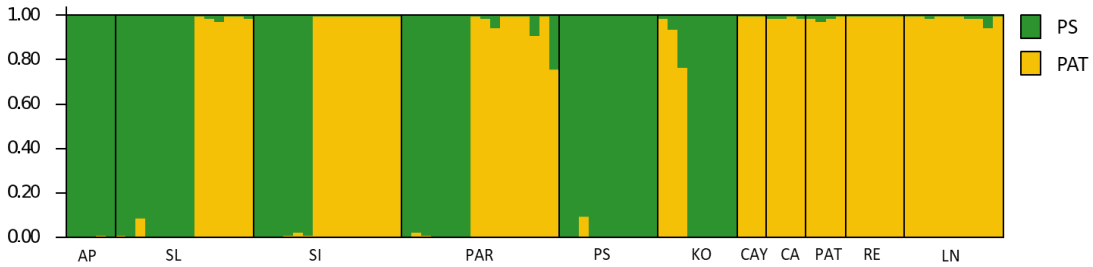


Figure S1.5: STRUCTURE analysis of 93 individuals of *Camponotus femoratus*, for which all four nuclear loci were sequenced successfully. For abbreviations of population names refer to Table 1.1.

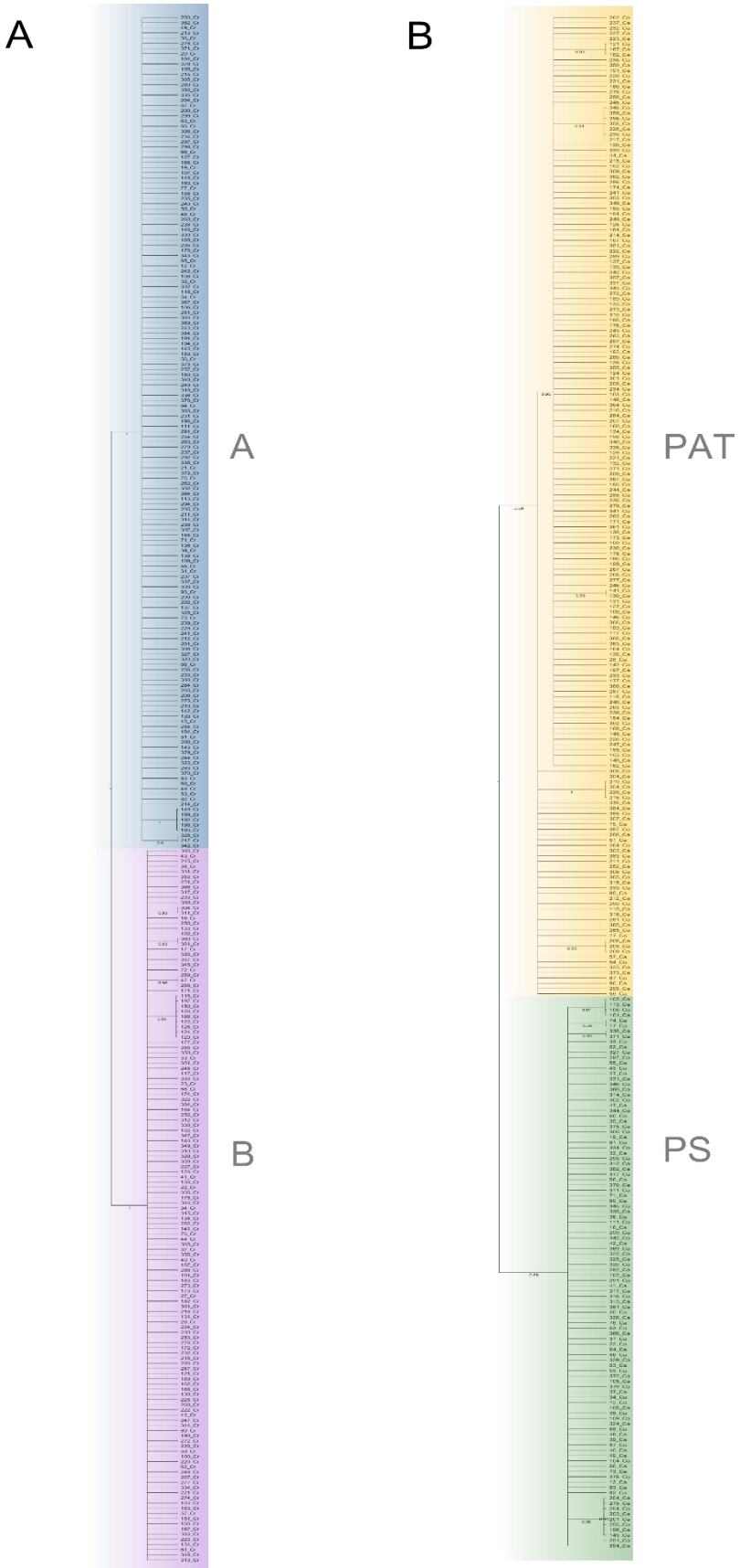


Figure S1.6: MrBayes Phylogeny of mitochondrial COI of (**A**) *Crematogaster levior s*pecies and (**B**) *Camponotus femoratus* species.

Table S1.1: Reliability of body measures. The table shows the variable codes (as described in Tables S1.2 and S1.3), the reliability R as well as the lower and upper 95 % confidence intervals (LCI and UCI).

variable code	R	LCI	UCI	variable code	R	LCI	UCI
Crematogaster				***Camponotus***			
spo.d	70.99	48.40	82.94	snp.d	73.32	52.22	84.55
ppt.l	71.30	48.91	83.13	pet.l	87.15	76.12	92.70
ocm.d	77.16	58.77	86.67	eye.b	92.73	86.27	95.90
pet.h	89.07	79.68	93.71	pet.h	96.45	93.23	98.01
hea.l	91.81	84.68	95.31	hea.l	97.52	95.25	98.61
eye.b	93.64	88.05	96.36	tb3.l	97.68	95.57	98.70
eye.h	94.19	89.07	96.68	sca.l	97.91	96.00	98.83
pet.l	94.91	90.40	97.09	mes.l	98.36	96.86	99.08
tb3.l	95.47	91.44	97.41	eye.h	98.69	97.48	99.27
ppt.h	96.04	92.52	97.74	fm3.l	98.92	97.92	99.39
ppt.b	96.97	94.25	98.27	ssm.d	99.23	98.52	99.57
fca.d	97.77	95.77	98.73	pet.b	99.28	98.61	99.60
fci.d	97.81	95.85	98.76	poc.d	99.28	98.62	99.60
pet.b	97.91	96.04	98.81	fci.d	99.41	98.86	99.67
fm3.l	98.01	96.23	98.87	ocm.d	99.45	98.94	99.69
poc.d	98.11	96.41	98.93	mes.b	99.73	99.49	99.85
mes.l	98.29	96.75	99.03	fca.d	99.74	99.50	99.85
spt.d	98.33	96.83	99.05	ssd.d	99.76	99.54	99.87
sca.l	98.56	97.26	99.18	eye.d	99.89	99.79	99.94
spi.l	98.57	97.29	99.19	hea.b	99.96	99.92	99.98
mes.b	98.97	98.03	99.41				
eye.d	99.00	98.09	99.43				
hea.b	99.16	98.40	99.52				

Table S1.2: Morphological traits measured in *Cr. levior* species. The table gives the name, definition, abbreviation code and reference of the measured traits as well as the perspective and scale the photos were taken at. Traits written in italics were excluded from the analysis because they had a reliability < 85%.

Sequence	Perspective	Code	Name	Definition	Scale	Reference
000	Lateral	eye.h	Eye length	Max. diameter of the eye; measure long distance from posterior margin (higher; point near head margin; direction to mesosoma) to anterior margin (lower point; near malar/ mandible margin); not outer ocellus; circle around them	200x	Csösz *et al.* 2014
001	Lateral	eye.b	Eye breadth	Min. diameter of the eye (measure in a 90° ankle to EL, posterior most point (directed to antennae) to anterior most point); not ocellus; light border	200x	Csösz *et al.* 2014
002	*Lateral*	*ocm.d*	*Ocular-malar-distance*	*Min. distance from the eye (center of the light border around last ocellus) to a dark tip above the mandibular junction (mandibular junction sometimes is moved; not constant); above the mandibular junction is a dark tip which is constant*	*200x*	*Csösz et al. 2014*
003	Lateral	mes.l	Mesosoma length	Diagonal measuring from highest (posterior most) point of the back (rounded) to the anterior most point on the left of the mesosoma (to petiolar junction)	200x	Seifert 2008
004	Lateral	spi.l	Spine length	Measured from the center of the mesosomal stigma to the tip of the spine	200x	Seifert 2008
005	Lateral	pet.h	Petiole height	Measured from the anterior most point to the horizontal border of petiole to the posterior most point belonging to the node of the petiole	200x	Seifert 2008
006	Lateral	pet.l	Petiole length	Measured diagonal, from the posterior most point of the border of the petiolar junction (connected to mesosoma) to the anterior most point to the junction between petiole and post petiole (diagonal line going top left to down right)	200x	Seifert 2008
007	Lateral	ppt.h	Postpetiole height	Measured from the anterior most point to the horizontal border of postpetiole to the posterior most point belonging to the node of the postpetiole (equally to petiole)	200x	Csösz *et al.* 2014
008	*Lateral*	*ppt.l*	*Postpetiole length*	*Diagonal length of postpetiole; measure from the anterior most point of postpetiolar junction (connection with petiole) to the posterior most point at the junction of the gaster*	*200x*	*Csösz et al. 2014*
009	Dorsal	mes.b	Mesosom	Max. mesosoma breadth measured in	200x	Seifert 2008

			a breadth	dorsal view; lay point of max. breadth in a straight horizontal line		
010	Dorsal	spt.d	Max. spine distance	Distance from tip of left to the tip of the right spine in the a horizontal line	200x	this study
011	*Dorsal*	*spo.d*	*Min. spine distance*	*From a "sitting" position looking dorsal/frontal face on the back of the ant; for the minimal distance between spines we measured the margin of the spines; from the left side of the dark border (center) to the right center of the border (border = connection with mesosoma)*	*200x*	*this study*
012	Dorsal	pet.b	Petiole breadth	Max. petiole breadth measured in dorsal view; lay point of max. breadth in same horizontal line (same as for mesosomal breadth)	200x	Seifert 2008
013	Dorsal	ppt.b	Postpetiole breadth	Max. Postpetiole breadth measured in dorsal view; lay point of max. breadth in same horizontal line (same as petiole and mesosomal breadth)	200x	Csösz *et al*. 2014
014	Frontal face	hea.l	Head length	Anterior most point of median clypeal margin to middle of posterior margin of the head, vertical line; full face view	200x	Csösz *et al*. 2014
015	Frontal face	poc.d	Postocular-distance	Measure the distance behind the eyes of the ant (in full face view; posterior most point of both eyes (not ocellus; border around eyes) connected with a horizontal line)	200x	Csösz *et al*. 2014
016	Frontal face	hea.b	Max. head breadth	Measure head breadth from the outer side of the eyes (max. head breadth includes eyes (again not ocellus), full face view)	200x	Seifert 2008
017	Frontal face	eye.d	Min. head breadth	Measure head breadth from the inner side of the eyes (min. head breadth excluding eyes (border around eyes), full face view)	200x	Seifert 2008
018	Frontal face	fca.d	Max. distance of frontal lobes	Max. distance of frontal carinae to examine max. distance between antennae (in full face view) measured from the center of the dark border (rounded condyle) on the left, to the center of the border (rounded condyle) on the right frontal lobe	200x	Csösz *et al*. 2014
019	Frontal face	fci.d	Min. distance of frontal lobes	Min. distance of frontal lobes measured on the basis (posterior to FL); take center of left dark border of the constriction, to center of the right dark border (point posterior to the margin of the torculus)	200x	Csösz *et al*. 2014
020	Antenna	sca.l	Scape length	Measured from proximal point (margin of scape) to distal end (and of scape,	200x	Csösz *et al*. 2014

				junction; start of funiculus) of scape of the antenna; points in horizontal line		
021	Legs	fm3.l	Femur length	Trochanter/femur knuckle (center) to femur/tibia knuckle (femur side) center in a horizontal line	200x	this study
022	Legs	tb3.l	Tibia length	Tibia tip (knuckle of femur/tibia) in lateral view to center of tibia/tarsi knuckle	200x	this study

Table S1.3: Morphological traits measured in *Ca. femoratus* species. The table gives the name, definition, abbreviation code and reference of the measured traits as well as the perspective and scale the photos were taken at. Traits written in italics were excluded from the analysis because they had a reliability < 85%.

Sequence	Perspective	Code	Name	Definition	Scale	Reference
000	Lateral	eye.h	Eye length	Max. diameter of the eye; measure long distance from posterior margin (higher; point near head margin; direction to mesosoma) to anterior margin (lower point; near malar/ mandible margin); not outer ocellus; circle around them	200x	Csösz *et al*. 2014
001	Lateral	eye.b	Eye breadth	Min. diameter of the eye (measure in a 90° ankle to EL, posterior most point (directed to antennae) to anterior most point); not ocellus; light border	200x	Csösz *et al*. 2014
002	Lateral	ocm.d	Ocular-malar-distance	Min. distance from the eye (center of the light border around last ocellus) to a dark tip above the mandibular junction (mandibular junction sometimes is moved; not constant); above the mandibular junction is a dark tip which is constant	200x	Csösz *et al*. 2014
003	Lateral	pet.h	Petiole height	Measure tip of the node of the petiole to the anterior end of the petiole (building a horizontal line)	200x	Seifert 2008
004	Lateral	pet.l	Petiole length	Diagonal length measured from posterior most point of petiolar junction to anterior most end of a tip on the dorsal end of the petiole	200x	Seifert 2008
005	*Lateral*	*snp.d*	*Stigma-spine of petiole distance*	*Center of stigma of petiole to tip of spine (petiolar spine)*	*200x*	*Seifert 2008*
006	Lateral	ssm.d	Stigma distance	Posterior (higher) stigma center to anterior (lower; near petiole) center of stigma (both stigma on mesosoma)	200x	Seifert 2008
007	Dorsal	mes.b	Mesosoma breadth	Max. mesosoma breadth measured in dorsal view; lay point of max. breadth in a straight horizontal line	200x	Seifert 2008

008	Dorsal	ssd.d	Mesosoma stigma distance	Small mesosoma max. stigma distance (measured of the center of stigma) building a horizontal line in dorsal view	200x	this study
009	Dorsal	pet.b	Petiole breadth	Max. petiole breadth measured in dorsal view; lay point of max. breadth in same horizontal line (same as for mesosomal breadth)	200x	Seifert 2008
010	Frontal face	poc.d	Postocular-distance	Measured the distance behind the eyes of the ant (in full face view; posterior most point of both eyes (not ocellus; border around eyes) connected with a horizontal line)	200x	Csösz *et al*. 2014
011	Frontal face	hea.b	Max. head breadth including eyes	Measure head breadth from the outer side of the eyes (max. head breadth includes eyes (again not ocellus), full face view)	200x	Seifert 2008
012	Frontal face	eye.d	Min. head breadth excluding eyes	Measure head breadth from the inner side of the eyes (min. head breadth excluding eyes (border around eyes), full face view)	200x	Seifert 2008
013	Frontal face	fca.d	Max. distance frontal lobes	Max. distance of frontal carinae to examine max. distance between antennae (in full face view) measured from the center of the dark border (rounded condyle) on the left, to the center of the border (rounded condyle) on the right frontal lobe	200x	Csösz *et al*. 2014
014	Frontal face	fci.d	Min. distance frontal lobes	Min. distance of frontal lobes measured on the basis (posterior to FL); take center of left dark border of the constriction, to center of the right dark border (point posterior to the margin of the torculus)	200x	Csösz *et al*. 2014
015	Dorsal	mes.l	Mesosoma length	Diagonal measuring from highest (posterior most) point of the back (rounded) to the anterior most point on the left of the mesosoma (to petiolar junction)	100x	Seifert 2008
016	Frontal face	hea.l	Head length	Anterior most point of median clypeal margin to middle of posterior margin of the head, vertical line; full face view	100x/150x	Csösz *et al*. 2014
017	Antenna	sca.l	Scape length of antennae	Measured from proximal point (margin of scape) to distal end (and of scape, junction; start of funiculus) of scape of the antenna; points in horizontal line	100x/150x	Csösz *et al*. 2014
018	Legs	fm3.l	Femur length	Trochanter/femur knuckle (center) to femur/tibia knuckle (femur side) center in a horizontal line	100x/150x	this study
019	Legs	tb3.l	Tibia	Tibia tip (knuckle of femur/tibia) in	100x/1	this study

			length	lateral view to center of tibia/tarsi knuckle	50x	

Table S1.4: PCR mastermix and sequencing reaction.

PCR mastermix	Volume	Sequencing reaction	Volume
10x Buffer B (molegene)	1.0 µl	Big Dye Terminator Mix	0.16 µl
$MgCL_2$ 25mM (molegene)	1.0 µl	5x Buffer B (molegene)	1.84 µl
dNTP-Mix 2mM each (molegene)	0.1 µl	ddH_2O	6.50 µl
Primer 10 pmol/µl)	0.2 µl of each primer	Primer (10 pmol/µl)	0.5 µl of each primer
Taq-Polymerase (molegene)	0.1 µl	PCR product	1.0µl
ddH_2O	6.5 µl		

Table S1.5: Overview of all test primer pairs. *(Ströher et al. 2013)*

Primer	Primer length	Primer sequence	
ant. 1F	27	CCTTCGTGCCTAYGAGAATAGYGTTAC	Resulting sequences were variable enough for further analyses
ant. 1R	21	AACGACGTCSGACGGTTCCAT	
ant. 389F	21	ACGGACCCCACATTGAGAAGAAC	
ant. 389R	21	CYTTACCCACCTCCTCCACCA	
ant. 1087 F	21	ACCAGCAGAGGCTGGACGTGA	
ant. 1087 R	27	GCCAAGTTGATTGTGTACGAACTTTCT	
ant. 1401F	22	GYAGGAAGGACGCTCTTAATCT	
ant. 1401R	26	AAGCTTATCTCTAGGAAACTCCCATC	
ant. 1225 F	26	TAATACRACTGAAGAGAGACCAGGAG	Not used for further analyses as resulting sequences were either not variable or did not amplify in PCR
ant. 1225 R	27	GACTAGATCCTAAGCTAGAGAGRCTGG	
ant. 1281 F	23	GACGCAGGTTGYAACGAAATCAC	
ant. 1281 R	24	GCCRCTAATATCCAGCTTCACGAG	
ant. 384 F	27	TAGTAGTCGAAGGAGTCATACCAAAGG	
ant. 384 R	20	TGYGTGTTCGATGCCGTTGA	
ant. 965 F	24	AGTTCAAGGTTCACCGGTGCCTAA	
ant. 965 R	25	GAGAAGGYGAAYTTAAAGACTGATG	
ant. 1503F	21	GRTTYGCCTTCCAGGAGATCA	
ant. 1503R	23	AAGTAGTCCAGGCAGAACCACAC	
ant. 202F	26	CCYATCAACTCTGTTAATATCGAACG	
ant. 202R	22	GACACAATGTTGGAAGCCCTTG	
ant. 263F	27	GACTAGCTCAGAATCACACTCTTCCAC	
ant. 263R	24	GTTGTTTTGGWGGCAATATTGGAG	
ant. 346F	23	GTGGTCCACCATCCGTKGGATCT	
ant. 346R	26	GGATTGTTTTGTGTAATCTGCGTTCG	
ant. 505F	24	CCTCAGATGAAGTTYCGAGTTCC	
ant. 505R	26	TAAYCCGRACACCCTCACTTTATACG	
ant. 839F	25	CAATGGCGATTTACAACGAATTTCT	
ant. 839R	22	CAGGCANAGCAGCAATGTGACG	

Table S1.6: Tajima's D statistics. Given are values within all sampled populations of *Crematogaster levior* B and *Camponotus femoratus* PS and PAT. Bold characters indicate statistical significance ($p < 0.05$) based on a permutation test. *Crematogaster levior* A is not shown as there was only a single haplotype found at all locations.

	***Cr. levior* B**		***Ca. femoratus* PS**		***Ca. femoratus* PAT**	
	Tajima's D	p	Tajima's D	p	Tajima's D	p
AP	0	1	0	1	0	1
SL	0	1	**-1.513**	**0.033**	**-1.486**	**0.04**
SI	-1.088	0.189	0	1	1.284	0.888
PS	0	1	0.545	0.764	0	1
PAR	-1.149	0.161	-1.142	0.145	**-2.072**	**0.003**
LN	0.713	0.869	0	1	**-2.107**	**0.002**
KO	1.225	0.943	-0.817	0.321	0	1
MT	-	-	0	1	0	1
CA	0	1	-	-	0.944	0.843
CAY	0	1	-	-	2.125	0.986
RE	-	-	0	1	2.192	0.993
PAT	-0.774	0.218	-	-	-0.657	0.297

CHAPTER 2

Influence of Mutualistic Lifestyle, Mutualistic Partner, and Climate on Cuticular Hydrocarbon Profiles in Parabiotic Ants

Philipp P. Sprenger, Juliane Hartke, Barbara Feldmeyer, Jérôme Orivel, Thomas Schmitt, Florian Menzel

Published as:
Sprenger PP, Hartke J, Feldmeyer B, Orivel J, Schmitt T, Menzel F (2019). Influence of Mutualistic Lifestyle, Mutualistic Partner, and Climate on Cuticular Hydrocarbon Profiles in Parabiotic Ants. Journal of Chemical Ecology. 45:741–754.

ABSTRACT

A vital trait in insects is their cuticular hydrocarbon (CHC) profile, which protects the insect against desiccation and serves in chemical communication. Due to these functions, CHC profiles are shaped by both climatic conditions and biotic interactions. Here, we investigated CHC differentiation in the neotropical parabiotic ant species *Crematogaster levior* and *Camponotus femoratus*, which mutualistically share a nest. Both consist of two cryptic species each (*Cr. levior* A and B and *Ca. femoratus* PAT and PS) that differ genetically and possess strongly different CHC profiles. We characterized and compared CHC profiles of the four cryptic species in detail. Our results suggest that *Cr. levior* A, *Ca. femoratus* PAT and *Ca. femoratus* PS adapted their CHC profiles to the parabiotic lifestyle by producing longer-chain CHCs. At the same time, they changed their major CHC classes, and produce more alkadienes and methyl-branched alkenes compared to *Cr. levior* B or non-parabiotic species. The CHC profiles of *Cr. levior* B were more similar to related, non-parabiotic species of the *Orthocrema* clade than *Cr. levior* A, and the chain lengths of B were similar to the reconstructed ancestral state. Signals of both the parabiotic partner (biotic conditions) and climate (abiotic conditions) were found in the CHC profiles of all four cryptic species. Our data suggest that mutualisms shaped the CHC profiles of the studied species, in particular chain length and CHC class composition. Beside this, signals of the parabiotic partners indicate potential impacts of biotic interactions, via chemical mimicry or chemical camouflage.

Keywords Adaptation. Chemical communication. Cryptic species. Formicidae. Mimicry. Mutualism. Parabiosis

INTRODUCTION

Cuticular hydrocarbons (CHCs) cover the surface of nearly all terrestrial arthropods and function as barriers against water loss and as agents in chemical communication. They can be classified in three major groups: straight-chain n-alkanes, methyl-branched alkanes and unsaturated alkenes (Blomquist 2010). The substances produced by an insect are usually species-specific but can differ in their relative proportions (Martin *et al*. 2008). The composition of CHC profiles is also shaped by abiotic selection pressures such as temperature or humidity, but also by communication requirements (Chung & Carroll 2015). Indeed, ant species from habitats with higher rainfall produce rather unusual substance classes like alkadienes and methyl-branched alkenes, which are less effective in preventing water loss (Menzel *et al.* 2017a; van Wilgenburg *et al*. 2011). However, several studies revealed that relative CHC profile composition is rather flexible and organisms are able to respond plastically to temperature and humidity variation on a rather short-term basis (Menzel *et al.* 2018; Sprenger *et al*. 2018; Stinziano *et al.* 2015; Wagner *et al*. 2001) and can be influenced by biotic factors like e.g. diet, parasites or pathogens (Otte *et al.* 2018).

Well-known examples of how biotic interactions shape the evolution of CHC profiles come from antagonistic interactions. Many parasites and predators mimic the CHC profiles of their hosts (or prey) to avoid being detected. This phenomenon is especially common in parasites (myrmecophiles) or predators of ants. Some species use chemical mimicry, i.e. the active production of similar substances, to impede nestmate recognition by social insects (Lenoir *et al*. 2001; Bagnères & Lorenzi 2010). As a second mechanism, the parasites can acquire their host's CHCs without producing them themselves either passively through the host or nesting material or actively like e.g. *Formicoxenus* ants that groom off their host's CHCs (chemical camouflage; Lenoir *et al*. 1997, 2001). A third possible way to avoid recognition is chemical insignificance, in which the parasite possesses only few recognition cues (low quantity of CHC), substances which are hard to distinguish (e.g. n-alkanes) or additional chemical substances (e.g. alkaloids) to cover the recognition cues (Kleeberg *et al*. 2017; Neupert *et al*. 2018). Some species also use CHCs or different chemical com- pounds as appeasement allomones to get accepted in foreign colonies or reduce interspecific aggression (Mori *et al.* 2000; Menzel *et al.* 2013; Elia *et al.* 2018). Exploitation by parasites can also be a selection pressure on the host leading to diversification in CHC profiles as a counter-adaptation (Jongepier & Foitzik 2016).

Beside host-parasite interactions, mutualisms can also exert selection on CHC profiles (Menzel & Schmitt 2012a). Mutualistic interactions are often drivers of the evolution of novel phenotypic traits and are expected to speed up their emergence (Herre *et al*. 1999; Hoeksema & Bruna 2000; Guimarães *et al*. 2011). An example of a remarkable mutualism is the so-called parabiosis in ants, in which two different species live in a common nest while

keeping their brood separated (Orivel *et al*. 1997; Menzel *et al.* 2008b). Parabioses are restricted to tropical habitats and often include species of the genus *Crematogaster* (Swain 1980; Davidson 1988; Orivel *et al.* 1997; Vantaux *et al*. 2007; Menzel *et al.* 2008b; but see Parmentier *et al*. 2017). They are characterized by high interspecific tolerance (Menzel *et al.* 2008a,b), which is unusual for ants of different species or colonies. The inter- specific tolerance in parabiotic ants is most likely achieved through adaptations of the CHC profile including increase in chain length and production of rare substance classes such as alkadienes or methyl-branched alkenes (Menzel *et al*. 2008a; Menzel & Schmitt 2012).

In a recent study, we presented strong evidence for cryptic speciation in the neotropical parabiotic species *Crematogaster levior* and *Camponotus femoratus* (Hartke *et al.* 2019a). Both exhibit two different chemical morphs, which also differ genetically and, albeit only slightly, in their morphology. In addition, *Cr. levior* was only quite recently taxonomically separated from *Cr. carinata*. Both belong to a species complex within the *Orthocrema* clade, one of three subgenera of *Crematogaster*, which is suspected to contain even more cryptic species (Longino 2003). The two species live in commonly shared nests, so-called ant gardens (Davidson 1988; Orivel & Leroy 2011). These consist of specialized epiphytic plants, whose seeds were carried in by *Ca. femoratus* (Youngsteadt *et al.* 2008) and whose root systems are coated with carton by the ants and serve as the actual ant nest. *Camponotus femoratus* profits from the ability of *Cr. levior* to discover prey and was shown to follow their trail pheromones (Menzel *et al*. 2014; Vantaux *et al.* 2007). *Crematogaster levior* on the other hand benefits from *Ca. femoratus* building nest structures, planting epiphytic seeds, responding to damages on the ant garden plants and aggressively defending the nest against attacking vertebrates (Vantaux *et al.* 2007; Youngsteadt *et al*. 2008; Vicente *et al.* 2014; Leal *et al.* 2017) including field biologists (personal observations).

In this study, we describe the chemical diversity of parabiotic *Camponotus* and *Crematogaster* species in French Guiana. We characterized the cryptic species in both genera and investigated in detail how they differed from each other. Furthermore, we report additional variants that did not fit into these cryptic species. By comparing the chemical profiles of the cryptic *Cr. levior* and *Ca. femoratus* species to other related species, we inferred CHC changes related to the parabiotic lifestyle and reconstructed the ancestral state for CHC chain length in *Crematogaster*. Finally, we investigated how the CHC profiles were influenced by climatic conditions (due to the need for waterproofing) and by the parabiotic partner (due to mimicry or substance transfer).

MATERIALS AND METHODS

Sample Collection

We collected workers of the arboreal ant species *Crematogaster levior* (N = 332 colonies) and *Camponotus femoratus* (N = 306 colonies) in 13 different locations in French Guiana from August to October 2016 (Table S2.1, Fig. S2.1). Furthermore, we also collected workers of *Crematogaster* found in parabiosis or trail-sharing with *Camponotus* species, as well as their partners to infer if there were more potentially cryptic species. We noted the GPS coordinates of the exact sampling locations for each colony. We freeze-killed the collected workers at -20°C and subsequently extracted the CHCs by immersing groups of 10 *Cr. levior* (small, dry mass: 0.074 ± 0.016 mg) or 5 *Ca. femoratus* workers (large, 0.922 ± 0.092 mg), in hexane. After 10 minutes the ants were removed from the extracts using freshly cleaned forceps.

Chemical Analyses

Crematogaster CHC samples were purified using silica columns (Chromabond, SiOH 1mL/100mg, Macherey-Nagel, Düren, Germany) before analysis since they contained hereto unknown polar compounds which concealed the hydrocarbon profile. All samples were concentrated to approximately 20 µl under a gentle N_2 stream and were analysed using gas chromatography-mass spectrometry (GC-MS). We used a Zebron Inferno DB5-MS capillary column (length 30 m, Ø 0.25 mm, 0.25 µm coating, Phenomenex Ltd., Aschaffenburg, Germany) in an Agilent gas chromatograph (7890A, Agilent Technologies, Santa Clara, CA, USA) with Helium as carrier gas at a flow rate of 1.2 mL per minute. The mass spectrometer (5975C, Agilent Technologies) ran with electron ionization (EI) at 70 eV.

We used a Programmed Temperature Vaporization (PTV) injection method for the *Cr. levior* CHC extracts. This injection method was used to inject higher extract quantities, which is necessary to achieve good CHC detection despite the low CHC quantities of this species. With the PTV method, the sample is injected, and the solvent is then rapidly vaporized. 4 µl of the extracts were injected at 40°C and this temperature was held constant for 2 minutes. Then, the oven was heated with 60°C per minute to 200°C and following this with 4°C per minute to 320°C. This temperature was then kept constant for 10 minutes. Similarly, 2 µl of *Ca. femoratus* extracts were injected at 60°C using the splitless method. The oven heated with 60°C per minute up to 200°C and then with 4°C per minute to 300°C, which were kept constant for 10 minutes.

We integrated the chromatograms manually using MSD ChemStation (E.02.02.1431, Agilent Technologies). Afterwards substances were aligned in Microsoft Excel 2010. We excluded all non-hydrocarbon substances as well as entire substances if their average abundance was below 0.1% or if they were present in less than 20% of the samples of the

respective cryptic species. However, we included substances that had multiple double bonds if other alkenes of the same chain length were present, even if they were present in less than 20% of the samples (as they might be varieties coming from the same biosynthetic pathway). CHCs were identified using diagnostic ions and Kovats indices calculated based on a standard series of *n*-alkanes (Carlson *et al.* 1998, Table S2.2 and S2.3). To identify their positions in alkenes and alkadienes, we performed methylthiolation of the carbon-carbon double bonds using DMDS (dimethyl-disulfide, Sigma-Aldrich, St. Louis, MO, USA) (Attygalle 1998). Methyl-branched alkenes were identified by their mass spectra as previously described in other parabiotic ant species (Menzel *et al*. 2008a). Diagnostic ions for methyl-branched alkenes and exemplary mass spectra are shown in the supplementary material (Tab. S2.2, S2.3; Fig. S2.2).

Statistical Analyses

We aimed to find out which cuticular hydrocarbons differentiated the two cryptic species within *Cr. levior* and *Ca. femoratus* (respectively) from each other, and if they differed in overall chain length or CHC class composition. Next, we compared the CHC profiles of the cryptic *Cr. levior* species to other species of the *Orthocrema* clade (including an ancestral state reconstruction for CHC chain length) and those of *Ca. femoratus* to other *Camponotus* species. Finally, we investigated the impact of biotic (parabiotic partner) or abiotic factors (climate) on the CHC profiles. All statistical tests were performed using R version 3.6.0 (R Core Team 2018), and conducted separately for *Cr. levior* and *Ca. femoratus* samples.

As a first step, we assigned the colonies according to their CHC profiles. To this end, we first did a non-metric multidimensional scaling ordination (NMDS) based on Bray-Curtis dissimilarities using the R-package *vegan* (Oksanen *et al.* 2016) on the proportions of each cuticular hydrocarbon. In addition, we performed principal component analyses (PCA) on the centered log-ratio transformed compositional data (Aitchison 1982), and compared the clusters produced by each ordination method. We performed both methods since they are very different concerning their assumptions and the underlying algorithms (Brückner & Heethoff 2017); hence comparing both methods was a way to assess the robustness of the obtained results.

Afterwards, we tested whether the colonies could be correctly assigned to cryptic species using a random forest approach with 10,000 permutations (R-package *randomForest* (Liaw & Wiener 2002)). An advantage of this machine learning method is that it calculates the importance of the single substances for group assignment (Brückner & Heethoff 2017). In addition to this 'variable importance', we report the OOB (out-of-bag) error rate which is the classification error averaged across all training subsets. The training subsets (used to

train the classification algorithm) consist of the full data set minus one of the samples; thus, the number of subsets equals the number of samples.

We created a list of CHCs ordered by importance for group assignment and used the upper third of this list to perform an average-linkage hierarchical cluster analysis. This allowed us to identify which substances were more common in either of the cryptic species and thus to describe which compounds contributed most to the species differences (and hence defined species-specific profiles). Then, we used these CHCs to test whether certain hydrocarbon classes were enriched in one of the two cryptic species per genus, or whether the species-specific hydrocarbons differed systematically in chain length. To this end, for each substance class, we compared the number of species-specific CHCs between cryptic species using χ^2 tests. Further, we compared their chain lengths between cryptic species using Wilcoxon rank-sum tests. Beside these qualitative tests (that were based on the presence or absence of certain CHCs), we incorporated the abundances of the characteristic CHCs, and tested for differences between the cryptic species using Wilcoxon rank-sum tests.

In two additional analyses we compared the cryptic species to the profiles of other related species. *Crematogaster levior* was compared with 13 other species of its *Crematogaster* clade, *Orthocrema* (Blaimer 2012) using data from Menzel *et al*. (2017b). *Camponotus femoratus* was compared with 37 other *Camponotus* species using data from Menzel *et al*. (2017a) (to our knowledge, there is no robust phylogeny of *Camponotus* available, thus making it impossible to identify the species closely related to *Ca. femoratus*). We added 10 randomly chosen colonies from each of our cryptic species (R command *sample*) to keep the analysis concise. We used an NMDS ordination based on Bray-Curtis dissimilarities to compare the cryptic species to the other related species and to investigate if our samples cluster together with their conspecifics from original dataset. Additionally, we conducted an average-linkage cluster analysis for each dataset (R-command *hclustCBI* from the *fpc* package; based on the same distance matrices). Afterwards, we assessed the cluster stability using the mean Jaccard similarity of 100 bootstrap iterations (R-command *clusterboot* from the *fpc* package). Clusters with a mean Jaccard coefficient larger than 0.75 are usually considered stable (Hennig 2007).

The most prominent difference between *Cr. levior* A and B is the average chain length. Since a phylogeny for the *Orthocrema* clade including *Cr. levior* is available (Bayesian phylogeny based on five nuclear genes; Menzel *et al*. (2017b), we used it to reconstruct the ancestral state of this trait. Average chain length (weighed according to CHC proportion) was calculated for each species in the phylogeny based on data from this study or from Menzel *et al*. (2017b) (using species averages if applicable). The analysis was done using the commands *fastAnc* for estimating the ancestral states of the character at each node and its

95% confidence intervals as well as *contMap* to visualize them on the phylogeny (R-package *phytools;* Revell 2012).

To detect species-specific CHCs as signals of the parabiotic partner in their CHC profile, we determined the substances with the strongest differences according to their parabiotic partner. To this end, we ran univariate PERMANOVA analyses based on Euclidean distance for each compound separately with 999 permutations and added the partner identity as fixed factor. This analysis yields in pseudo-F values equivalent to univariate F-statistics as effect sizes (Anderson 2017). We identified the 10% compounds with the highest F-values, i.e. the highest differentiation according to the identity of the parabiotic partner. In the PERMANOVAs, minute random values with a normal distribution (mean ± SD: $10^{-8} \pm 10^{-8}$) were added to avoid samples with distances of zero in the distance matrix. The analysis was done for each of the four cryptic species separately.

Finally, we obtained Bioclim variables from CHELSA (Karger *et al.* 2017), which provides data with a resolution of about 1 km accuracy. Subsequently, we created subsets of the climatic variables according to the coordinates of colonies present for each species and performed a principal component analysis (PCA) to reduce the number of climate variables. Most of the variance was explained by annual precipitation and annual average temperature, which were negatively correlated (Fig. S2.3). Overall effects of partner, climate and their interaction were analysed using multivariate PERMANOVAs based on Bray-Curtis dissimilarities with 999 permutations on entire CHC profiles (separately for *Crematogaster* and *Camponotus*, but for the two cryptic species together in each genus) (R command *adonis*, package *vegan*). As fixed factors we used species identity, partner identity and the loadings of climate PC1. Similarly to the analysis for signals of the parabiotic partners, we also performed univariate PERMANOVAs based on Euclidean distances on each single CHC for the climate effects.

RESULTS

Differentiation Between CHCs of the Cryptic Species

In *Crematogaster levior*, we differentiated three different groups. *Crematogaster levior* A and B were already known from previous studies (Emery & Tsutsui 2013; Menzel *et al.* 2014; Hartke *et al*. 2019), and a third group of *Crematogaster* colonies, which in the following will be referred to as *Cr. levior* C (n = 10; Fig. 2.1 A+B). The CHC profile of *Cr. levior* C was rich in *n*-C27 (mean ± SD: 11.34 ± 8.84%), C27-alkenes (15.83 ± 14.07%) and C29-alkenes (13.36 ± 6.64%) (Table S2.2). These colonies were sometimes found in parabiosis with *Odontomachus mayi* or were sharing a trail with unidentified *Camponotus* species.

Morphological examination (by B. Blaimer, Raleigh NC, USA) showed that most C individuals had the carinae (keel-like crests) typical for *Cr. carinata* on their pronotum. One colony showed characteristics of both groups, A and B. It clustered with A in the PCA (Fig. 2.1 C) but with B in an NMDS ordination (Fig. 2.1 A), making unambiguous assignment difficult.

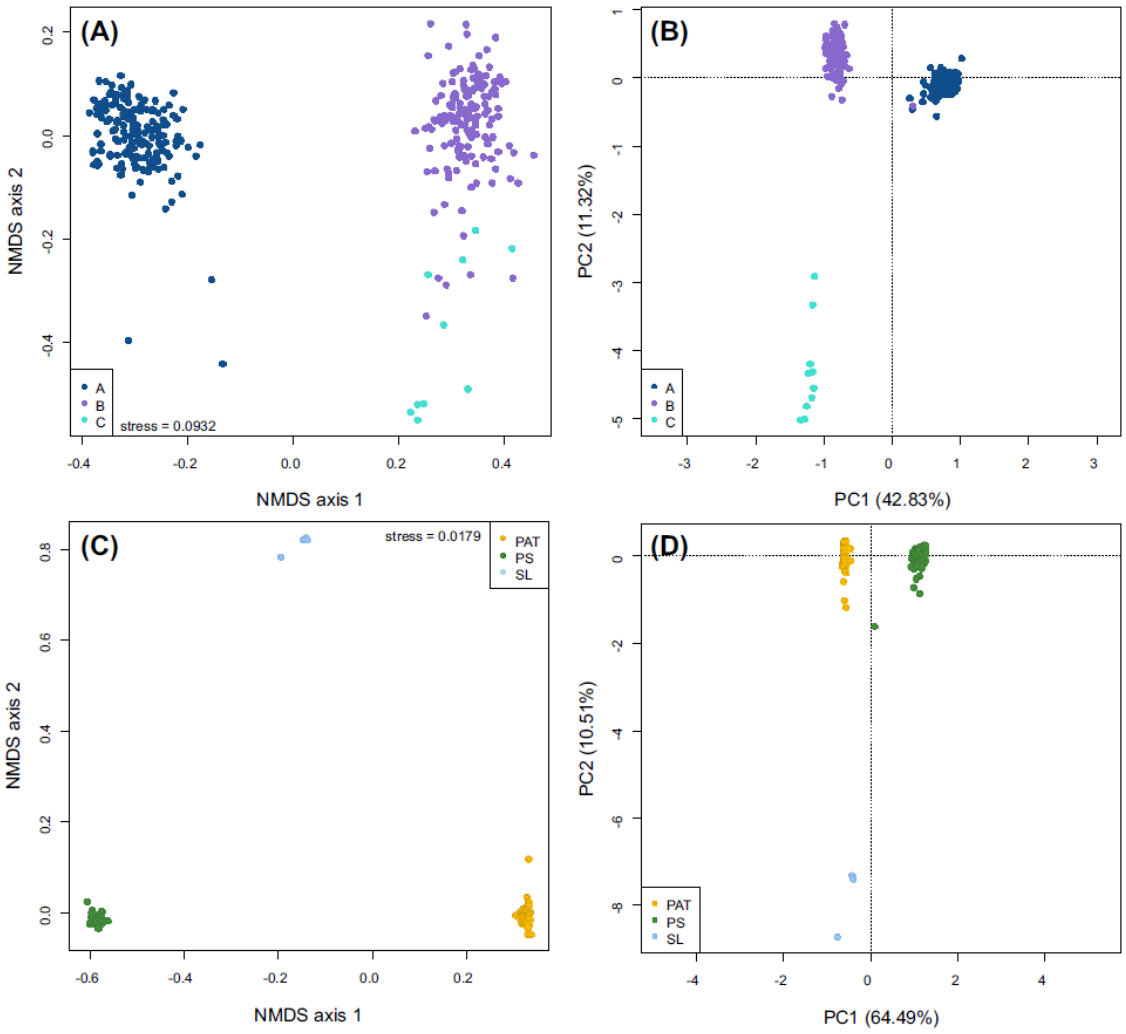


Figure 2.1: Differentiation of CHCs in *Cr. levior* and CHCs in *Ca. femoratus*. (**A**) & (**C**) Show non-metric multidimensional scaling (NMDS) ordinations based on Bray-Curtis dissimilarities of the CHC profile of *Cr. levior* and *Ca. femoratus*, respectively. Each data point represents the profile of one colony. (**B**) & (**D**) Show principal component analysis (PCA) ordinations of the CHC profiles of *Cr. levior* and *Ca. femoratus*. Similarly, each dot represents the profile of one colony.

In *Camponotus femoratus*, we could confirm the clear differentiation of the two cryptic species *Ca. femoratus* Patawa (PAT) and Petit Saut (PS) that were described in earlier studies (Menzel *et al.* 2014; Hartke *et al.* 2019a). However, four colonies found together with *Cr. levior* had profiles with several CHCs being not present in the cryptic *Ca. femoratus* species that formed a third cluster (Fig. 2.1 C, D). These colonies were tentatively identified as a

different species as they differ morphologically and genetically (at the COI locus) from *Ca. femoratus* (data not shown). However, they were still found in association with *Cr. levior*, which is why we present these data. As these colonies were restricted to our sampling locations in Saint-Laurent-du-Maroni (in the west of French Guiana, Fig. S2.1), we will refer to those as *Camponotus* sp. SL. The most common substances in their CHC profiles were *n*-C31 (31.33 ± 4.15%), 11-MeC31 (21.17 ± 4.41%), 2-;4-MeC30 (8.63 ± 1.47%) and an unknown unsaturated CHC (9.35 ± 0.69%) (Table S2.3).

Characterization of the CHC Profiles

For the following analyses, we used a dataset excluding *Cr. levior* C (n = 10) and *Camponotus* sp. SL (n = 4), due to their low sample sizes. For *Crematogaster levior* A and B (total N = 322), the random forest algorithm classified all A individuals correctly (n = 174) but misclassified one out of 148 colonies of *Cr. levior* B as A (error rate 0.0068%). Based on the random forest, we identified 28 CHCs defining *Cr. levior* A and 17 substances defining *Cr. levior* B (Fig. 2.2 A). Among these, a C37-alkadiene (variable importance: 28.06), a C39-alkadiene (27.81), *n*-C25 (27.27), a C27-alkene (26.72) and 9-C35-alkene (27.58) were most important. The C37- and C39-alkadiene and 9-C35-alkene were characteristic for *Cr. levior* A, while *n*-C25 and C27-alkene were more common in B. The out-of-bag (OOB) estimate of error rate was low with 0.31%.

The 28 substances characterizing *Cr. levior* A had overall longer chains compared to the 17 substances from B (Wilcoxon rank sum test: W = 20, $p < 0.001$). The number of characteristic hydrocarbons per substance class did not differ between groups (χ^2-test: $\chi^2_7 = 11.96$, $p = 0.10$). However, when we included the relative abundance of substances, we found that *Cr. levior* A had higher proportions of alkadienes, alkenes, dimethyl alkanes and methyl-branched alkenes in the characterizing substances (Wilcoxon tests: all four $W > 20000$, $p < 0.001$). On the other hand, *Cr. levior* B possessed relatively more characteristic monomethyl alkanes ($W = 222$, $p < 0.001$) and *n*-alkanes ($W = 46$, $p < 0.001$).

For *Camponotus femoratus* PAT and PS (n = 195 and 107, respectively), the random forest algorithm classified all colonies correctly (OOB estimated error rate: 0%). For PS, 30 CHCs were characteristic, the most important ones being a C36-alkene (variable importance: 19.68), C34-alkene (19.59), C40-alkadiene (19.40), C33-alkene (19.27), 13-MeC37-alkene (18.86) and 13-MeC41-alkene (18.84). In contrast, a set of 9 CHCs sufficed to define *Ca. femoratus* PAT, with a C40-alkadiene (11.42) and cf. 9,33-C41-alkadiene (11.18) as most important ones (Fig. 2.2 B).

In contrast to *Cr. levior* A and B, the compounds that characterized *Ca. femoratus* PAT and PS (respectively) on average did not differ in chain lengths nor substance class membership (chain length: Wilcoxon test; $W = 163.5$, $p = 0.35$; CHC class: χ^2 test; $\chi^2_6 = 9.06$, $p = 0.17$).

However, when we took their relative abundances into account, *Ca. femoratus* PAT had higher proportions of dimethyl alkanes (Wilcoxon test: W = 20808, $p < 0.001$) and *n*-alkanes (W = 18220, $p < 0.001$) among the characteristic substances. In contrast, *Ca. femoratus* PS had higher proportions of alkadienes (W = 53, $p < 0.001$) and also possessed alkenes (W = 0, $p < 0.001$), monomethyl alkanes (W = 0, $p < 0.001$) and methyl-branched alkenes (W = 0, $p < 0.001$), all three of which lacked in the characterizing substances of *Ca. femoratus* PAT.

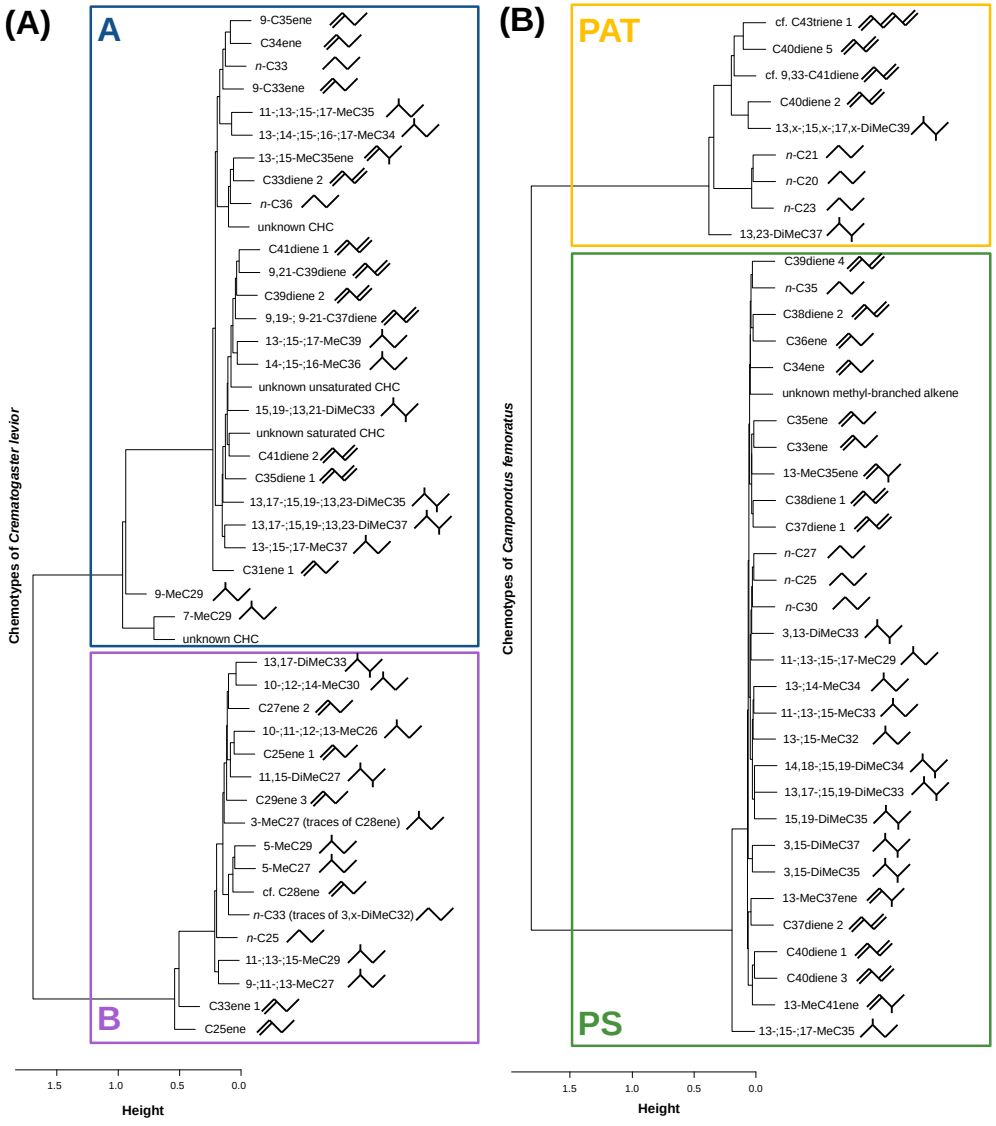


Figure 2.2: Hierarchical cluster analysis of substances characterizing the chemotypes of parabiotic ants. (**A**) Shows the different CHCs of the cryptic *Cr. levior* species with A in one cluster (top – blue) and B in one cluster (bottom – purple). (**B**) Shows the same for *Ca. femoratus* with PAT defined by fewer substances (top – yellow) compared to PS (bottom – green). Structures behind the substances indicate their substance class identities - note that these structures are just simplified symbols that do not reflect the actual structures (i.e. double bonds are most likely not conjugated and methyl groups usually do not occur on adjacent carbons in insect CHCs).

Comparison of *Crematogaster levior* With Other *Orthocrema* Species

Crematogaster levior A was clearly separated from all other species in both the NMDS ordination (Fig. 2.3A) and the cluster analysis (Fig. 2.3B). The closest other species was *Cr. brasiliensis*, most likely caused by the high chain lengths of both species. In contrast, *Cr. levior* B grouped more closely to the related non-parabiotic species and the third chemotype *Cr. levior* C and formed three groups rather than a single group within a bigger cluster (Fig. 2.3B). The assignment of the third chemotype *Cr. levior* C was also ambiguous as they did not group evenly in the cluster.

The ancestral state reconstruction for the average CHC chain length in the *Orthocrema* clade revealed that the ancestral chain length was 27.91 ± 2.82 carbon atoms (reconstructed ancestral state ± SD; Table S2.4). The weighted average chain length of *Cr. levior* A was 31.91 and thus above this range, while the one of *Cr. levior* B was 27.19 (Fig. 2.3C). This indicates that the chain length of *Cr. levior* A is a derived state, while for *Cr. levior* B, the trait does not differ from its ancestral state.

Comparison of *Camponotus femoratus* With Other *Camponotus* Species

In the ordination, colonies of *Ca. femoratus* PAT and PS were clearly separated and located most closely to other parabiotic species (Fig. 2.4 A). The cluster analysis resulted in two stable parabiotic and one big stable non-parabiotic clusters (Fig. 2.4 B; Jaccard coefficient > 0.75). Three clusters were not stable and in total contained only four species (one parabiotic and three non-parabiotic; Jaccard coefficient < 0.75), leaving one species-rich parabiotic, one smaller parabiotic and one species-rich non-parabiotic as stable clusters. *Camponotus femoratus* PAT and PS were grouped next to each other in the species-rich parabiotic cluster together with three other parabiotic species (Menzel *et al.* 2017a).

Signals of Parabiotic Partner

For each cryptic species separately, we searched in detail for substances which differed according to the identity of their parabiotic partner, based on pseudo-F values from multiple PERMANOVA analyses for each substance. Indeed, several of the identified CHCs were exclusively found or more abundant in the profiles of the respective mutualistic partners (Table 2.1).

In *Cr. levior* A, 9,19-; 9,21-C37-alkadiene was more abundant in colonies associated with *Ca. femoratus* PS (PERMANOVA: pseudo-F_1 = 15.57, p = 0.001). These substances were common in *Ca. femoratus* PS (mean abundance ± SD: 9.91 ± 3.65) but absent in *Ca. femoratus* PAT. In *Cr. levior* B, we only found substances more common when living together with PAT (i.e. no CHCs were more abundant in colonies living with PS). Here, 13,25-dimethyl C39

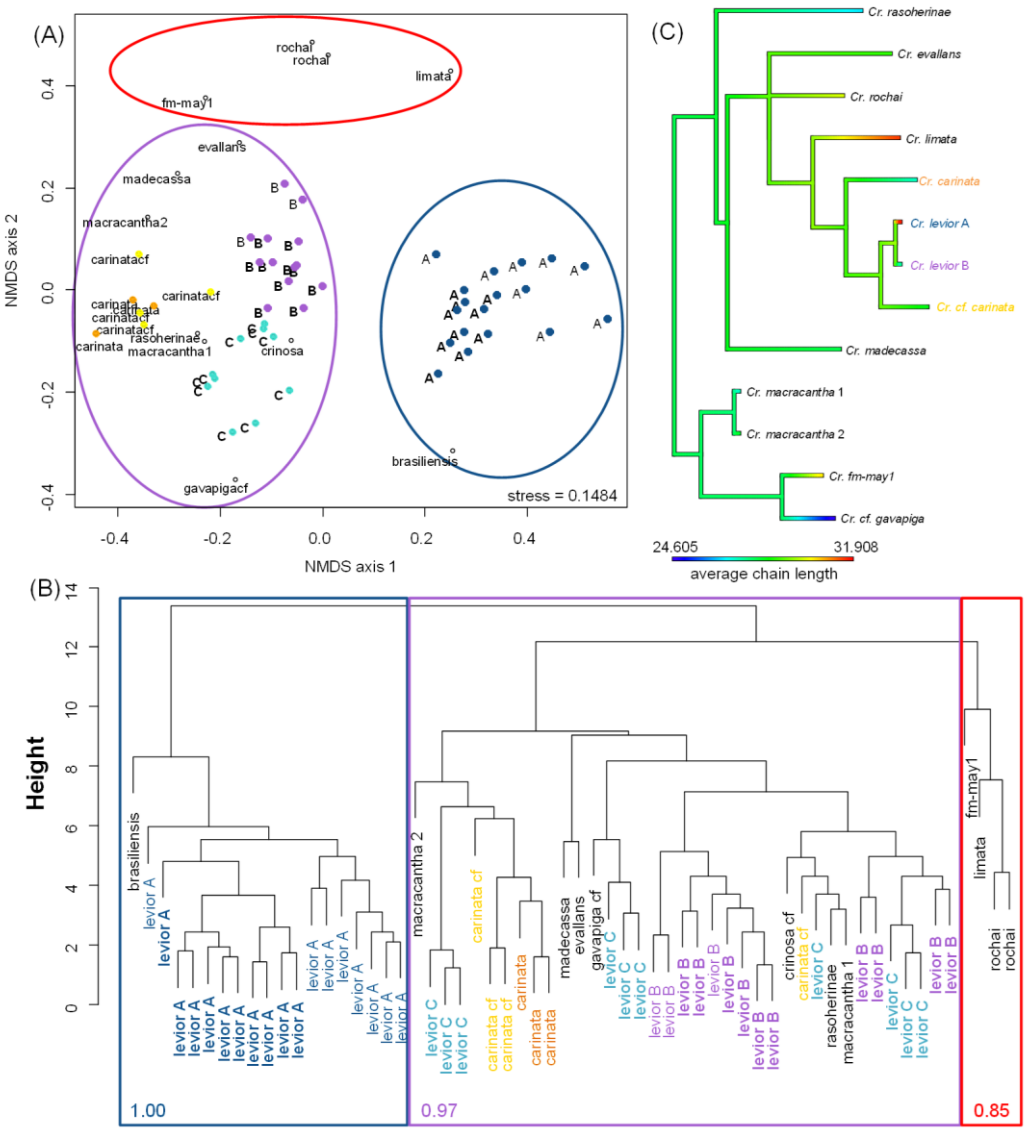


Figure 2.3: Ordination, cluster analysis and CHC chain length ancestral state of *Orthocrema* species. (**A**) Shows an NMDS ordination using Bray-Curtis dissimilarities of *Orthocrema* species based on the CHC data from (Menzel *et al.* 2017b) and ten colonies each of the three chemotypes we found (printed in bold). *Cr. levior* A clusters together with *Cr. brasiliensis* (blue circle), but is clearly separated from all other species, including *Cr. levior* B and C (purple circle). *Cr. limata* and *Cr. rochai* also form an own cluster (red circle). (**B**) Shows an average linkage hierarchical cluster analysis based on the same data. Numbers in the clusters show the mean Jaccard similarity between 100 bootstrap iterations. Values larger than 0.75 indicate that a cluster is stable. Again *Cr. levior* A as well as *Cr. limata* and *Cr. rochai* are clearly separated from all other *Orthocrema* species. Within the big group, *Cr. levior* B is in a single cluster, while *Cr. levior* C is distributed across two bigger branches. The *Cr. levior* A and B samples from this study group together with those of the previously published dataset in both analyses. (**C**) Shows an ancestral state reconstruction for the average CHC chain length based on the previously published phylogeny of the *Orthocrema* clade (Menzel *et al.* 2017b). The chain lengths of *Cr. levior* A and *Cr. limata* are strongly elongated compared to the ancestral state.

(pseudo-F_1 = 11.15, p = 0.003) was particularly interesting, because this substance was exclusively present in *Ca. femoratus* PAT (2.82 ± 1.08%).

For *Ca. femoratus* PAT, the substance which differed most depending on the parabiotic partner was a C40-alkadiene (pseudo-F_1 = 13.72, p = 0.001). However, this substance is neither present in *Cr. levior* A nor B. Under the aspect of signals of the mutualistic partner, *Ca. femoratus* PS was more interesting. The strongest effect was found in 11-;13-;15-monomethyl C29, which was more common in *Ca. femoratus* PS colonies associated with *Cr. levior* B (pseudo-F_1 = 19.02, p = 0.001), and which at the same time was one of the most abundant substances in this species compared to *Cr. levior* A (17.91 ± 7.73% vs. 0.36 ± 0.30%). In contrast, a C39-alkadiene was more common in *Camponotus* PS colonies living with *Cr. levior* A (pseudo-F_1 = 14.60, p = 0.001)

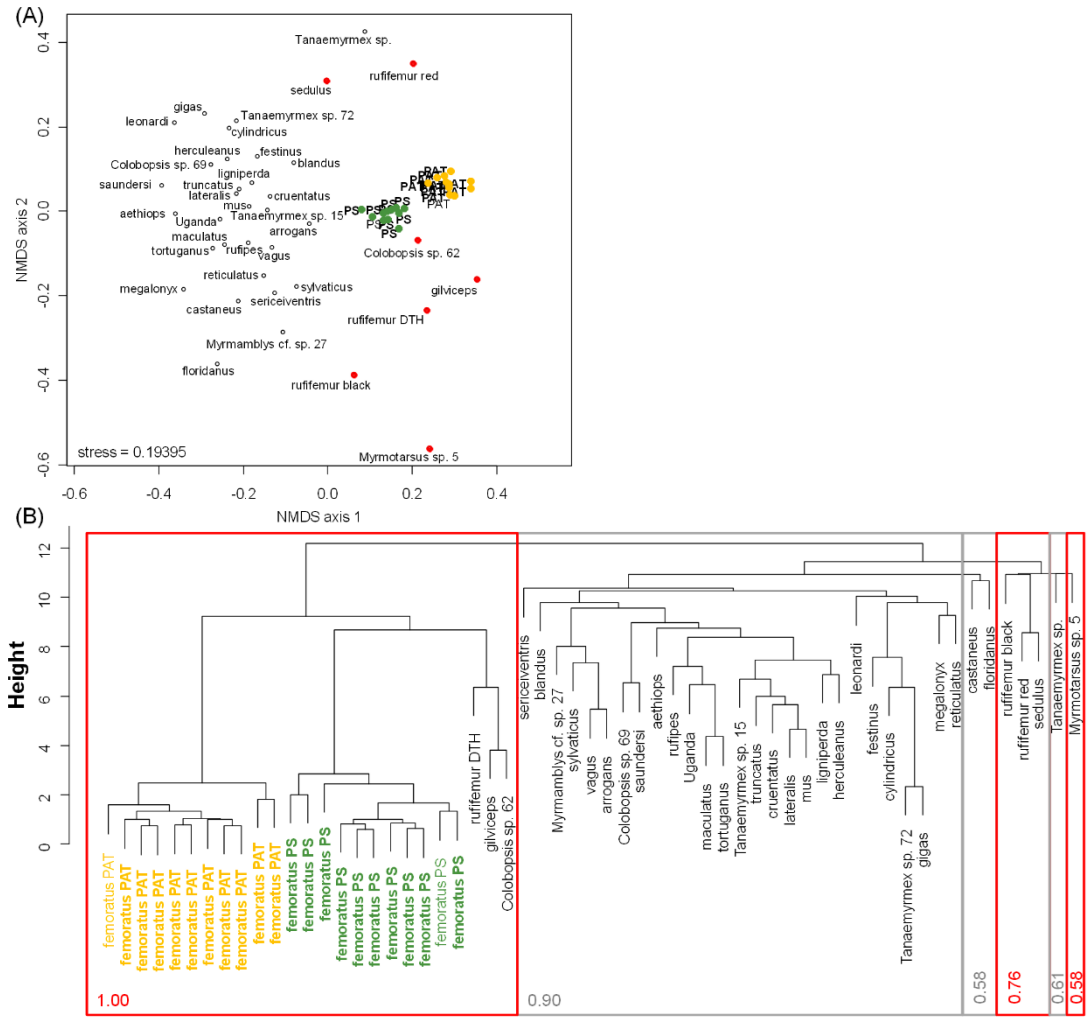


Figure 2.4: Ordination and cluster analysis of *Camponotus* species. (**A**) Shows an NMDS ordination using Bray-Curtis dissimilarities of *Camponotus* species based on the CHC data from (Menzel *et al.* 2017a) and each ten colonies of the *Ca. femoratus* species we found. *Ca. femoratus* PAT is depicted in orange, *Ca. femoratus* PS in green. Other parabiotic species are marked red. (**B**) Shows an average linkage hierarchical cluster analysis based on the same data. Numbers in the clusters show the mean Jaccard similarity between 100 bootstrap iterations. Values bigger than 0.75 indicate that a cluster is stable. Clusters with parabiotic species are framed in red, non-parabiotic ones in grey. New data points for *Ca. femoratus* PAT and PS reassemble the data from the previous study.

than in PS colonies living with *Cr. levior* B. C39-alkadienes were common in *Cr. levior* A (4.85 ± 5.13%) while lacking in B.

Effects of Species Identity, Partner and Climate on the Entire CHC Profile

As expected, the CHC profile differed the strongest between *Cr. levior* A and B (PERMANOVA based on Bray-Curtis dissimilarities: pseudo-F_1 = 461.39, p = 0.001). Besides, we also found significant effects of their parabiotic partner (PAT vs PS; pseudo-F_1 = 4.27, p = 0.013) and of the climate (pseudo-F_1 = 4.20, p = 0.018). Interactions between species identity, partner and climate indicate that the cryptic species responded differently to climate and partner species, which is why they were analysed separately below (interaction species:partner: pseudo-F_1 = 3.33, p = 0.030; interaction species:climatePC1: pseudo-F_1 = 4.85, p = 0.010; interaction partner:climatePC1: pseudo-F_1 = 3.23, p = 0.029).

Ca. femoratus PAT and PS profiles likewise differed most strongly (pseudo-F_1 = 1233.83, p = 0.001). There was a trend towards an influence of climate (pseudo-F_1 = 2.99, p = 0.067), but the *Cr. levior* partner had no impact on the CHC profile in its entirety (pseudo-F_1 = 1.64, p = 0.174). We also found effects of the climate on single CHCs in both genera (Supplement Table S2.5).

DISCUSSION

With our large-scale sampling, we confirmed that both previously described chemotypes of *Cr. levior* and *Ca. femoratus* were stable and present over a wide geographical range (Emery & Tsutsui 2013; Menzel *et al.* 2014; Hartke *et al*. 2019a). In the following, we will discuss the chemical diversity between the parabiotic cryptic species and how they differ from related non-parabiotic species, the influence of biotic interactions (parabiotic partner) and the impact of climate on their CHC profiles.

Chemical Diversity and Influence of Parabiotic Lifestyle

The parabiotic lifestyle was shown to promote certain characteristics of the CHC profile like higher chain length or higher abundances of alkadienes and methyl-branched alkenes (Menzel & Schmitt 2012b). This is probably the case because it requires a reduction of recognition cues that can be perceived by the mutualistic partner to facilitate interspecific tolerance. By increasing the chain length of hydrocarbons, their biophysical properties will change: Usually the melting point increases with chain length of structurally similar substance classes (Gibbs & Pomonis 1995). However, insects should need to maintain a sufficiently high fluidity in their CHC layer for functions like communication, foot adhesion or lubrication (Sprenger *et al*. 2018), which is why chain elongation in hydrocarbons is constrained. To keep a semi-liquid state of the CHC layer, the insertion

Table 1: CHCs influenced by the identity of the parabiotic partner. The table lists the CHCs that differ among colonies of a single species with the two respective parabiotic partners, including the corresponding pseudo-F and p values from univariate PERMANOVAs based on Euclidean distance. For each substance, proportions in the focal species according to partner identity are given (relative proportion of CHC (%) of species living with partner; with higher proportion in bold), and the proportions in the respective partner species (relative proportion of CHC (%) in partner species), with substances printed bold if they fit the trend in the focal species. Numbers behind alkenes or alkadienes indicate different substances eluting at different retention times.

species	substance	pseudo-F	p	% of *Cr. levior* profile living with partner PAT	% of *Cr. levior* profile living with partner PS	% in *Ca. femoratus* PAT	% in *Ca. femoratus* PS
***Crematogaster levior* A**	13,17-;15,19-DiMeC35	21.042	0.001	1.16 ± 0.46	**1.55 ± 0.61**	0 ± 0	**0.88 ± 0.65** [3]
	15-MeC37ene 1	19.711	0.001	**0.57 ± 0.68**	0.17 ± 0.23	0 ± 0	13.44 ± 4.01 [4]
	C39diene 3	19.253	0.001	**0.17 ± 0.12**	0.09 ± 0.10	8.05 ± 2.24 [1]	9.75 ± 2.37 [1]
	unknown CHC (RT 11.44)	18.780	0.001	**0.22 ± 0.17**	0.12 ± 0.12	0 ± 0	0 ± 0
	13-;14-;15-;16-MeC32	15.694	0.001	0.23 ± 0.07	**0.28 ± 0.07**	0 ± 0	**0.70 ± 0.31** [2]
	9,19-; 9,21-C37diene (1)	15.570	0.001	3.24 ± 2.75	**5.26 ± 3.77**	0 ± 0	**9.91 ± 3.65** [1]
	15-MeC37ene 2	14.150	0.001	0.18 ± 0.25	**0.36 ± 0.35**	0 ± 0	**13.44 ± 4.01** [4]
	C33diene 2	14.025	0.001	**7.43 ± 7.04**	3.45 ± 5.90	0 ± 0	0 ± 0
	unknown CHC (RT 13.48)	12.367	0.002	**0.13 ± 0.09**	0.09 ± 0.05	0 ± 0	0 ± 0
	unknown CHC (RT 11.10)	12.358	0.001	**0.20 ± 0.18**	0.11 ± 0.12	0 ± 0	0 ± 0
	unknown CHC (RT 12.10)	12.077	0.002	**0.27 ± 0.22**	0.16 ± 0.14	0 ± 0	0 ± 0
***Crematogaster levior* B**	13,17-;15,19-DiMeC37	30.728	0.001	**0.33 ± 0.20**	0.15 ± 0.10	0 ± 0	0.27 ± 0.25 [3]
	unknown CHC (RT 11.10)	16.337	0.001	**0.01 ± 0.01**	0.01 ± 0.01	0 ± 0	0 ± 0
	13,25-DiMeC39	11.148	0.003	**0.13 ± 0.09**	0.08 ± 0.06	**2.82 ± 1.08** [5]	0 ± 0
	cf. 7-MeC23	10.152	0.005	**0.10 ± 0.09**	0.05 ± 0.07	0 ± 0	0 ± 0
	unknown CHC (RT 11.72)	9.557	0.002	**0.16 ± 0.10**	0.10 ± 0.09	0 ± 0	0 ± 0
	unknown CHC (RT 11.44)	9.555	0.002	**0.15 ± 0.11**	0.09 ± 0.09	0 ± 0	0 ± 0
	cf. 5-MeC23	9.392	0.002	**0.25 ± 0.13**	0.17 ± 0.14	0 ± 0	0 ± 0
	unknown CHC (RT 11.92)	8.727	0.002	**0.20 ± 0.17**	0.12 ± 0.10	0 ± 0	0 ± 0
	unknown CHC (RT 12.10)	8.006	0.007	**0.36 ± 0.21**	0.25 ± 0.18	0 ± 0	0 ± 0
	unknown CHC (RT 13.48)	7.960	0.007	**0.20 ± 0.09**	0.16 ± 0.06	0 ± 0	0 ± 0
	9-;11-MeC23	7.802	0.004	**0.41 ± 0.19**	0.32 ± 0.14	0 ± 0	0 ± 0
	unknown CHC (RT 11.60)	7.785	0.006	**0.17 ± 0.10**	0.12 ± 0.09	0 ± 0	0 ± 0
	unknown CHC (RT 12.48)	6.900	0.009	**0.25 ± 0.13**	0.20 ± 0.12	0 ± 0	0 ± 0
species	**substance**	**pseudo-F**	**p**	**% of *Ca. femoratus* profile living with partner A**	**% of *Ca. femoratus* profile living with partner B**	**% in *Cr. levior* A**	**% in *Cr. levior* B**
***Camponotus femoratus* PAT**	C40diene 4	13.721	0.001	**1.16 ± 0.40**	0.96 ± 0.36	0 ± 0	0 ± 0
	13,x-;15,x-;17,x-DiMeC39	6.473	0.010	**3.01 ± 1.00**	2.62 ± 1.14	**0.36 ± 0.20** [6]	0.11 ± 0.09 [6]
	n-C23	4.392	0.038	0.95 ± 1.23	**1.35 ± 1.42**	0.30 ± 0.20	**0.89 ± 0.66**
	C41diene 4	4.406	0.045	**3.00 ± 1.06**	2.71 ± 0.89	**0.93 ± 1.14** [1]	0 ± 0
	C41diene 5	5.699	0.015	**0.65 ± 0.45**	0.52 ± 0.28	**0.93 ± 1.14** [1]	0 ± 0
***Camponotus femoratus* PS**	11-;13-;15-MeC29	19.018	0.001	0.12 ± 0.13	**0.33 ± 0.36**	0.36 ± 0.30	**17.91 ± 7.73**
	n-C25	16.592	0.001	0.11 ± 0.09	**0.22 ± 0.18**	0.26 ± 0.20	**11.20 ± 5.69**
	C39diene 1	14.603	0.001	**0.09 ± 0.14**	0 ± 0	**4.84 ± 5.12** [1]	0 ± 0
	15,19-DiMeC41	5.886	0.016	0.08 ± 0.11	**0.15 ± 0.21**	0 ± 0	0 ± 0
	C33ene	5.598	0.003	**0.79 ± 0.84**	0.48 ± 0.27	**14.25 ± 4.13** [1]	1.14 ± 1.50 [1]
	unknown CHC (RT 30.40)	5.464	0.019	0.02 ± 0.04	**0.05 ± 0.07**	0 ± 0	0 ± 0
	C39diene 2	5.364	0.031	**6.17 ± 2.12**	5.24 ± 1.93	**4.84 ± 5.12** [1]	0 ± 0
	C37diene 1	5.302	0.011	**0.62 ± 0.35**	0.48 ± 0.22	**5.53 ± 4.08** [1]	0 ± 0
	unknown CHC (RT 30.60)	5.296	0.020	**0.12 ± 0.11**	0.08 ± 0.08	0 ± 0	0 ± 0
	C38diene 1	4.994	0.021	**1.01 ± 0.28**	0.89 ± 0.28	0 ± 0	0 ± 0

[1] combined percentages of all alkenes or alkadienes; [2] only 13-;15-MeC32; [3] only 15,19-DiMe; [4] 13-MeC37ene – different methyl-group position; [5] mixed with other dimethyl-alkanes; [6] 13,x-;13,25-DiMeC39

of disrupting features, either methyl-branches or double bonds, that lower the melting temperature are required (Menzel *et al.* 2017a). At ambient temperatures, long-chained hydrocarbons with such disrupting features should be similarly liquid but less volatile (and thus less perceivable) than shorter-chained ones as the difference between melting and boiling point increases with chain length (Li *et al.* 2006).

Although both *Cr. levior* species live in parabiosis, only the CHC profiles of *Cr. levior* A showed elongated carbon chains and high abundances of alkadienes. In contrast, *Cr. levior* B had CHCs of shorter chain length, which require less disrupting features (here, a high abundance of mono-methyl alkanes; Fig. 2.2A). In our comparison with other species of the *Orthocrema* clade *Cr. levior* A was chemically most distant from all others (with only *Cr. brasiliensis* in the same cluster; Fig. 2.3B). This separation is probably due to longer CHC chains in *Cr. levior* A. *Cr. levior* B on the other hand was more similar to most other *Orthocrema* species, and their chain lengths were similar to the ancestral state.

The lack of characteristic adaptations of the CHC profile in *Cr. levior* B despite living in parabiosis is surprising. It calls for additional studies to investigate whether *Cr. levior* A and B differ in how much they are tolerated by *Camponotus femoratus*. For example, due to their shorter chain lengths, *Cr. levior* B might be attacked more frequently. Alternatively, there might be other mechanisms than modification of the CHC profile to achieve interspecific tolerance. For example, beside CHCs, *Cr. levior* A and B also possess species-specific polar substances on their cuticles (Hartke *et al.* 2019a). Their function is currently unknown, but they might function as appeasement allomones, similarly to other parabiotic *Crematogaster* species in South-East Asia (Menzel *et al*. 2013). Additional behavioural tests using fractions of these polar compounds could give insights if they indeed function as appeasement allomones.

Interestingly, we found an additional chemotype in *Crematogaster* (*Cr. levior* C) that greater CHC similarity to *Cr. levior* B compared to A (see Fig. 2.1A, 2.3A). Morphological examination revealed that the individuals had carinae (keel-like crests on the pronotum) comparable to *Cr. carinata*, which is otherwise morphologically very similar to *Cr. levior*. Both were only quite recently divided into separate species (Longino 2003). Nevertheless, genetically, the *Cr. levior* C colonies were different from *Cr. carinata* and also from *Cr. levior* A and B (concerning the COI locus; data not shown). They either lived in parabiosis with *Odontomachus mayi* or were found trail-sharing with unidentified *Camponotus* ants. Similarly, *Cr. carinata* is also facultatively parabiotic with different *Dolichoderus* and *Odontomachus* species (Longino 2003). This finding suggests that there might be an even higher chemical diversity within the *Cr. levior/Cr. carinata* complex, with possibly more cryptic species to be discovered.

In *Camponotus*, all parabiotic species in our comparison differed from non-parabiotic congeners (Fig. 2.4). The cryptic species *Ca. femoratus* PAT and PS possessed profiles that differed substantially from each other. *Ca. femoratus* PS had higher proportions of alkadienes and methyl-branched alkenes, which are generally more common in parabiotic species (Martin & Drijfhout 2009; Menzel & Schmitt 2012b; Menzel *et al*. 2017a). The mean chain length, however, was increased in both species as well as other parabiotic species, which could again indicate adaptations to the mutualism. Nevertheless, *Ca. femoratus* PAT and PS seem to use slightly different strategies to allow this chain elongation, i.e. to introduce disruptive features into the CHC chains: While both have characteristic alkadienes, *Ca. femoratus* PAT has high proportions of di-methyl alkanes, whereas *Ca. femoratus* PS produces methyl-branched alkenes, a quite rare substance group in ants (Martin & Drijfhout 2009) (Fig. 2.2B). The compounds with a high chain length thus differ in their chemical structure between the two cryptic species.

Biotic Interactions Influence CHC Profiles

Although the two partner species had consistently different CHC profiles, we detected some signals of major compounds of one species in the respective parabiotic partner. In several cases, we found a major compound in one partner to be more abundant in those colonies of the other species that lived with this partner. However, only few CHCs showed this signal, indicating that there is no simple substance transfer between parabiotic partners. Rather, there might be either selective transfer of few compounds (via trophallaxis or via nest material), or partial chemical mimicry where one species produces higher quantities of a certain CHC only if it is associated to one of the two cryptic species. Although the relative abundances of these substances could still be high in one partner while they were only traces in the other partner, in both cases these shared CHCs might promote the interspecific tolerance of the parabiotic ants and make it easier to familiarize with the profile of the associated species (Orivel *et al.* 1997; Lenoir *et al.* 2001).

Based on our data, we cannot conclusively distinguish between chemical mimicry (i.e. active production of similar CHCs) and chemical camouflage (i.e. active or passive acquisition of the partner CHCs). However, not only the most common substances were exchanged, which would be the most likely scenario if the mechanism was chemical camouflage. Only a subset of certain substances could be found, and they were not always present in the parabiotic partner, which tentatively hints towards an active production of CHCs according to the associated partner. Interestingly, we found that *Cr. levior* B has higher proportions of 13,25-DiMeC39, a substance only common in *Ca. femoratus* PAT if associated with them. On the other hand, *Cr. levior* A has higher proportions of C41-dienes if they live with *Ca. femoratus* PAT colonies. This indicates that in *Cr. levior* B, itself richer in monomethyl alkanes, profiles change in abundance of a dimethyl alkane, while in the

alkadienes-rich *Cr. levior* A alkadienes change most. In *Camponotus*, we found almost no signal of the *Cr. levior* partner in PAT colonies, while *Ca. femoratus* PS showed signals of very prominent CHCs of their respective parabiotic partners (with A: C39-dienes; with B: 11-;13-;15-methyl nonacosane; Table 2.1). According to these trends, it seems that *Crematogaster* needs to have changed profiles to co-habit with *Ca. femoratus* PAT (which itself remains unchanged), while *Ca. femoratus* PS shows signals of its respective *Crematogaster* partner.

In *Cr. levior* A one of the CHCs differing depending on the parabiotic partner was 15-MeC37ene. *Ca. femoratus* PAT however produced 13-MeC37ene, a similar substance with a different methyl-group position. Although we do not know the position of the double bond in both cases, this could indicate an incomplete chemical mimicry by *Cr. levior* A. Such imperfect mimicry with CHC profiles that do not entirely fit the host profile can be found in two parasitoid wasp species and the ant *Ectatomma parasiticum* that both parasitize the ant *E. tuberculatum* (Savarit & Fénéron 2014; Pérez-Lachaud *et al*. 2015).

The chemical interactions in this parabiotic association are likely to be even more complex since both species also interact with the ant garden plants. *Camponotus femoratus* was found to actively collect seeds of specific ant garden plants and recognize them based on chemical signals (Youngsteadt *et al*. 2008, 2009). However, the seeds of different plants did not share common signals eliciting this carrying behaviour (Youngsteadt *et al*. 2010). Furthermore, at least *Ca. femoratus* recruits to damaged leaves of the ant garden plant, indicating that they actively defend the plants against herbivory (Vantaux *et al.* 2007; Vicente *et al*. 2014; Leal *et al.* 2017). Further studies should investigate if the chemically different cryptic species described here differ in their reactions to certain plant odours or show differences in recruiting to leaf damage. Such experiments would give insight if certain plants coevolve with one or the other cryptic species of *Ca. femoratus* or *Cr. levior*.

Climate Influences on CHC Profiles

We found that the climate affected the whole CHC profile of both species of *Cr. levior*. Climate adaptations were found in *Drosophila melanogaster*, populations along the east coast of the United States (Rajpurohit *et al*. 2017) but also in several ant species (Menzel *et al.* 2017a). Furthermore, ants are able to acclimate their CHC profile on a short-term basis e.g. to higher temperatures (Menzel *et al.* 2018; Sprenger *et al.* 2018) which can lead to differences in populations between seasons (Buellesbach *et al.* 2018). In *Ca. femoratus*, we only found a non-significant trend when also accounting for species identity and partner in the PERMANOVA. Probably, this is because the distribution of the cryptic *Ca. femoratus* species was associated with the climatic gradient in French Guiana. *Ca. femoratus* PS was very rare in eastern French Guiana, where precipitation is higher and temperature slightly

lower (Hartke *et al.* 2019a; Fig. S2.1), while both PS and PAT occur in the drier west. This difference in distribution could probably hint at niche partitioning according to different climate adaptations.

CONCLUSIONS

In this study, we demonstrate that the cryptic species of *Cr. levior* and *Ca. femoratus* strongly differ in their CHC profiles as well as in the proportions of certain substance classes. The parabiotic lifestyle led to hydrocarbon backbone elongations in *Cr. levior* A, *Ca. femoratus* PAT and *Ca. femoratus* PS. These results suggest that differences in certain substance classes might represent different strategies to allow CHC chain elongation towards the very high chain lengths present in parabiotic ants that promotes the highly unusual interspecific tolerance. In contrast, *Cr. levior* B has a CHC profile comparable to other non-parabiotic *Orthocrema* clade-species and thus is closer to the ancestral state concerning the average CHC chain length. It remains to be studied whether this CHC differences results in differences in interspecific tolerance that *Cr. levior* A and B experience from their parabiotic partner.

Importantly, we also demonstrate that the CHC profiles of parabiotic ants are influenced by their mutualistic partners as well as by the climate. Depending on the association, one or the other partner seems to adjust its CHC profile either actively or passively, which could facilitate the co-habituation. Although we cannot determine the mechanism behind the shared CHCs among the parabiotic partners (i.e. mimicry or substance transfer) yet, the interspecific tolerance in this parabiotic system offers further chances to investigate the establishment of mutualistic interactions and the chemical mechanisms facilitating them.

ACKNOWLEDGEMENTS

Firstly, we would like to thank Heike Stypa for their support in the lab. Further on, we want to acknowledge Jérôme Chave, Philippe Gaucher and Dorothée Deslignes for permission to sample at the Nouragues Ecological Research Station, Aurélie Dourdain for allowing us working at the Paracou Research Station and similarly, the late Philippe Cerdan for the Hydreco Lab Petit Saut. Next, we thank Bonnie Blaimer and Manfred Verhaagh for their species identification. Finally, we want to thank Marina Psalti and two anonymous reviewers for useful comments on an earlier version of this manuscript.

SUPPLEMENTARY MATERIAL

Effects of Climate on single CHCs

The CHC profiles of all four species were affected by climate, however with different consequences. In *Cr. levior* A climate affected the CHC profile (multivariate PERMANOVA based on Bray-Curtis dissimilarities: pseudo-F_1 = 6.13, p = 0.001) via an increase of three alkadienes and two methyl-branched alkenes at higher precipitation and lower temperature. On the other hand, one *n*-alkane, two monomethyl alkanes, one dimethyl alkane, one alkene and one alkadiene with higher chain length were more abundant at lower precipitation and higher temperatures (Table S2.5). Climate also affected the CHC profile of *Cr. levior* B (pseudo-F_1 = 5.01, p = 0.001). The relative proportion of two alkenes increased with higher precipitation and lower temperatures while two alkadienes decreased with increasing rainfall and cooler temperatures (Table S2.5).

In *Camponotus*, the CHCs of *Ca. femoratus* PAT were significantly influenced by climatic conditions (pseudo-F_1 = 3.21, p = 0.022). Here, one alkene, four alkadienes and one methyl-branched alkene were affected most strongly, and all these hydrocarbons increased at higher temperature with less precipitation while two dimethyl alkanes decreased (Table S2.5). The effect of climate was even stronger in *Ca. femoratus* PS (pseudo-F_1 = 4.86, p = 0.001). One alkadiene increased with higher precipitation and lower temperature, whereas four methyl-branched alkenes decreased along this gradient (Table S2.5).

Table S2.1: Sampling locations. Note that the coordinates reflect the actual sampling locations, not the cities the locations are named after.

Location	Latitude	Longitude	Elevation (m)	# *Crematogaster*	# *Camponotus*
Apatou (AP)	5.200783	-54.312017	28	16	16
Saint-Laurent-du-Maroni (SL)	5.463902	-53.997322	63	37	36
Angoulême (AN)	5.409200	-53.650933	64	1	1
Sinnamary (SI)	5.352035	-53.077604	45	19	20
Petit Saut (PS)	5.061213	-52.988772	93	22	18
Paracou (PAR)	5.265905	-52.933605	41	51	49
Les Nouragues (LN)	4.039650	-52.673933	63	73	61
Kourou (KO)	5.083106	-52.643022	23	12	11
Montsinéry-Tonnegrande (MT)	4.866000	-52.538483	26	4	4
Cacao (CA)	4.557416	-52.463067	71	23	20
Cayenne (CAY)	4.793831	-52.317594	20	6	5
Régina (RE)	4.181286	-52.131963	82	16	14
Camp Patawa (PAT)	4.546067	-52.130483	282	52	51
Total				**332**	**306**

Table S2.2: Cuticular hydrocarbons of the cryptic *Crematogaster levior* species. In this table we present the identified CHCs, their mean abundances and the retention indices (RI) diagnostic ions for identification. Numbers in italics are the M^+ ions; diagnostic ions in brackets were obtained after DMDS derivatization of unsaturated hydrocarbons. Note that, for unsaturated hydrocarbons, position of double bonds could only be determined for *Cr. levior* A.

CHC	RI	mean ± SD in *Cr. levior* A	mean ± SD in *Cr. levior* B	mean ± SD in *Cr. levior* C	diagnostic ions and M^+
C20ene	20.00	2.36 ± 1.74	1.69 ± 0.97	2.17 ± 1.51	*280*
n-C20	20.06	4.72 ± 3.2	3.49 ± 1.86	3.53 ± 1.09	*282*
unknown CHC (RT 8.84)	20.14	0.57 ± 0.46	0.39 ± 0.23	0.48 ± 0.25	
C22ene	21.99	1.14 ± 0.77	0.82 ± 0.41	0.72 ± 0.21	*308*
n-C22	22.05	2.47 ± 1.6	1.84 ± 0.83	1.77 ± 0.55	*310*
unknown CHC (RT 11.10)	22.12	0.63 ± 0.45	0.44 ± 0.23	0.47 ± 0.17	
unknown CHC (RT 11.44)	22.38	0.18 ± 0.19	0.11 ± 0.08	0.11 ± 0.08	
unknown CHC (RT 11.48)	22.41	0.18 ± 0.16	0.13 ± 0.11	0.14 ± 0.07	
unknown CHC (RT 11.60)	22.50	0.37 ± 0.26	0.25 ± 0.14	0.26 ± 0.1	
unknown CHC (RT 11.72)	22.59	0.24 ± 0.17	0.16 ± 0.1	0.17 ± 0.07	
unknown CHC (RT 11.94)	22.76	0.25 ± 0.27	0.15 ± 0.1	0.1 ± 0.12	
unknown CHC (RT 12.10)	22.88	0.24 ± 0.21	0.17 ± 0.15	0.22 ± 0.09	
n-C23	23.05	0.3 ± 0.2	0.88 ± 0.66	3.28 ± 3.84	*324*
unknown CHC (RT 12.48)	23.15	0.4 ± 0.35	0.33 ± 0.21	0.33 ± 0.16	
unknown CHC (RT 12.74)	23.33	0.36 ± 0.24	0.24 ± 0.13	0.23 ± 0.12	
cf. 9-;11-MeC23	23.41	0.42 ± 0.28	0.39 ± 0.18	0.44 ± 0.17	140/224; 168/196
cf. 7-MeC23	23.47	0.12 ± 0.16	0.09 ± 0.09	0.14 ± 0.11	112/252
cf. 5-MeC23	23.54	0.35 ± 0.34	0.22 ± 0.14	0.24 ± 0.12	85/281
3-MeC23	23.77	0.08 ± 0.09	0.07 ± 0.07	0.15 ± 0.11	57/309
unknown CHC (RT 13.48)	23.84	0.13 ± 0.09	0.09 ± 0.05	0.08 ± 0.05	
C24ene	23.99	0.43 ± 0.29	0.31 ± 0.15	0.32 ± 0.09	*336*
C25ene 1	24.79	0 ± 0	0.03 ± 0.26	0.12 ± 0.11	*350*
C25ene 2	24.88	0 ± 0	0.12 ± 0.1	2.04 ± 2.65	*350*
C25ene 3	25.04	0 ± 0	0.08 ± 0.27	0.25 ± 0.29	*350*
n-C25	25.05	0.26 ± 0.2	11.2 ± 5.69	10.08 ± 7.58	*352*
unknown CHC (RT 15.64)	25.27	0.2 ± 0.13	0.19 ± 0.08	0.15 ± 0.04	
cf. 9-;11-;13-MeC25	25.41	0.35 ± 0.31	0.68 ± 0.36	0.86 ± 0.5	140/252; 168/224; 197/197
7-MeC25	25.46	0 ± 0	0 ± 0	0.34 ± 0.43	112/280
unknown CHC (RT 16.22)	25.64	0.25 ± 0.14	0.17 ± 0.09	0.19 ± 0.06	
C26ene	26.01	0.18 ± 0.11	0.12 ± 0.07	0.12 ± 0.08	*364*
n-C26	26.06	0.66 ± 0.3	1.14 ± 0.36	0.92 ± 0.28	366
unknown CHC (RT 16.98)	26.12	0.12 ± 0.09	0 ± 0	0 ± 0	
8-;9-;10-MeC26	26.37	0 ± 0	0 ± 0	0.25 ± 0.22	126/280; 140/266; 154/252
11-;12-;13-MeC26	26.40	0 ± 0	0.14 ± 0.05	0 ± 0	168/238; 182/224; 196/210
C27diene 1	26.58	0 ± 0	0 ± 0	0.06 ± 0.09	*376*
C27diene 2	26.62	0 ± 0	0 ± 0	0.07 ± 0.06	*376*
C27diene 3	26.75	0 ± 0	0 ± 0	0.12 ± 0.15	*376*
C27ene 1	26.78	0 ± 0	0.29 ± 0.58	10.17 ± 9.65	*378*
C27ene 2	26.81	0 ± 0	0.81 ± 1.64	4.71 ± 8.39	*378*
C27ene 3	27.02	0 ± 0	0.25 ± 0.83	0.95 ± 0.98	*378*
n-C27	27.06	1.69 ± 1.32	5.17 ± 2.01	11.43 ± 8.84	*380*
9-;11-;13-MeC27	27.39	0.25 ± 0.11	2.35 ± 1.57	0.64 ± 0.79	140/280; 168/252; 196/224
unknown CHC (RT 19.08)	27.44	0.17 ± 0.16	0.18 ± 0.37	4.85 ± 5.83	
7-MeC27	27.51	0 ± 0	0 ± 0	1.83 ± 2.72	112/308
5-MeC27	27.58	0 ± 0	0.17 ± 0.24	0.12 ± 0.07	85/336
11,15-DiMeC27	27.69	0 ± 0	0.12 ± 0.06	0.22 ± 0.21	168/266, 238/196
cf. C28ene	27.75	0 ± 0.01	0.19 ± 0.16	0.54 ± 0.61	*392*
3-MeC27 (with traces of C28ene)	27.80	0 ± 0	0.27 ± 0.17	0.02 ± 0.05	57/365
n-C28	28.07	0.58 ± 0.21	0.92 ± 0.22	0.52 ± 0.23	*394*
unknown CHC (RT 20.70)	28.46	0 ± 0	0 ± 0	0.26 ± 0.31	
C29diene 1	28.66	0 ± 0	0 ± 0	0.93 ± 0.9	*404*
C29ene 1	28.72	0 ± 0	0 ± 0	0.9 ± 1.61	*406*
C29ene 2	28.75	0 ± 0	0 ± 0	0.85 ± 1.08	*406*
C29diene2	28.81	0.02 ± 0.09	0.25 ± 1.5	0.63 ± 1.38	*404*
C29ene 3	28.82	0.11 ± 0.13	4.12 ± 5.75	7.55 ± 6.5	*406*
C29diene 3	28.89	0 ± 0	3.28 ± 4.06	0 ± 0	*404*

C29ene 4	28.9	0 ± 0	2.76 ± 4.06	3.73 ± 5.14	*406*
C29ene 5	28.92	0 ± 0	0.46 ± 0.32	0.05 ± 0.1	*406*
C29ene 6	29.04	0.06 ± 0.26	0 ± 0	0.28 ± 0.58	*406*
n-C29	29.08	5.46 ± 2.23	5.35 ± 1.85	3.19 ± 1.47	*408*
11-; 13-;15-MeC29 (11-MeC29 only in *Cr. levior* B)	29.38	0.36 ± 0.3	17.97 ± 7.73	1.72 ± 1.85	168/280; 196/252; 224/224
9-MeC29	29.43	0.01 ± 0.03	0.12 ± 0.51	2.25 ± 1.92	140/308
7-MeC29	29.47	0.06 ± 0.04	0.49 ± 1.01	1 ± 0.96	112/336
5-MeC29	29.57	0 ± 0	0.95 ± 1.26	0.08 ± 0.04	85/365
13,17-DiMeC29 (traces of 11,x)	29.65	0 ± 0	0.62 ± 0.77	0.85 ± 0.75	196/266, 266/196
11,x-DiMeC29	29.68	0 ± 0	0.21 ± 0.46	0 ± 0	168/294, x
C30diene (with traces of 3-MeC29)	29.86	0 ± 0	0.31 ± 0.26	0.01 ± 0.03	*418*
C30ene	29.88	0 ± 0	0.16 ± 0.26	0.3 ± 0.38	*420*
3-MeC29	29.89	0 ± 0	0 ± 0	0.33 ± 0.43	57/393
n-C30	30.06	0.4 ± 0.1	0.45 ± 0.16	0.14 ± 0.04	*422*
10-;12-;14-MeC30	30.38	0 ± 0	0.9 ± 0.54	0 ± 0	154/308; 182/280; 210/252
cf. 12,16-;13,17-DiMeC30	30.64	0 ± 0	0.05 ± 0.09	0 ± 0	182/294, 252/224; 196/280, 266/210
C31triene	30.66	0.12 ± 0.78	0.05 ± 0.27	0 ± 0	*430*
C31diene 1	30.68	0 ± 0	0.01 ± 0.07	2 ± 2.18	*432*
C31ene 1	30.69	0.12 ± 0.16	0 ± 0	0.12 ± 0.2	*434*
C31diene 2	30.80	0 ± 0	0.6 ± 1.61	0.86 ± 1.01	*432*
C31ene 2	30.81	0 ± 0	5.29 ± 6.04	2.52 ± 3.14	*434*
C31diene 3	30.85	0.27 ± 1.5	1.54 ± 2.07	0 ± 0	*432*
C31ene 3 (9-C31ene)	30.86	2.53 ± 1.32	1.71 ± 2.69	0.54 ± 0.66	*434* [173/355]
C31diene 4	30.91	0.07 ± 0.86	0.02 ± 0.14	0 ± 0	*432*
C31ene 4	30.93	0.09 ± 0.11	0.18 ± 0.28	0.09 ± 0.14	*434*
n-C31	31.07	2.01 ± 0.63	0.9 ± 0.4	0.43 ± 0.39	*436*
cf. 5-MeC31ene	31.29	0 ± 0	0.12 ± 0.2	0 ± 0	83/390, 97/406, *448*
11-;13-;15-MeC31 (11-MeC31 only in *Cr. levior* B and C)	31.37	1.11 ± 0.48	6.59 ± 4.32	0.54 ± 0.54	168/308; 196/280; 224/252
9-MeC31	31.48	0 ± 0.01	0.18 ± 0.45	0.2 ± 0.25	140/336
7-MeC31	31.50	0 ± 0	0.13 ± 0.25	0.26 ± 0.3	112/364
5-MeC31; 13,17-DiMeC31	31.58	0.27 ± 0.11	1.3 ± 0.85	0 ± 0	85/393; 196/294, 266/224
unknown CHC (RT 25.94)	31.87	0.44 ± 0.12	0.16 ± 0.12	0 ± 0	
n-C32	32.07	0.26 ± 0.12	0.17 ± 0.09	0 ± 0	*450*
13-;14-;15-;16-MeC32	32.37	0.25 ± 0.07	0.2 ± 0.13	0 ± 0	196/294; 210/280; 224/266; 238/252
C33diene 1	32.53	0.23 ± 0.26	0 ± 0	0 ± 0	*460*
12-,16-DiMeC32	32.61	0 ± 0	0.11 ± 0.17	0 ± 0	182/322, 252/252
C33diene 2	32.63	5.69 ± 6.86	0 ± 0	0 ± 0	*460*
C33triene 1	32.64	0 ± 0	0.01 ± 0.06	0 ± 0	*458*
C33diene 3	32.71	0.47 ± 0.54	0.18 ± 0.62	0.03 ± 0.08	*460*
C33ene 1	32.74	0.16 ± 0.3	1.04 ± 1.45	0.05 ± 0.1	*462*
C33triene 2	32.79	0.04 ± 0.56	0 ± 0	0 ± 0	*458*
C33diene 4	32.81	1.69 ± 2.61	0.06 ± 0.63	0 ± 0	*460*
C33ene 2	32.82	0 ± 0	0.07 ± 0.22	0.07 ± 0.19	*462*
C33diene 5	32.87	0.06 ± 0.78	0.08 ± 0.1	0 ± 0	*460*
C33ene 3 (9-C33ene)	32.89	14.09 ± 4.08	0.03 ± 0.13	0 ± 0	*462* [173/383]
n-C33 (traces of 3,x-DiMeC32)	33.05	0 ± 0	0.16 ± 0.09	0 ± 0	*464*
n-C33	33.07	1.08 ± 0.4	0 ± 0	0 ± 0	*464*
15-MeC33ene	33.14	0.35 ± 0.38	0 ± 0	0 ± 0	223/278, 238/294, *476*
9-;11-;13-;15-;17-MeC33	33.37	5 ± 1.62	0.99 ± 0.97	0.15 ± 0.46	140/364; 168/336; 196/308; 224/280; 252/252
C34diene 1	33.54	0.03 ± 0.06	0 ± 0	0 ± 0	*474*
13,17-DiMeC33	33.6	0 ± 0	1.43 ± 1.95	0 ± 0	196/322, 266/252
15,19-; 13,21-DiMeC33	33.61	1.18 ± 0.5	0.02 ± 0.21	0 ± 0	224/294, 294/224;

					196/322, 322/196
C34diene 2	33.88	0 ± 0.04	0 ± 0	0 ± 0	*474*
C34ene	33.89	0.28 ± 0.09	0 ± 0	0 ± 0	*476*
n-C34	34.08	0.2 ± 0.07	0.12 ± 0.07	0.05 ± 0.02	*478*
13-;14-;15-;16-;17-MeC34	34.37	0.32 ± 0.11	0 ± 0	0 ± 0	196/322; 210/308; 224/294; 238/280; 252/266
C35diene 1	34.58	3.22 ± 2	0 ± 0	0.01 ± 0.03	*488*
C35diene 2	34.67	0.87 ± 0.81	0 ± 0	0 ± 0	*488*
C35triene	34.73	0.01 ± 0.14	0 ± 0	0 ± 0	*486*
C35diene 3 (cf. 9,19-;9,21-C35diene) [1]	34.74	3.65 ± 4.46	0 ± 0	0 ± 0	*488* [173/409, 271/311; 243/339]
C35ene 1 (14-;16-;17-C35ene)[2]	34.76	1.02 ± 1.18	0.24 ± 0.53	0.02 ± 0.05	*490* [243/341; 271/313; 285/299]
C35diene 4	34.87	0.02 ± 0.16	0 ± 0	0 ± 0	*488*
C35ene 2 (9-C35ene)	34.88	2.53 ± 0.82	0 ± 0	0 ± 0	*490* [173/411]
13-;15-MeC35ene 1	35.07	0.61 ± 0.56	0 ± 0	0 ± 0	223/306, 238/322, *504*
13-;15-MeC35ene 2	35.16	0.09 ± 0.1	0 ± 0	0 ± 0	223/306, 238/322, *504*
13-;15-MeC35ene 3	35.21	0.45 ± 0.82	0 ± 0	0 ± 0	223/306, 238/322, *504*
11-;13-;15-;17-MeC35	35.37	3.78 ± 1.32	0.25 ± 0.24	0.03 ± 0.09	168/364; 197/336; 224/308; 252/280
cf. 13,17-;15,19-; 13,23-DiMeC35	35.59	1.33 ± 0.57	0.22 ± 0.21	0 ± 0	196/350, 266/280; 224/322, 294/252; 196/350, 350/196
n-C36	36.08	0.14 ± 0.08	0 ± 0	0 ± 0	*506*
14-;15-;16-MeC36	36.37	0.19 ± 0.11	0 ± 0	0 ± 0	210/336; 224/322; 238/308
C37diene 1 (cf. 9,19-; 9,21-C37diene)	36.60	4.09 ± 3.43	0 ± 0	0 ± 0	*516* [173/437, 299/311;271/339]
C37diene 2	36.73	1.38 ± 2.02	0 ± 0	0 ± 0	*516*
C37ene 1 (16-;18-C37ene)	36.75	0.83 ± 1.15	0 ± 0	0 ± 0	*518* [271/341; 299;313]
C37diene 3	36.90	0.06 ± 0.14	0 ± 0	0 ± 0	*516*
C37ene 2	36.91	0.27 ± 0.22	0 ± 0	0 ± 0	*518*
15-MeC37ene 1	37.07	0.4 ± 0.56	0 ± 0	0 ± 0	223/334, 238/350, *532*
15-MeC37ene 2	37.16	0.24 ± 0.3	0 ± 0	0 ± 0	223/334, 238/350, *532*
13-;15-;17-MeC37	37.36	1.26 ± 0.82	0.18 ± 0.25	0 ± 0	196/364; 224/336; 252/308
13,17-;15,19-; 13,23-DiMeC37	37.58	1.21 ± 0.64	0.27 ± 0.19	0 ± 0	196/378, 266/308; 224/350, 294/280; 196/378/350/224
C39diene 1 (cf. 9,21-C39diene)	38.58	4.12 ± 4.55	0 ± 0	0 ± 0	*544* [173/465, 299/339]
C39diene 2	38.72	0.58 ± 0.65	0 ± 0	0 ± 0	*544*
C39ene 1	38.73	0.01 ± 0.14	0 ± 0	0 ± 0	*546*
C39diene 3	39.01	0.14 ± 0.12	0 ± 0	0 ± 0	*544*
unknown CHC (RT 35.94)	39.15	0.09 ± 0.24	0 ± 0	0 ± 0	
13-;15-;17-MeC39	39.33	0.28 ± 0.21	0 ± 0	0 ± 0	196/392; 224/364; 252/336
13,25-; 13,x-DiMeC39	39.57	0.36 ± 0.2	0.11 ± 0.09	0 ± 0	196/406, 378/224
C41diene 1	40.58	0.71 ± 0.97	0 ± 0	0 ± 0	*572*
C41diene 2	40.71	0.22 ± 0.26	0 ± 0	0 ± 0	*572*
unknown saturated CHC	41.7	0.16 ± 0.09	0 ± 0	0 ± 0	

(RT 40.02) unknown unsaturated CHC (RT 42.36)	43.29	0.15 ± 0.18	0 ± 0	0 ± 0	

[1] assignment to peak is tentative

[2] identical to 17-; 19-; 21-C35ene (counted from the other end of the molecule)

Table S2.3: Cuticular hydrocarbons of the cryptic *Camponotus femoratus* species and *Camponotus spec.* SL. In this table we present the identified CHCs, their mean abundances and the retention indices (RI) diagnostic ions for identification. Numbers in italics are the M^+ ions; diagnostic ions in brackets were obtained after DMDS derivatization of unsaturated hydrocarbons. Note that, for unsaturated hydrocarbons, position of double bonds could only be determined for *Ca. femoratus* PAT.

CHC	**RI**	**mean ± SD in *Ca. femoratus* PAT**	**mean ± SD in *Ca. femoratus* PS**	**mean ± SD in *Ca. spec.* SL**	**diagnostic ions and M^+**
n-C20	20.07	0.39 ± 0.41	0 ± 0	0.22 ± 0.12	*282*
n-C21	21.03	7.97 ± 8.02	0 ± 0	4.47 ± 2.63	*296*
n-C22	22.07	0.18 ± 0.19	0 ± 0	0 ± 0	*310*
C23ene	22.83	0.23 ± 0.25	0 ± 0	0 ± 0	*322*
n-C23	23.00	1.14 ± 1.33	0 ± 0	0 ± 0	*324*
C25ene	25.01	0 ± 0	0 ± 0.03	0 ± 0	*350*
n-C25	25.06	0 ± 0	0.16 ± 0.15	0 ± 0	*352*
unknown unsaturated CHC (RT 13.22)	25.31	0 ± 0	0 ± 0.35	0 ± 0	
9-MeC25	25.35	0 ± 0	0 ± 0.24	0 ± 0	140/252
7-MeC25	25.41	0 ± 0	0 ± 0.09	0 ± 0	112/280
2-;4-MeC25	25.61	0 ± 0	0 ± 0.02	0 ± 0	43/351; 71/323
C26ene	25.94	0 ± 0	0 ± 0.02	0 ± 0	364
n-C26	26.00	0 ± 0	0 ± 0.09	0 ± 0	366
10-MeC26	26.34	0 ± 0	0 ± 0.09	0 ± 0	154/252
8-MeC26	26.37	0 ± 0	0 ± 0.06	0 ± 0	126/280
4-MeC26	26.63	0 ± 0	0 ± 0.52	0 ± 0	71/337
C27en	26.94	0 ± 0	0 ± 0.28	0 ± 0	*378*
n-C27	27.00	0 ± 0	0.31 ± 0.25	0.21 ± 0.12	*380*
9-MeC27en	27.36	0 ± 0	0 ± 2.37	0 ± 0	138/278, 153/293, *392*
n-C28	28.00	0 ± 0	0 ± 0.02	0 ± 0	*394*
4-MeC28	28.70	0 ± 0	0 ± 0	0.2 ± 0.16	71/365
n-C29	29.01	0.23 ± 0.19	0.75 ± 0.53	0.55 ± 0.14	*408*
11-; 13-;15- MeC29	29.39	0 ± 0	0.21 ± 0.27	2.19 ± 0.77	168/280; 196/252; 224/224
7-MeC29	29.43	0 ± 0	0 ± 0	0.28 ± 0.04	112/336
11,15-DiMeC29	29.61	0 ± 0	0 ± 0	2.66 ± 1.65	168/294, 238/224
7,11-DiMeC29	29.70	0 ± 0	0 ± 0	0.12 ± 0.12	112/350, 182/280
3-MeC29 (traces of C30ene)	29.81	0 ± 0	0 ± 0	0.39 ± 0.15	57/392
n-C30	30.00	0 ± 0	0.6 ± 0.33	0.57 ± 0.34	*422*
10-MeC30 (traces of 11-;12-MeC30)	30.35	0 ± 0	0.05 ± 0.09	0.52 ± 0.9	154/308; 168/294; 182/280
2-; 4-MeC30	30.64	0 ± 0	0.19 ± 0.56	8.63 ± 1.47	43/421; 71/393
C31ene 1	30.86	0.16 ± 0.11	0.24 ± 0.18	2.94 ± 1.63	*434*
C31ene 2	30.93	0 ± 0	0 ± 0	0.14 ± 0.11	*434*
C31ene 3	31.01	0 ± 0	0 ± 0	1.49 ± 0.48	*434*
n-C31	31.11	0.19 ± 0.17	2.13 ± 1.27	1.34 ± 0.17	*436*
9-;11-;13-;15-MeC31 (in *Ca. spec.* SL only 11-MeC31)	31.42	0.07 ± 0.66	1.04 ± 1.31	31.33 ± 4.15	140/336; 168/308; 196/280; 224/252
11,15-DiMeC31	31.64	0 ± 0	0 ± 0	21.17 ± 4.41	168/322, 238/252
11,15-;13,17-;13,21-;13,25-DiMeC31	31.66	0.03 ± 0.31	0.46 ± 0.74	0 ± 0	168/322, 238/252; 196/294, 266/224; 196/294, 322/168; 196/294,

					238/252
cf. DiMeC31	31.73	0 ± 0	0.09 ± 0.19	0 ± 0	
3-MeC31;15,21-DiMeC31	31.83	0.01 ± 0.16	0.47 ± 0.34	0.8 ± 1.38	57/421; 224/266, 322/168
C32ene	31.89	0 ± 0	0.01 ± 0.07	0.1 ± 0.17	*448*
11,15,19-TriMeC31	31.93	0 ± 0	0.32 ± 0.23	0 ± 0	168/336, 238/266, 308/196
unknown unsaturated CHC (RT 24.30)	32.12	0 ± 0	0.06 ± 0.38	9.35 ± 0.69	
n-C32	32.13	0 ± 0	0.58 ± 0.34	0 ± 0	*450*
10-;11-MeC32	32.34	0 ± 0	0 ± 0	0.78 ± 0.19	154/336; 168/322
13-;15-MeC32	32.42	0 ± 0	0.7 ± 0.31	0 ± 0	196/294; 224/266
C33diene	32.79	0 ± 0	0 ± 0.01	0.63 ± 0.27	*460*
C33ene	32.91	0 ± 0	0.67 ± 0.68	0.22 ± 0.37	*462*
n-C33	33.12	0.07 ± 0.06	0.83 ± 0.45	0.17 ± 0.29	*464*
11-;13-;15-MeC33 (in *Ca. spec.* SL mainly 11-MeC33)	33.44	0 ± 0	9.55 ± 2.75	1.71 ± 0.66	168/336; 196/308; 224/280
13,17-;15,19-DiMeC33	33.66	0 ± 0	0.87 ± 0.48	0.34 ± 0.58	196/322, 266/252; 224/294, 294/224
unknown saturated CHC (RT 26.62)	33.72	0 ± 0	0.15 ± 0.22	0.08 ± 0.14	
3-MeC33	33.89	0 ± 0	0.11 ± 0.14	0 ± 0	57/449
C34ene	33.94	0 ± 0	0.53 ± 0.23	0.09 ± 0.16	*476*
3,13-DiMeC33	34.14	0 ± 0	0.31 ± 0.15	0.19 ± 0.33	57/463, 210/308
13-;14-MeC34	34.42	0 ± 0	0.61 ± 0.15	0.04 ± 0.06	196/322; 210/308
cf. 16-MeC34	34.43	0 ± 0	0 ± 0	1.35 ± 2.33	238/280
14,18-;15,19-DiMeC34	34.63	0 ± 0	0.21 ± 0.13	0 ± 0	210/322, 280/252; 224/308, 294/238
C35ene	34.95	0 ± 0	2.75 ± 1.22	0.05 ± 0.09	*490*
n-C35	35.13	0 ± 0	0.45 ± 0.32	0 ± 0	*492*
13-MeC35ene	35.28	0 ± 0	0.28 ± 0.11	0 ± 0	195/334, 210/350, *504*
13-;15-;17-MeC35	35.44	0.12 ± 0.08	2.31 ± 0.62	0.66 ± 0.9	196/336; 224/308; 252/280
15,19-DiMeC35	35.65	0 ± 0	0.88 ± 0.65	0 ± 0	224/322, 294/252
13,x-DiMeC35	35.71	0 ± 0	0.13 ± 0.12	0.06 ± 0.1	196/350, x
3-MeC35	35.90	0 ± 0	0.12 ± 0.11	0 ± 0	57/477
C36ene	35.96	0 ± 0	1.12 ± 0.43	0.04 ± 0.07	*504*
C36diene	36.02	0 ± 0	0.16 ± 0.15	0 ± 0	*502*
3,15-DiMeC35	36.17	0 ± 0	0.45 ± 0.22	0 ± 0	57/491, 238/308
13-MeC36ene	36.24	0 ± 0	0.53 ± 0.15	0 ± 0	195/348, 210/364, *518*
15-;16-;17-;18-MeC36	36.39	0 ± 0	0 ± 0	0.29 ± 0.5	224/322; 238/308; 252/294; 266/280
12-;13-;14-MeC36	36.42	0 ± 0	0.24 ± 0.11	0 ± 0	182/364; 196/350; 210/336
unknown CHC (RT 30.40)	36.48	0 ± 0	0.03 ± 0.06	0 ± 0	
unknown CHC (RT 30.60)	36.63	0 ± 0	0.11 ± 0.1	0 ± 0	
12,x; 13,x; 14,x; 15,x-DiMeC36 (tentative x = 22-25)	36.68	0.24 ± 0.13	0 ± 0	0 ± 0	182/378, x; 196/364, x; 210/350, x; 224/336, x
C37diene 1	36.76	0 ± 0	0.57 ± 0.31	0 ± 0	*516*
C37ene 1 (9-C37ene; traces of C37diene)	37.04	0.7 ± 0.31	3.09 ± 2.05	0.11 ± 0.18	*518* [173/439]
C37diene 2	37.09	0 ± 0	9.34 ± 3.65	0.22 ± 0.37	*516*
unknown CHC (RT 31.26)	37.14	0.06 ± 0.05	0 ± 0	0 ± 0	
unknown unsaturated CHC (RT 31.38)	37.23	0.11 ± 0.12	0 ± 0	0 ± 0	
13-MeC37ene	37.29	0 ± 0	13.44 ± 4.01	0.58 ± 1	195/362, 210/378, *532*
13-; 15-; 17-MeC37	37.45	1.97 ± 0.88	0.8 ± 0.18	0.05 ± 0.09	196/364;

					224/336; 252/308
unknown unsaturated CHC (RT 31.70)	37.48	0.11 ± 0.16	0 ± 0	0 ± 0	
unknown methyl-branched alkene (RT 31.72)	37.49	0 ± 0	0.25 ± 0.1	0 ± 0	
15,21-; cf. 15,19-DiMeC37	37.63	0 ± 0	0.27 ± 0.25	0 ± 0	224/350, 322/252; 224/350, 294/280
13,23-DiMeC37	37.77	22.47 ± 10.21	0.43 ± 0.25	0.06 ± 0.1	196/378, 350/224
C38diene 1	37.80	0 ± 0	0.96 ± 0.28	0.03 ± 0.06	*530*
C38ene 1 (9-; 10-C38ene)	37.99	0.28 ± 0.15	0.42 ± 0.19	0 ± 0	*532* [173/453; 187/439]
C38diene 2	38.05	0 ± 0	0.61 ± 0.16	0.03 ± 0.06	*530*
3,15-DiMeC37	38.18	0 ± 0	1.22 ± 0.65	0.06 ± 0.11	57/518, 238/336
13-MeC38ene	38.24	0.12 ± 0.08	1.21 ± 0.39	0.08 ± 0.13	195/376, 210/392, *546*
13-; 15-; 17-MeC38 (in *Ca. femoratus* PS only 13-MeC38)	38.40	0.17 ± 0.07	0.13 ± 0.11	0 ± 0	196/378; 224/350; 252/322
8-MeC38	38.45	0 ± 0	0 ± 0	0.08 ± 0.14	126/448
C39diene 1	38.63	0 ± 0	0.06 ± 0.12	0 ± 0	*544*
12,x-;13,x-;14,x-;15,x-DiMeC38	38.71	0.44 ± 0.26	0 ± 0	0 ± 0	182/406, x; 196/392, x; 210/378, x; 224/364, x
C39diene 2 (cf. 9,29-C39diene)	38.81	4.11 ± 1.2	5.79 ± 2.08	0.16 ± 0.27	*544* [173/465, 187/451]
C39diene 3	38.85	0.04 ± 0.18	1.64 ± 0.71	0.05 ± 0.09	*544*
C39diene 4	38.96	0 ± 0	1.18 ± 0.48	0.05 ± 0.08	*544*
C39ene 1 (9-C39ene; traces of C39diene)	39.01	3.47 ± 1.89	0 ± 0	0 ± 0	*546* [173/467]
C39diene 5	39.09	3.9 ± 1.3	1.07 ± 0.42	0.38 ± 0.39	*544*
13-MeC39ene	39.27	1.85 ± 1.1	6.89 ± 1.92	0.41 ± 0.71	195/418, 210/434, *588*
9-MeC39	39.34	0 ± 0	0 ± 0	0.1 ± 0.17	140/448
7-MeC39	39.37	0 ± 0	0 ± 0	0.06 ± 0.1	112/476
13-; 15-; 17-MeC39	39.38	0.9 ± 0.38	0.39 ± 0.18	0.03 ± 0.05	196/392; 224/364; 252/336
cf. DiMeC39	39.47	0 ± 0	0.03 ± 0.08	0 ± 0	
13,x-;15,x-DiMeC39	39.52	0 ± 0	0 ± 0	0.04 ± 0.06	196/406, x ; 224/378, x
15,19-; 13,23-DiMeC39	39.60	0 ± 0	0.35 ± 0.27	0.04 ± 0.06	224/378, 294/308; 196/406, 350/252
13,x-;15,x-;17,x-DiMeC39 (tentative x = 23; 25; 27)	39.70	2.82 ± 1.08	0 ± 0	0 ± 0	196/406, x; 224/378, x; 252/350, x
C40diene 1	39.77	0 ± 0	1.04 ± 0.38	0.06 ± 0.1	*558*
C40diene 2	39.82	1.69 ± 0.53	0 ± 0	0 ± 0	*558*
C40diene 3	39.84	0 ± 0	1.09 ± 0.37	0.06 ± 0.11	*558*
C40diene 4	39.88	0 ± 0	0.14 ± 0.12	0 ± 0	*558*
unknown CHC (RT 35.16)	39.91	0.02 ± 0.07	0 ± 0	0 ± 0	
C40diene 5	40.00	1.06 ± 0.39	0 ± 0	0 ± 0	*558*
unknown CHC (RT 35.42)	40.09	0.02 ± 0.04	0 ± 0	0 ± 0	
13-MeC40ene	40.25	0.17 ± 0.11	0.94 ± 0.35	0.08 ± 0.14	195/404, 210/420, *574*
8-MeC40	40.53	0 ± 0	0 ± 0	0.04 ± 0.07	126/476
C41diene 1	41.16	13.24 ± 4.42	5.53 ± 1.83	0.29 ± 0.51	*572*
C41diene 2 (cf. 9,29-; 9,31-C41diene) [1]	41.25	11.61 ± 3.42	3.71 ± 0.95	0.2 ± 0.35	*572* [173/493, 215/451; 187/479]
C41diene 3 (cf. 9,33-C41diene) [1]	41.32	7.46 ± 1.81	0 ± 0	0 ± 0	*572* [173/493, 159/507]
C41diene 4 (cf. 6,32-C41diene) [1,2]	41.41	2.86 ± 0.99	0 ± 0	0 ± 0	*572*
C41diene 5	41.51	0.59 ± 0.38	0 ± 0	0 ± 0	*572*
C41diene 6	41.63	0.73 ± 0.65	0 ± 0	0 ± 0	*572*
13-MeC41ene	41.70	0 ± 0	2.82 ± 1.22	0.19 ± 0.32	195/418, 210/434, *588*
C41diene 7	41.79	0.1 ± 0.12	0 ± 0	0 ± 0	*572*
13,29-; x,y-DiMeC41	42.29	0.45 ± 0.31	0 ± 0	0 ± 0	196/434, 434/196

15,19-DiMeC41	42.34	0 ± 0	0.11 ± 0.16	0 ± 0	224/406, 294/336
C42diene 1	42.40	0.23 ± 0.11	0 ± 0	0 ± 0	*586*
unknown CHC (RT 38.70)	42.54	0 ± 0	0.02 ± 0.06	0 ± 0	
C42diene 2	42.61	0.42 ± 0.21	0.67 ± 0.36	0.06 ± 0.11	*586*
13-MeC42ene	43.08	0 ± 0	0.13 ± 0.14	0 ± 0	195/432, 210/448, *644*
cf. C43triene 1	44.27	0.84 ± 0.47	0 ± 0	0 ± 0	*598*
cf. C43triene 2	44.46	2.16 ± 0.99	1.28 ± 0.74	0.08 ± 0.13	*598*
cf. C43triene 3	44.61	1.8 ± 0.86	0 ± 0	0 ± 0	*598*
unknown unsaturated CHC (RT 41.78)	44.78	0 ± 0	0.09 ± 0.11	0 ± 0	
13-MeC43ene	44.88	0 ± 0	0.08 ± 0.14	0 ± 0	195/446, 210/462, *658*

[1] assignment to peaks is tentative

[2] identical to 9,35-C41diene (counted from the other end of the molecule)

Table S2.4: Reconstructed ancestral states of average CHC chain length in the *Orthocrema* clade. The table shows the average CHC chain length reconstructed for each node. Nodes are numbered from the most ancestral to the most recent one.

Node	Ancestral average chain length	95% Confidence intervals
Node 1	27.91	[22.39; 33.43]
Node 2	27.79	[23.31; 32.26]
Node 3	27.81	[23.64; 31.99]
Node 4	27.63	[26.13; 29.14]
Node 5	28.10	[22.95; 33.24]
Node 6	28.26	[23.25; 33.28]
Node 7	29.02	[24.50; 33.53]
Node 8	29.02	[24.50; 33.53]
Node 9	29.38	[24.62; 34.14]
Node 10	28.83	[24.46; 33.20]
Node 11	29.11	[25.93; 32.28]
Node 12	29.49	[28.01; 30.97]

Table S2.5: CHCs influenced by the climate. The table lists the 20% of substances differing most according to climate and being abundant more than 0.5% for each of the cryptic species. Shown are pseudo-F and p-values from univariate PERMANOVAs based on Euclidean distances, plus the direction of the effect.

species	substance	pseudo-F	p	increasing/decreasing with higher precipitation (+/-)
***Crematogaster levior* A**	13,17-;15,19-DiMeC35	25.445	0.001	-
	13-;15-MeC35ene	24.500	0.001	+
	C33diene 2	21.180	0.001	+
	15-MeC37ene	22.933	0.001	+
	C35ene 2 (9-C35ene)	10.524	0.001	-
	C37diene 1 (cf. 9,19-; 9,21-C37diene)	14.176	0.002	-
	13-;15-;17-MeC37	7.995	0.005	-
	C35diene 2	19.269	0.001	+
	11-;13-;15-MeC31	5.885	0.014	-
	C33diene 3	15.573	0.001	+
	n-C27	8.084	0.011	-
***Crematogaster levior* B**	C29diene 2	17.803	0.001	-
	C31ene 3	14.508	0.002	+
	C33ene 1	11.520	0.002	+
	C31diene 2	8.264	0.004	-
***Camponotus femoratus* PAT**	C39diene 5	19.210	0.001	-
	13-MeC39ene	12.457	0.001	-
	12,x-;13,x-;14,x-;15,x-DiMeC37	8.485	0.001	+
	C41diene 4 (cf. 6,32-C41diene)	8.171	0.008	-
	C41diene 6	7.380	0.009	-
	C39ene 1	7.224	0.009	-
	C41diene 5	6.926	0.007	-
	13,23-DiMeC37	5.341	0.025	+

***Camponotus femoratus* PS**	13-MeC38ene	21.718	0.001	-
	13-MeC41ene	19.905	0.001	-
	13-MeC40ene	17.999	0.001	-
	C39diene 2	16.773	0.001	+
	13-MeC39ene	13.236	0.002	-

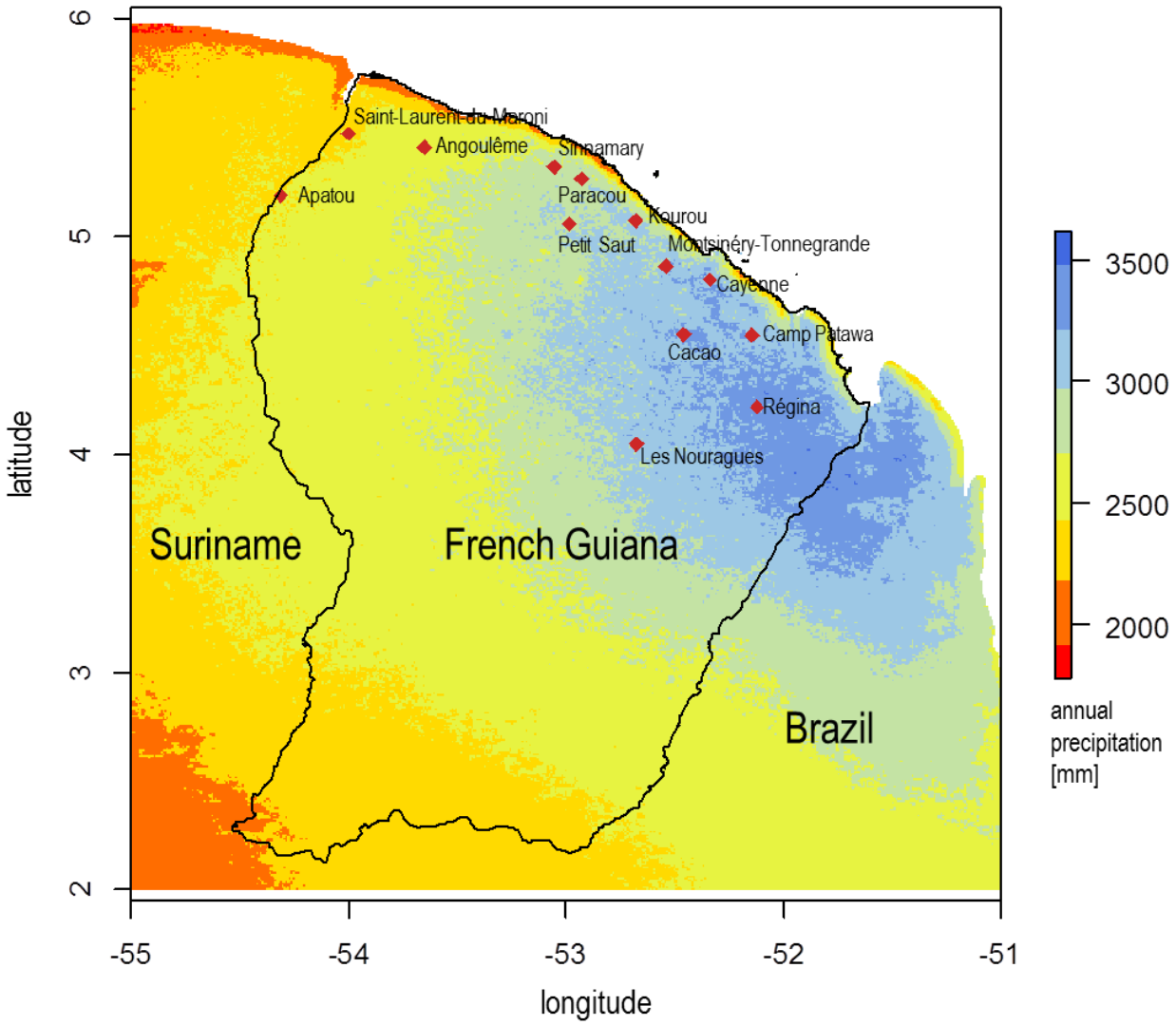


Figure S2.1: Map of our sampling locations in French Guiana. The sampling locations are indicated by red diamonds. Annual precipitation in French Guiana was downloaded from WorldClim and is indicated by colors. For coordinates and numbers of samples see Table S1.

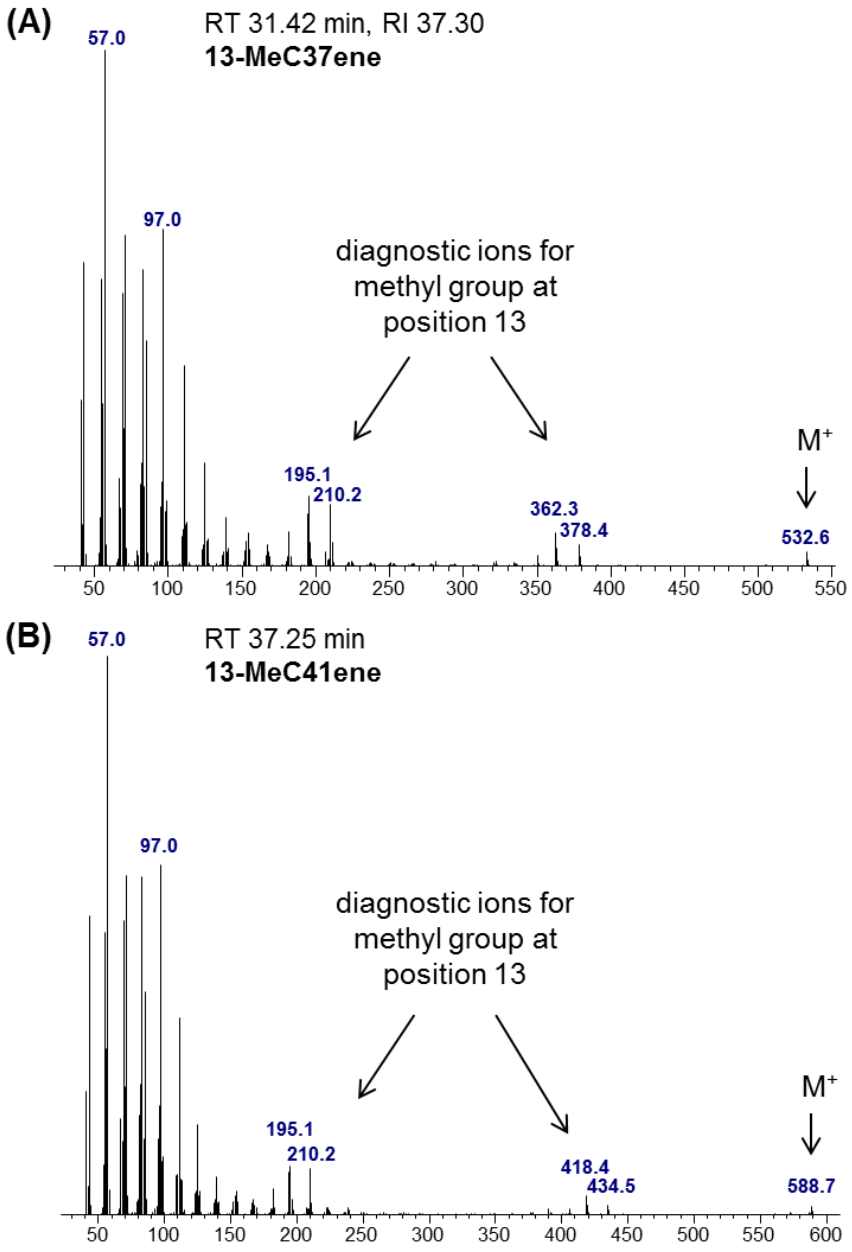


Figure S2.2: Representative mass spectra for two methyl-branched alkenes of *Ca. femoratus* PS. Methyl-branched alkenes eluted shortly before the saturated methyl-branched alkanes with the same methyl-branch position. We identified them based on the molecular peak (M^+) that is two amu below that expected for monomethyl alkanes of the same chain length. Because of the double bond, molecules break in two different ways leading to four diagnostic ions instead of two. The first pair of ions is similar to those expected for the saturated molecule minus two amu (here 195/362 m/z in (A) and 195/418 m/z in (B)) and the second pair is 15 to 16 amu heavier than the first ones (here 210/378 and 210/434 m/z). For comparison, the diagnostic ions of (saturated) 13-MeC37 are 196 and 365 m/z; those of 13-MeC41 are 196 and 420 m/z.

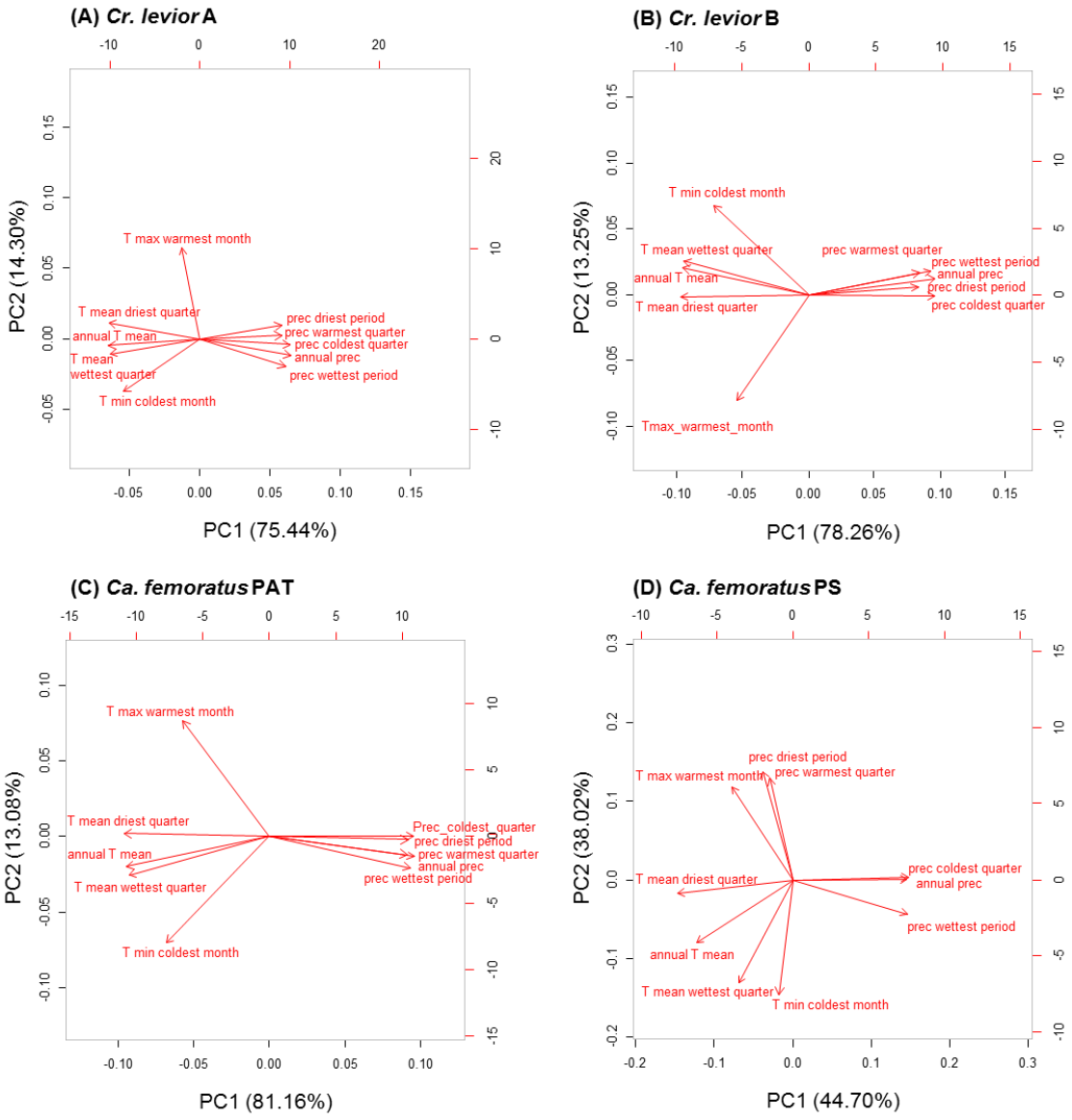


Figure S2.3: Factor loadings of the climate variables on CHC profiles. The arrows indicate factor loadings of each of the 10 climate variables concerning principal components 1 and 2. Temperature (annual mean temperature, minimum temperature in the coldest month, maximum temperature in the warmest month, mean temperature in the driest quarter and mean temperature in the wettest quarter) is abbreviated with "T" and precipitation (annual precipitation, precipitation in the coldest quarter, precipitation in the warmest quarter, precipitation in the driest period and precipitation in the wettest period) with "prec".

CHAPTER 3

Hybrid Genome Assembly of a Neotropical Mutualistic Ant

Juliane Hartke, Tilman Schell, Evelien Jongepier, Hanno Schmidt, Philipp P. Sprenger, Juraj Paule, Erich Bornberg-Bauer, Thomas Schmitt, Florian Menzel, Markus Pfenninger, Barbara Feldmeyer

Published as:
Hartke J, Schell T, Jongepier E, Schmidt H, Sprenger PP, Paule J, Bornberg-Bauer E, Schmitt T, Menzel F, Pfenninger M, Feldmeyer B (2019). Hybrid Genome Assembly of a Neotropical Mutualistic Ant. Genome Biology and Evolution. 11:2306–2311.

ABSTRACT

The success of social insects is largely intertwined with their highly advanced chemical communication system that facilitates recognition and discrimination of species and nest-mates, recruitment, and division of labour. Hydrocarbons, which cover the cuticle of insects, not only serve as waterproofing agents but also constitute a major component of this communication system. Two cryptic *Crematogaster* species, which share their nest with *Camponotus* ants, show striking diversity in their cuticular hydrocarbon (CHC) profile. This mutualistic system therefore offers a great opportunity to study the genetic basis of CHC divergence between sister species. As a basis for further genome-wide studies high-quality genomes are needed. Here we present the annotated draft genome for *Crematogaster levior* A. By combining the three most commonly used sequencing techniques – Illumina, PacBio and Oxford Nanopore – we constructed a high-quality *de novo* ant genome. We show that even low coverage of long reads can add significantly to overall genome contiguity. Annotation of desaturase and elongase genes, which play a role in CHC biosynthesis revealed one of the largest repertoires in ants and a higher number of desaturases in general than in other Hymenoptera. This may provide a mechanistic explanation for the high diversity observed in C. *levior* CHC profiles.

Keywords: Cuticular Hydrocarbons, Assembly, MinION, Formicine, Elongase, Desaturase

INTRODUCTION

The genomic basis of chemical communication is still mostly unknown, despite its importance in animal behaviour. A prime example are social insects, in which cuticular hydrocarbons (CHCs) represent the most important means of communication and facilitate the functioning of complex social organization. They enable the expression and recognition of various attributes, such as species and nest-mate status, caste, sex and fertility (Lahav *et al.* 1999; Dietemann *et al*. 2003; Leonhardt *et al.* 2016). CHCs cover the cuticle of all insects and originally evolved as a protection against desiccation (Blomquist & Bagnères 2010; Menzel *et al.* 2018). Due to their function in both ecological adaptation and mate signalling, they were proposed as drivers of speciation (Smadja & Butlin 2009; Thomas & Simmons 2009; Chung & Carroll 2015), and thus may have driven the high diversity witnessed today in social insects.

One of the most successful families of social insects are ants with about 13,000 recognised species (Chomicki & Renner 2017). They occur in virtually all terrestrial habitats, barring the polar regions, and evolved a striking diversity in life-history traits, morphology and behaviour. This diversity, however, is not reflected in the number of published genomes so far (n = 19).

The Neotropical ant species *Crematogaster levior* and *Camponotus femoratus* are representative of the remarkable diversity within this family, as they mutualistically share a nest, a so-called ant garden (Davidson 1988). Obligate mutualisms that are characterized by a benefit for both partners are rare. Here *Crematogaster* benefits from strong defence capabilities of *Camponotus*, whereas the latter benefits from *Crematogasters* efficiency in finding resources (Vantaux *et al.* 2007). Both species show unusually high diversity in their cuticular hydrocarbon profiles (Menzel *et al.* 2014) that were now shown to represent cryptic species (Hartke *et al.* 2019a). This mutualism therefore offers the unique chance to study the underlying genomic basis of CHC complexity and their putative function in species divergence in two closely related species. Here, we present the first annotated draft genome for one of the cryptic *Crematogaster* species, *Crematogaster levior* A, and compare the number of genes with putative function in communication to other available ant and hymenopteran genomes.

MATERIAL AND METHODS

Sample collection and sequencing:

Specimens for sequencing were collected from a single nest in, French Guiana (4°33'14.5"N 52°09'02.4"W), in September 2016. The ants were stored in 96% ethanol until DNA

isolation. We followed a hybrid approach, acquiring sequences from three different sequencing platforms. To obtain sufficient amounts of DNA for sequencing, we pooled 70 larvae for HiSeq 2000 (Illumina Inc, CA, USA) paired-end sequencing, 110 larvae for two SMRT cells on PacBio Sequel (Pacific Biosciences, CA, USA) and > 300 larvae for a total of six sequencing runs on an Oxford Nanopore Technologies (ONT), UK, MinION. Illumina and PacBio sequencing were conducted at the Beijing Genomics Institute (BGI), Hong Kong, and Oxford Nanopore sequencing inhouse.

DNA for Illumina sequencing was isolated with the DNeasy Blood and Tissue Kit (Qiagen), following manufacturer's instructions. DNA isolation and library preparation for PacBio sequencing were partly conducted by BGI, Hong Kong, plus additional DNA isolated from our lab by DNeasy Blood and Tissue kit. We constructed four different libraries for a total of six ONT MinION runs, for which we tested different DNA isolation and library preparation protocols. We isolated two DNA samples following the Qiagen Blood and Tissue Protocol, and two samples following Urban *et al*. (2015 PREPRINT), which is optimized for long high molecular weight DNA. The library preparation was conducted three times following the latest ONT protocol and once using the Urban *et al*. (2015 PREPRINT) protocol (details in Supplementary Information M1 and Supplementary Table S3.1).

For transcriptome sequencing, specimens of the same nest were freeze killed at -80°C. We isolated RNA from different worker stages (newly emerged and old workers, young and old worker pupae). We furthermore isolated RNA from eggs of an additional colony. Extraction protocol followed Alleman *et al.* (2018). Sequencing on a HiSeq 2000 was conducted by BGI, Hong Kong. For extraction, pre-assembly processing and assembly protocol please refer to supplement M2. We furthermore assembled transcriptomes of the sister species, *C. levior* B (BioProject PRJNA540400; Sprenger *et al*. submitted).

Assembly strategy

Illumina reads were quality-trimmed and filtered for adapter sequences with the BBDuk algorithm from BBMap v36.92 (Bushnell 2014), screened for contamination using FastQ Screen v0.10.0 (Wingett *et al.* 2018), and filtered for mtDNA with BBDuk. Before and after every processing step, read quality was checked with FastQC v0.11.3. PacBio reads were quality corrected with Proovread v2.14.0 (Hackl *et al.* 2014), using the Illumina read set to obtain high-quality reads. MinION reads were base called and quality-filtered with the Nanopore basecaller Albacore v2.0 (ONT, UK) and subsequently filtered for mtDNA with BBDuk. For more details see supplementary material M3.

The Illumina read set was assembled with SPAdes v3.10.0 (Bankevich *et al.* 2012) using default settings, and the resulting assembly was triplicated to a coverage of 3x to be included

by the algorithm of the next assembler. This set of contigs, together with ONT and PacBio reads was assembled with the long-read assembler Ra (github.com/rvaser/ra; commit ID: 65bedfe). The resulting assembly was scaffolded with SSPACE-LongRead v1.1 (Boetzer & Pirovano 2014) using ONT and PacBio long reads (see supplementary methods M4). We assessed repeat content within our Illumina read set using RepeatExplorer (Novák *et al.* 2013), and checked for the completeness of gene space with BUSCO v2.0 (Simão *et al.* 2015) with the provided database for hymenopteran orthologous genes.

Genome size estimation

We estimated genome size by dividing the total number of nucleotides used in the Illumina assembly by the peak coverage resulting from mapping those reads back to the assembly (Schell *et al.* 2017). Additionally, genome size was also estimated using flow cytometry with three individuals of *C. levior* A, and *Glycine max* cv. Polanka as an internal standard (see supplementary methods M5).

Annotation strategy

Before annotation, we masked all regions that were covered only by uncorrected PacBio or MinION reads with bedtools *maskfasta* (Quinlan & Hall 2010), to base gene predictions only on high-quality information throughout the assembly. Gene annotation was conducted using the MAKER2 pipeline v2.31.8 (Holt & Yandell 2011). As evidence, we used transcriptomes from *C. levior* A; additional ESTs from the sister species, *C. levior* B (worker; BioProject PRJNA540400**;** see Sprenger *et al.* submitted); *ab initio* models from SNAP v2006-07-28 (Korf 2004), Augustus v3.2.2 (Stanke *et al.* 2006) and GeneMark v4.32 (Lomsadze *et al.* 2005); and the repeat library. As protein homology evidence, we used the SwissProt Database (accessed September 22nd, 2017) and an annotated protein set of *Cardiocondyla obscurior*, which is the most closely related ant species with a published genome (Schrader *et al.* 2014). For a more detailed protocol refer to supplement M6. Moreover, we manually annotated elongases and desaturases (supplementary methods M7). We also searched for elongases and desaturases in 43 annotated Hymenoptera genomes via a blastp v2.5.1 (Camacho *et al.* 2009) and PfamScan v1.6 (Punta *et al*. 2012) workflow (see supplementary methods M8).

RESULTS AND DISCUSSION

Genome sequencing and assembly

An overview of raw sequences obtained from each sequencing strategy and number of trimmed reads can be found in supplement Table S3.3. Genome size, assessed by the peak coverage approach (Schell *et al.* 2017), was estimated to be 355.52 Mbp. This estimate is at

the higher end but still within range compared to other ant genomes (Table S3.12). Genome size (2C-value) was also estimated by flow cytometry (see supplement M4). When correcting the original *G. max* calibration (Doležel *et al.* 1994) for the newest human reference genome assembly (GRCh38.p13), the 2C value corresponds to 409.96 Mbp (1 pg = 978 Mbp, Doležel *et al.* 2003), which is within range of previously reported estimates, although significantly larger than estimates for the same genus (*Crematogaster hespera*: 275.9 Mbp; Tsutsui *et al.* 2008). The difference in size estimates from flow cytometry and peak coverage might be explained by the loss of sequences during library preparation. Regions in the DNA with long stretches of repeats are prone to harbour breakage points or form secondary structures, such as hairpins (De Bustos *et al.* 2016), that hinder sequencing in those regions and thereby lead to faulty coverage estimations by read distribution.

Table 3.1: Overview of different assembly approaches for *Crematogaster levior* A using different combinations of Illumina, MinION and PacBio reads.

Read type	#Contigs	N50	Length [Mbp]	Recovered BUSCO [%]
Illumina	52,838	15,083	259.9	95.4
MinION	3,420	39,345	114.3	2.8
PacBio	3,270	142,016	319.9	0
PacBio polished	3,615	104,646	298.8	90.5
MinION & PacBio	1,898	361,377	326.6	10.1
MinION & PacBio polished	2,207	260,013	325.9	11.7
PacBio polished & Illumina (3x)	3,311	120,772	299.9	92.4
PacBio polished & MinION & Illumina (3x)	2,298	242,096	324.2	98.0

NOTE.—Illumina (3x): Illumina reads were added as triplicates to the hybrid assembly. All assemblies were conducted with Ra, except for the Illumina only assembly that was assembled using Spades.

Assembly and scaffolding resulted in 1,523 scaffolds with a N50 length of 383,244 bp and a total length of 326.2 Mbp (peak coverage: 92% of the estimated size, flow cytometry: 80% of the estimated size). To assess gene-space completeness of the draft genome, BUSCO v2.0 was used with the provided Hymenoptera dataset of core orthologues, of which 98.0% could be retrieved (N = 4,415; complete: 95.9%, fragmented: 2.1%, missing: 2.0%), suggesting a high level of completeness and contiguity of coding regions.

Approximately 12.2% of the genome assembly consist of repeats, with the largest portion being labelled as unclassified (65%), followed by LINEs and LTRs (both 11%) (Figure S3.1). Most ant genomes sequenced so far, have higher reported repeat contents (mean = 24%; Table S3.12). Especially when regarding the fact that up to 20% of the estimated genome size could not be assembled, which is most likely due to repeat regions, the estimates by RepeatExplorer (10.5%) and RepeatModeler (3.2%) seem too low, which is in line with the above given reasoning of either break points and/or secondary structures of the DNA in

repeat regions, which leads to lower representation of these regions in the sequences used for assembly. Backmapping rates are very high with over 96% for each sequencing method (Table S3.11), indicating that over 95% of the actually sequenced reads are represented in the final assembly.

Table 3.2: Genome statistics of final assembly, containing all three read types, after scaffolding

Parameter	**Value**
#Scaffolds	1,523
Assembly length	326.2 Mbp
N50	383,244 bp
Gaps (N)	0.63%
BUSCO orthologous genes present	98.00%
#Genes	17,855
Gene space (UTR, Exons, Introns etc.)	103 Mbp (31.66% of assembly)
Mean distance between genes	6,479 bp
#Exons	117,323
Exon space	36 Mbp (11.27% of assembly)
Exons/gene	6.6

Comparison of assembly strategies

We used different combinations of our read data as input for Ra and are thus able to compare the influence of single read types on the accuracy and contiguity of the assembly (Table 3.1). From all single read type assemblies, the one from uncorrected PacBio reads seemed to be the most continuous, but it lacks in accuracy with 0% of BUSCO orthologues found. Prior correction with Illumina data improved the assembly immensely (90.5% found orthologues). When combining corrected PacBio reads with the Illumina assembly, quality metrics improved further, albeit only slightly. The MinION only assembly also lacked in accuracy and compared to the PacBio assembly, also in completeness (32% of final assembly length). A combination of the corrected PacBio reads with MinION reads lead to a substantial drop in accuracy (11.7% found orthologues) compared to the assembly without MinION reads. By combining all three read types, we obtained the best results in terms of length and accuracy (98% of orthologues). Especially, when comparing this 3-way assembly to the one lacking MinION reads, the difference in contiguity and accuracy is striking. N50 increased by more than 120 kbp and we found 6% more BUSCO orthologues. This shows that even a coverage of MinION reads as low as 9x can significantly increase assembly contiguity, although this only held true when Illumina reads were added.

Finally, we analysed which fraction of the final assembly was uniquely covered by single read types (Table S3.11). Only 1.05% of the draft was covered solely by Illumina reads. For PacBio, the percentage was higher with 2.33%, including 1.31% of the assembly that was

covered by uncorrected PacBio reads only. Genome positions that were only covered by MinION reads made up 2.42% of the final assembly.

Annotation report

MAKER2 annotation resulted in 17,855 genes that comprise 31% of the assembly space (Table 3.2). The number of annotated genes is within the same range as other annotated ant genomes (Table S3.12). Using a blastp search against the NCBI non-redundant invertebrate database (accessed March 2019), we were able to retrieve 14,713 genes, indicating 3,142 putative taxonomically restricted genes within *C. levior* A. This number is lower than previously found in other Hymenoptera species (Simola *et al*. 2013), however, the number of available genomes and thereby the number of similar genes increased in the meantime, which may explain the discrepancy. Mean GC content genome-wide (36%), within exons (43%) and within introns (30%) was similar to other reports on invertebrates (Jiang *et al*. 2014).

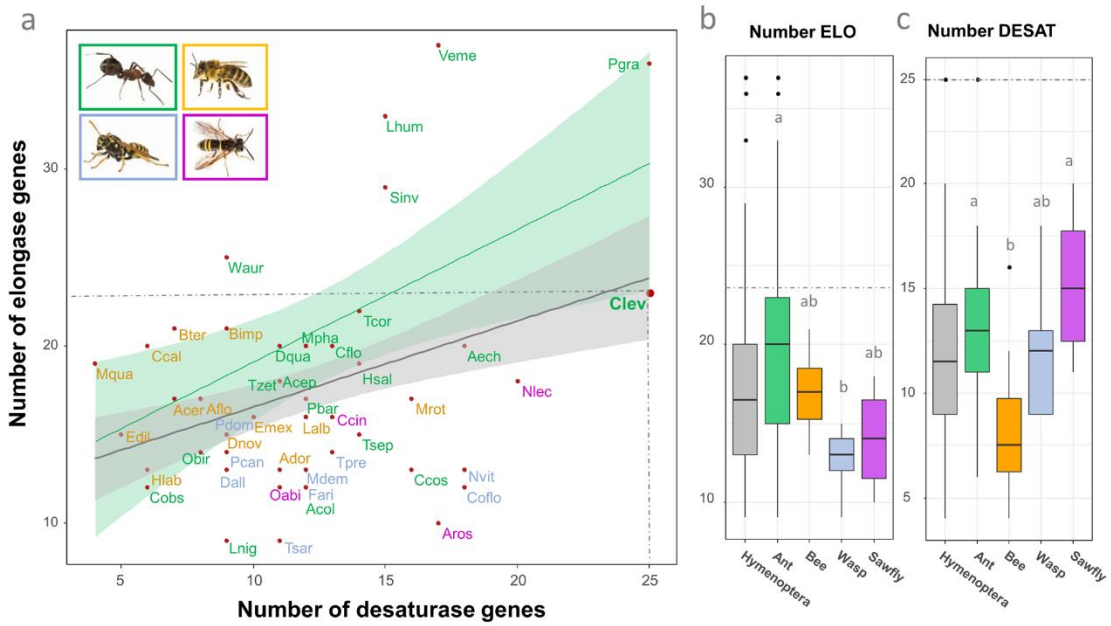


Figure 1: (a) Relationship between the number of elongase and desaturase genes across 48 hymenopterans (see also Table S3.8). The different colours depict the different families (green: ants, yellow: bees, red: wasps, purple: sawflies). Similarly, the green regression was calculated based on ants, while the grey regression was calculated based on all Hymenoptera. Pictures show exemplary species for each family (ant, bee, wasp (all ©Barbara Feldmeyer), sawfly (©Alex Hyde)). Comparison of the number of **(b)** elongases and **(c)** desaturases across hymenopteran families. Different letters indicate significant difference in number of genes (significance level: $p < 0.05$; One-way ANOVA, Tukey's HSD, Table S3.9). The dotted lines indicate the number of genes found in *Crematogaster levior* A.

Comparison of gene families

Elongases and desaturases are among the gene families that play key roles in the biosynthesis of CHCs (Falcón *et al.* 2014). To detect differences within gene family sizes

between closely related species, high quality genomes are needed. By manually annotating elongases and desaturases, we moreover tested the contiguity of our assembly, and found 23 elongases and 25 desaturases in the *C. levior* A draft genome (Table S3.6). We compared these values to 47 other hymenopteran draft genomes (Figure 3.1, Table S3.8) and found significant differences between groups (elongases: $p = 0.015$; desaturases: $p = 0.008$, one-way ANOVA). Ants had significantly more elongases than wasps (Figure 3.1b, Table S3.9) and bees had significantly fewer desaturases than ants and sawflies (Figure 3.1c, Table S3.9). Among all analysed species, *C. levior* A and *Pseudomyrmex gracilis* had the highest number of desaturases (mean number in ants: 13.7, Hymenoptera: 12.0). In line with increased chain elongation in *C. levior* A (Sprenger *et al.* 2019), their number of elongases was higher than the general mean in ants (20.6, Hymenoptera: 17.5). On the one hand this increased number of elongases and desaturases may be a major part of the genomic basis of high intraspecific CHC variation reported in *C. levior* (Menzel *et al.* 2017b), on the other hand it might be reflective of a highly contiguous and complete assembly within coding regions.

CONCLUSION

Here we present the annotated draft genome of *Crematogaster levior* A. By using a hybrid assembly approach encompassing three different sequencing techniques, and by combining high-quality short reads with long reads, we were able to produce a high-quality *de novo* ant genome assembly. Even rather low coverages of long reads significantly increased accuracy and contiguity and are a good and cost-effective way to obtain high-quality draft genomes. A comparison to other Hymenoptera yielded strong differences between species in the total number of desaturase and elongase genes. Among all analysed species, *C. levior* A (together with *P. gracilis*) showed the highest number of desaturases, which may be reflective of their high intraspecific diversity in cuticular hydrocarbon profiles.

ACKKNOWLEDGEMENTS

This work was supported by the German Science Foundation (DFG) as a grant to BF (FE 1333/7-1), ThS (SCHM 2645/7-1) and FM (ME 3842/5-1). This work was additionally supported by the LOEWE Centre Translational Biodiversity Genomics (TBG), funded by the state of Hesse, Germany. EBB and EJ were supported by a DFG grant (BO2544/12-1). This study is compliant with the Nagoya protocol (permission number: TREL1734890A/13).

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY METHODS

M1: ONT MinION sequencing protocol

Sequencing was conducted using six flowcells and four different DNA isolations and library preparations. The first isolation and library preparation followed standard DNeasy Blood and Tissue kit (Qiagen) and standard Nanopore protocol for 2D reads. The second isolation and library preparation were done following the protocol of Urban *et al*. (2015 Preprint) specifically designed to obtain long reads. For the third run, we used DNA isolated with the Urban *et al*. protocol and prepared the library following standard MinION 1D protocol. The fourth run was prepared as the first one, but with the MinION 1D protocol, and we used library left over from the first run to top up the flowcell during sequencing. For the fifth run, we used libraries prepared for the third and fourth run, and for the sixth run, we used a left-over library from the third run.

Table S3.1: Overview of Oxford Nanopore MinION sequencing runs and their output

	DNA isolation	Library prep	Base pairs	Reads	N50 [bp]	Max length [bp]
Run 1	Qiagen	Nanopore 2D	479,650,038	373,578	2,674	32,955
Run 2	Urban *et al*.	Urban *et al*.	13,110,799	5,219	4,944	69,407
Run 3	Urban *et al*.	Nanopore 1D	274,508,112	47,050	10,304	133,542
Run 4	Qiagen	Nanopore 1D + Run 1	1,080,604,289	653,752	2,156	73,473
Run 5		Run 3 + 4	800,540,339	316,154	4,036	277,903
Run 6		Run 3	696,465,807	120,491	9,497	445,755

M2: Transcriptomes – extraction, processing, and assembly

Frozen samples of different worker stages (eggs, young pupae, old pupae, newly emerged worker and older worker) were ground in Trizol until homogenized. After adding 200 µl of chloroform the sample was shaken vigorously for 15 seconds. The mixture was centrifuged for 15 min at 4 °C at 11,600 rcf. The resulting aqueous phase was transferred to a new RNase-free tube and precipitated by adding 0.5 volume of absolute ethanol. The solution was pipetted four times to mix and then transferred to a RNeasy mini-spin column (Qiagen). The following steps follow step three onwards of the RNeasy Clean-Up manual (Qiagen). The column was centrifuged for 30 seconds at 10,000 rcf. RW-Buffer was added, and the column was centrifuged again. A wash step with RPW buffer followed and the column was then placed in a new 1.5 ml collection tube. For elution, 30 µl RNase free water

was pipetted directly onto the column and incubated for 1 min at room temperature. A final centrifugation step for 1 minute at 10,000 rcf. followed.

Transcriptomes were sequenced by BGI, Hong Kong, on an Illumina HiSeq 4000 to obtain paired-end reads with an insert size of 200 bp. The resulting sequences plus additional sequences from the same species (Sprenger *et al.* submitted) were trimmed using the autotrim v0.5 wrapper script based on Trimmomatic (Bolger *et al*. 2014; Waldvogel *et al.* 2018) (Trimmomatic-0.36/adapters/TruSeq3-PE-2.fa:2:30:10 TRAILING:3 SLIDINGWINDOW:4:15 -nok -rn). The resulting reads were quality checked with FastQC v0.11.3 (Andrews 2010) and assembled with Trinity v2.0.6 (Grabherr *et al.* 2013) (--seqType fq --min_contig_length 300 --full_cleanup --no_bowtie). We furthermore obtained sequences of the cryptic sister species, *Crematogaster levior* B (Sprenger *et al.* submitted), and followed the same assembly protocol with Trinity.

M3: Preparing of sequences prior to assembly

M3.1 Quality processing Illumina:

First, Illumina adapters were trimmed using BBDuk (part of BBMap 36.92) (minlen=25 qtrim=rl trimq=20 ktrim=r k=25 mink=11 ref=/opt/bbmap/resources/adapters.fa hdist=1 tpe tbo). Then, BBDuk was used to discard bases below a quality threshold of 20 (minlen=25 qtrim=rl trimq=20 k=25 mink=11). Only reads that were still properly paired after these filtering steps were retained, meaning if one read of the pair was removed during trimming, the other one was removed as well. We checked our read set for contaminations with FastQ Screen 0.10.0 (Wingett *et al.* 2018) using a custom database (Table S3.2). The mitochondrial genome was assembled with MITObim (Hahn *et al.* 2013) with default parameters by using the mitochondrial genomes of *Atta cephalotes (HQ415764.1)*, *Pristomyrmex punctatus (NC_015075)* and *Vollenhovia emeryi* (NC_030176.1) as a reference. The resulting assembly was then used in BBDuk (k=41 hdist=2) to filter our Illumina read set for mitochondrial DNA before the assembly.

Table S3.2: Reference species and databases used for contamination filtering with FastQ Screen

Species / database	Accession numbers / version
Homo sapiens	GCF_000001405.35
Mus musculus	GCF_000001635.25
Escherichia coli	U00096.2
PhiX	NC_001422.1
Ananas comosus	GCA_001540865.1
Boea hygrometrica	GCA_001598015.1
Orchesella cincta	GCA_001718145.1
Spirodela polyrhiza	GCA_000504445.1
Wolbachia	GCF_000204545.1
Deconseq bact/vir	V0.4.3

M3.2 Quality processing MinION Nanopore reads:

MinION reads were basecalled using the Nanopore basecaller Albacore with standard settings. This resulted in a total of 3.34 gbp distributed on 1.51 million reads. N50 was highly variable between runs and ranged from 2.1 kb to 10.3 kb. As for Illumina reads, MinION reads were filtered for mitochondrial DNA using BBDuk and subsequently for adapters used during sequencing (k=17 mink=11 hdist=2 ref=hairpin.fasta).

M3.3 Quality processing PacBio reads:

Sequencing with PacBio Sequel results in reads that were already filtered for adapters. Since these reads do not have any quality scores, we did not perform any filtering steps regarding quality as we did for Illumina reads. Instead, we used Proovread (Hackl *et al.* 2014) with standard settings to correct the data set with Illumina reads, that are less error-prone. For the final hybrid assembly, we used quality corrected, but untrimmed PacBio reads. This means that reads that could only partly be corrected by Illumina reads retained their original length and were only corrected for those positions with enough Illumina coverage.

Table S3.3: Sequence data before and after processing steps

	Before processing		**after processing**	
	Reads	Nucleotides	Reads	Nucleotides
Illumina	256,243,982	35,057,868,150	233,435,232	30,980,846,066
PacBio	1,138,469	6,653,714,229	656,310	4,211,462,945
MinION	1,516,244	3,344,879,393	1,483,045	3,204,107,205

M4: Assembly

a) Illumina only

For the Illumina only assembly we used SPAdes v3.10.0 (Bankevich *et al.* 2012)(`--only_assembler 12`), and afterwards retained only those contigs that were more than 500 bp long.

b) Hybrid-assembly

For the combined assembly we used the Illumina assembly that we duplicated to an artificial coverage of 3x to represent long reads, the corrected PacBio reads and the filtered MinION reads. We used Ra (https://github.com/rvaser/ra) with default options as assembler.

c) Scaffolding

The resulting assembly was scaffolded with the SSPACE-LongRead.pl v1.1 (Boetzer & Pirovano 2014) script using MinION and quality-corrected but untrimmed PacBio reads and default options.

M5: Estimating genome size

a) Peak coverage

We used the total number of trimmed nucleotides that were used for the Illumina assembly (38,396,416,216) and divided this by the maximum peak coverage of the per-position coverage frequency distribution (108) as an approximation of genome size. This method assumes an evenly distributed sequencing coverage throughout the genome (Schell *et al*. 2017).

b) Flow cytometry

Genome size (2C-value; Greilhuber *et al*. 2005) was estimated by flow cytometry using 3 individuals of *Crematogaster levior* and the Partec CyFlow Space (Partec, Münster, Germany) equipped with a green solid-state laser (Partec, 532 nm, 30 mW). Sample preparation followed two-step Otto protocol (Otto 1990), with an internal standard *Glycine max* cv. Polanka (2C = 2.50 pg; Doležel *et al.* 1994). Either the whole *C. levior* body or dissected head was mixed with ca. 1 cm^2 leaf of an internal reference standard and homogenized with a razor blade in a Petri dish containing 1 ml of ice-cold Otto I buffer (0.1 M citric acid, 0.5% Tween 20; Otto 1990). The suspension was filtered through a 42-µm nylon mesh and incubated for approximately 15 min at room temperature. The staining solution consisted of 1 ml of Otto II buffer (0.4 M $Na_2HPO_4 \cdot 12\ H_2O$), β-mercaptoethanol (final concentration of 2 µl/ml), intercalating fluorochrome propidium iodide (PI) and RNase IIA (both at final concentrations of 50 µg/ml). Fluorescence intensities of 15,000 particles (nuclei) were recorded for three replicates. Sample/standard ratios were calculated from the means of the sample and standard fluorescence histograms, and only histograms with coefficients of variation <3% for the G0/G1 sample peak were considered. The genome size was calculated by multiplying the sample/standard ratios with the genome size of the internal standard. Additionally, the genome size was corrected by approximately 4.41 %. This is the difference in assigned genome size of human male leukocytes (2C = 7.00 pg = 6,846 Mbp; Doležel *et al*. 1994, 2003) in the original calibration of *G. max* cv. Polanka and the most recent human reference genome assembly (GRCh38.p13; 2C = 6,544.234 Mbp; GenBank assembly accession: GCA_000001405.28).

M6: Annotation

RepeatModeler v1.0.4 (Smit *et al*. 2008) was run with default parameters on the final, scaffolded genome assembly, resulting in a repeat library that was used by MAKER2 (Holt & Yandell 2011) to mask repeat regions in the genome. For annotation, we used the MAKER2 v2.31.8 pipeline, which is not a gene prediction tool as such but a pipeline, that comprises multiple *ab initio* gene prediction tools such as SNAP (Korf 2004), Augustus (Stanke *et al.* 2006) and GeneMark (Lomsadze *et al*. 2005), and uses evidence alignments to find the best suited gene model for a given location.

We first created an Augustus (v3.2.2) species model based on the assembly, ESTs from the same species, and annotations of conserved orthologous genes computed with BUSCO v2.0 (insect dataset, option --long; Simão *et al.* 2015). We also created a SNAP v2006-07-28 (Korf 2004) model based on a CEGMA v2.5 (Parra *et al.* 2007) run on the assembly, by using the cegma2zff script (MAKER2) and the SNAP scripts fathom (fathom genome.ann genome.dna -categorize 1000 && fathom -export 1000 -plus uni.ann uni.dna), forge (export.ann export.dna) and hmm-assembler.pl. GeneMark v4.32 (Lomsadze *et al.* 2005) was used in self-training mode (--ES) on the assembly.

As input for MAKER we used the assembly; the Augustus model; the HMM models from SNAP and Genemark; ESTs from the same species, *C. levior* A; ESTs from *C. levior* B as alternative evidence; the Swiss-Prot/UniProt Database (accessed September 22nd, 2017); annotated proteins from *Cardiocondyla obscurior*; and the repeat library. Protein2genome and est2genome options were switched off and only proteins >30 AA were retained.

After the first iteration, the script gff3_merge was used to obtain a gff3 file and from this we built another Hidden Markov model with SNAP, as specified above. The SNAP model was converted to gff3 format using zff2gff2.pl (genome.ann | perl -plne 's/^t(\S+)$/^t\.\t$1/') and the Augustus model was retrained based on this file. Then MAKER2 was run for a second iteration as specified above, but with updated SNAP and Augustus models. For a second retraining, we repeated those steps and ran MAKER2 for a third time.

To assess completeness of our annotated proteome, we ran DOGMA (Dohmen *et al*. 2016), which uses the Pfam v31 (El-Gebali *et al.* 2019) database of conserved protein domains. In addition, DOGMA was also run on other annotated ant genomes available at the time of the analysis (Table S3.12).

M7: Manual annotation of specific gene families in *Cr. levior*

Elongase (ELO), desaturase (DESAT), odorant binding protein (OBP), gustatory receptor (GR), chemosensory protein (CSP) and odorant receptor (OR) gene families were manually annotated using a two-pass tblastn/Exonerate (Slater & Birney 2005) + GeMoMa (Keilwagen

et al. 2016) + Web Apollo (Lee *et al*. 2013) workflow. In the first round, manually annotated (i.e. high quality) target genes of *Temnothorax longispinosus* (Kaur *et al*. 2019) were blasted against the *Cr. levior* A assembly using tblastn v2.2.30 (e-value=1e-03). Target genes were annotated with Exonerate v2.2.0 (parameter settings: --model protein2genome --percent 50) in those genomic regions with blast hits. In parallel, GeMoMa v1.4.2 was used to annotate target genes, based on the annotation of eight reference ant genomes (Table S3.4) as well as the worker RNAseq libraries of *Cr. levior* A for intron predictions. All protein sequences of the reference species were annotated with PfamScan v31 (Punta *et al*. 2012) and those with the target domains (Table S3.5) were selected for annotation with GeMoMa. GeMoMa annotations were filtered using the GeMoMa Annotation filter (GAF), either retaining only complete gene models or all predictions (i.e. parameter settings: -r 0 -e 0). The Exonerate, GeMoMa (i.e. all predictions) and GAF gene models, as well as the mapped worker RNAseq reads (HISAT2 (Kim *et al.* 2015)) were used as evidence tracks for further manual annotation using Web Apollo v2.0.8. For each gene family, manual annotations in *Cr. levior* and associated orthologs in the reference species were aligned with MAFFT v7.123b (parameter settings: --maxiterate 1000 --localpair). Protein trees were obtained with FastTree (Price *et al*. 2009) (parameter settings: --pseudo). Alignments were visually checked to identify potentially fragmented annotations, which were further curated in Web Apollo. In the second round, Exonerate and GeMoMa were run again, now using the manual annotations from round one as queries. Predictions were used as evidence track in Web Apollo to manually annotate gene models that were missed in the first round. The second round yielded an additional 14 gene models, including 5 ORs, 4 ELOs, 1 OBP, 1 CSP, 1 DESAT and 2 GRs. The manually annotated protein predictions were functionally annotated with PfamScan and only those that contained the defining domains (Table S3.5) were retained (Table S3.6).

Table S3.4: Reference species set, and versions used for GeMoMa annotations

Abbr.	Species	genome.fa	genome.gff
aech	*Acromyrmex echinatior*	Aech_2.0	aech_OGSv3.8
acep	*Atta cephalotes*	Acep_1.0	acep_OGSv1.2
cobs	*Cardiocondyla obscurior*	Cobs_1.4	Cobs_1.4
cflo	*Camponotus floridanus*	Cflo_3.3	cflo_OGSv3.3
hsal	*Harpegnathos saltator*	Hsal_3.3	hsal_OGSv3.3
lhum	*Linepithema humile*	Lhum_1.0	lhum_OGSv1.2
pbar	*Pogonomyrmex barbatus*	Pbar_1.0	pbar_OGSv1.
tlon	*Temnothorax longispinosus*	tlon_1.0	tlon_1.0

Table S3.5: Target gene families and associated Pfam domains used for identification

gene family	abbr.	domain
odorant receptors	OR	7tm_6
gustatory receptors	GR	7tm_7 + Trehalose_recp
elongases	ELO	ELO
desaturase	DESAT	FA_desaturase
odorant binding proteins	OBP	PBP_GOBP
chemosensory proteins	CSP	OS-D

M8: Annotation of elongases and desaturases in other Hymenoptera

We used a reference set of nine high-quality insect genome assemblies (Table S3.7) to annotate elongases and desaturases in the available hymenopteran genomes (Table S3.9). To build the databases for the two gene families, we ran PfamScan on the protein sequences and subsequently filtered the output for those containing the defining ELO and DESAT domains (Table S3.5). The resulting target databases were then used to blast against the remaining 43 annotated Hymenoptera protein sets (Table S3.9) that were not contained within the reference set with blastp v2.5.1 (e-value < 1e-03). On proteins for which blast hits were found, we ran PfamScan to confirm the presence of the target domain. We then filtered the resulting sets of ELO and DESAT genes for identical duplicates to correct for putative assembly mistakes. We tested for significant differences in ELO and DESAT gene number between different families of Hymenoptera with a one-way ANOVA followed by Tukey's HSD, using R v3.5.2 (R Core Team 2018). Results can be seen in Table S3.8.

Table S3.6: Total number of target genes per family

gene family	total count
OR	282
GR	57
ELO	23
DESAT	25
OBP	14
CSP	17

Table S3.7: Reference species set for identification of ELO and DESAT genes

species	genome version	GeneBank Assembly Accession
Apis mellifera	Amel_4.5	GCF_000002195.4
Drosophila melanogaster	Dmel_6plusISO1MT	GCA_000001215.4
Bombus impatiens	Bimp_2.0	GCF_000188095.1
Tribolium castaneum	Tcas_5.2	GCF_000002335.3
Musca domestica	Mdom_2.0.2	GCF_000371365.1
Crematogaster levior	Clev_1.0	ERS3409635
Acromyrmex echinatior	Aech_3.9	GCF_000204515.1
Nasonia vitripennis	Nvit_2.1	GCF_000002325.3
Copidosoma floridanum	Coflo_2.0	GCF_000648655.2

Table S3.8: Results of a one-way ANOVA followed by Tukey's HSD comparing total ELO and DESAT numbers between different families of Hymenoptera. Bold type indicates significance ($p < 0.05$).

		Ant				**Bee**				**Wasp**			
		diff	lwr	upr	padj	diff	lwr	upr	padj	diff	lwr	upr	padj
ELO	**Ant**												
	Bee	-3.64	-8.75	1.47	0.24								
	Wasp	**-7.79**	**-13.7**	**-1.89**	**0.005**	-4.15	10.48	2.18	0.31				
	Sawfly	-6.57	-14.7	1.51	0.15	-2.93	-11.3	5.47	0.79	-1.22	-10.1	7.68	0.98
DES AT	**Ant**												
	Bee	**-5.31**	**-9.08**	**-1.54**	**0.003**								
	Wasp	-1.33	-5.69	3.03	0.85	3.98	-0.7	8.65	0.12				
	Sawfly	1.58	-4.38	7.55	0.89	**6.89**	**0.69**	**13.1**	**0.024**	-2.92	-9.49	3.66	0.64

Table S3.9: Overview of the different protein sets used in ELO and DESAT comparison. Species marked with an asterisk were used as reference species

Family	Species	abbr.	genome version	#ELO	#DESAT
ant	*Acromyrmex echinatior**	aech	Aech_3.9	20	18
ant	*Atta cephalotes*	acep	Acep_1.0	18	11
ant	*Atta colombica*	acol	Acol_1.0	12	12
ant	*Camponotus floridanus*	cflo	Caflo_1.0	20	13
ant	*Cyphomyrmex costatus*	ccos	Ccos_1.0	13	16
ant	*Cardiocondyla obscurior*	cobs	Cobs_1.4	12	6
ant	*Crematogaster levior**	clev	Clev_1.0	23	25
ant	*Dinoponera quadriceps*	dqua	Dqua_1.0	20	11
ant	*Harpegnathos saltator*	hsal	Hsal_1.0	19	14
ant	*Lasius niger*	lnig	Lnig_1.0	9	9
ant	*Linepithema humile*	lhum	Lhum_04	33	15
ant	*Monomorium pharaonis*	mpha	Mpha_2.0	20	12
ant	*Ooceraea biroi*	obir	Obir_1.0	14	8
ant	*Pogonomyrmex barbatus*	pbar	Pbar_03	17	12
ant	*Pseudomyrmex gracilis*	pgra	Pgra_1.0	36	25
ant	*Solenopsis invicta*	sinv	Sinv_Si_gnG	29	15
ant	*Trachymyrmex cornetzi*	tcor	Tcor_1.0	22	14
ant	*Trachymyrmex septentrionalis*	tsep	Tsep_1.0	15	14
ant	*Trachymyrmex zeteki*	tzet	Tzet_1.0	18	11
ant	*Vollenhovia emeryi*	veme	Veme_1.0	37	17
ant	*Wasmannia auropunctata*	waur	Waur_1.0	25	9
bee	*Apis cerana*	acer	Acer_2.0	17	7
bee	*Apis dorsata*	ador	Ador_1.3	13	11
bee	*Apis florea*	aflo	Aflo_1.0	17	8
bee	*Apis mellifera**	amel	Amel_4.5	17	7
bee	*Bombus impatiens**	bimp	Bimp_2.0	21	9
bee	*Bombus terrestris*	bter	Bter_1.0	21	7
bee	*Ceratina calcarata*	ccal	Ccal_1.0	20	6
bee	*Dufourea novaengliae*	dnov	Dnov_1.0	15	9
bee	*Eufriesea mexicana*	emex	Emex_1.0	16	10
bee	*Euglossa dilemma*	edil	Edil_1.0	15	5
bee	*Habropoda laboriosa*	hlab	Hlab_1.0	13	6
bee	*Lasioglossum albipes*	lalb	Lalb_5.4	16	12
bee	*Melipona quadrifasciata*	mqua	Mqua_1.0	19	4
bee	*Megachile rotundata*	mrot	Mrot_1.0	17	16
wasp	*Copidosoma floridanum**	coflo	Coflo_2.0	12	18
wasp	*Diachasma alloeum*	dall	Dall_1.0	13	9
wasp	*Fopius arisanus*	fari	Fari_1.0	12	12
wasp	*Microplitis demolitor*	mdem	Mdem_2.0	13	12
wasp	*Nasonia vitripennis**	nvit	Nvit_2.1	13	18
wasp	*Polistes canadensis*	pcan	Pcan_1.0	14	9
wasp	*Polistes dominula*	pdom	Pdom_r1.2	15	9
wasp	*Trichogramma pretiosum*	tpre	Tpre_2.0	14	13
wasp	*Trichomalopsis sarcophagae*	tsar	Tsar_1.0	9	11
sawfly	*Athalia rosae*	aros	Aros_1.0	10	17
sawfly	*Cephus cinctus*	ccin	Ccin_1.0	16	13
sawfly	*Neodiprion lecontei*	nlec	Nlec_1.0	18	20
sawfly	*Orussus abietinus*	oabi	Oabi_2.0	12	11

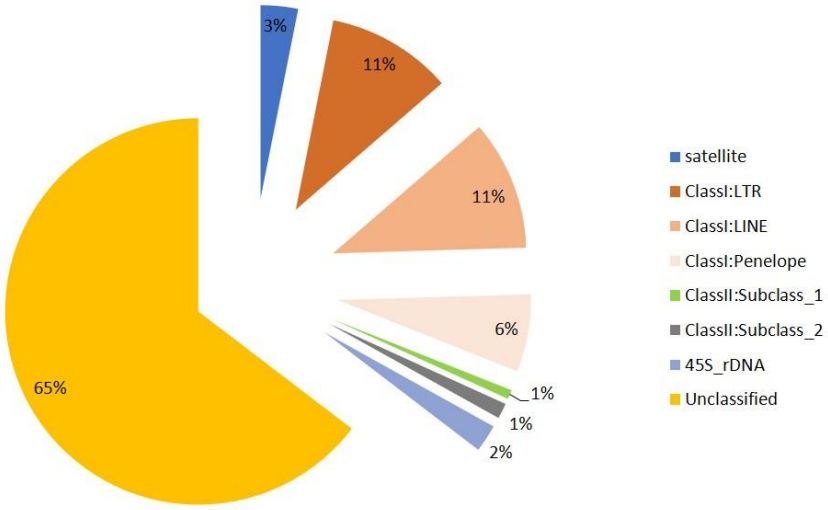


Figure S3.1: Proportion of different repeat classes in the *C. levior* genome

Table S3.10: Backmapping rates of different read sets on the final assembly

Read type	Percentage of mapped reads
Illumina reads	97.31%
Illumina contigs	76.90%
PacBio reads	97.49%
PacBio reads corrected	96.27%
MinION reads	96.89%

Table S3.11: Percentage of the genome that is only covered by one type of sequencing technique

Read type	Percentage of genome
Illumina contigs	1.05%
PacBio	2.33%
PacBio uncorrected	1.31%
MinION	2.42%

Table S3.12: Overview of annotated ant genomes and according assembly statistics. Assembly statistics were analysed using own summary pipeline. Values might therefore differ from respective publications. For analysed genome versions refer to Table S3.9.

Species	Estimated Genome size [Mbp]	Assembly Length [Mbp]	# Scaffolds	N50	# Genes	# Dogma CDAs	Dogma [%]	Repeats [%] of assembly	Publication	Source Genome size
Acromyrmex echinatior	335.0	295.9	4,339	1,110,580	17,278	5945	93.30%	27.70%	Nygaard *et al.* 2011	Gadau *et al.* 2012
Atta cephalotes	303.18	317.7	2,835	5,154,485	18,093	5606	88.00%	25.00%	Suen *et al.* 2011	Gadau *et al.* 2012
Atta colombica	303.0	291.3	1,550	2,037,154	14,345	6265	98.30%	31.80%	Nygaard *et al.* 2016	Tsutsui *et al.* 2008
Camponotus floridanus	303 - 323	232.7	10,791	451,320	17,064	6028	94.60%	15.10%	Bonasio *et al.* 2010	Gadau *et al.* 2012
Cardiocondyla obscurior		177.9	1,854	3,105,814	17,552	6084	95.50%		Schrader *et al.* 2014	
Crematogaster levior A	409.96	326.2	1,523	383,244	17,855	5652	88.70%	10.50%		
Ooceraea biroi		212.8	4,579	1,350,650	17,263	6019	94.50%	13.80%	Oxley *et al.* 2014	
Cyphomyrmex costatus	318.5	300. 3	15,379	1,159,032	16,468	6230	97.80%	34.00%	Nygaard *et al.* 2016	Nygaard *et al.* 2016
Dinoponera quadriceps		259,7	14,123	1,361,239	13,688	6269	98.40%	6.00%	Patalano *et al.* 2015	
Harpegnathos saltator	330.00	294.5	8,893	601,965	18,564	5940	93.20%	26.90%	Bonasio *et al.* 2010	Gadau *et al.* 2012
Lasius niger		236.2	36,804	17,057	19,989	4920	77.20%	1.40%	Konorov *et al.* 2017	
Linepithema humile	254.28	219.5	3,030	1,402,257	16,123	5914	92.80%	16.50%	Smith *et al.* 2011	Gadau *et al.* 2012
Monomorium pharaonis		258.0	12,136	75,377	13,616	6261	98.20%		Mikheyev & Linksvayer 2015	
Pogonomyrmex barbatus	245 - 280	235.6	4,645	819,605	17,177	5711	89.60%	11.60%	Chris R Smith *et al.* 2011	Chris R Smith *et al.* 2011
Pseudomyrmex gracilis	391.20	282.8	6,556	317,681	16,069	6264	98.30%	41.00%	Rubin *et al.* 2016	Ardila-Garcia *et al.* 2010
Solenopsis invicta	463 - 753.3	399.0	66,904	621,039	16,569	4146	65.10%	23.00%	Wurm *et al.* 2011	Gadau *et al.* 2012
Trachymyrmex cornetzi	396.1	369.4	19,761	760,749	19,827	6242	97.90%	42.10%	Nygaard *et al.* 2016	Nygaard *et al.* 2016
Trachymyrmex septentrionalis	294.4	291.7	5,836	2,520,094	15,575	6288	98.70%	23.90%	Nygaard *et al.* 2016	Nygaard *et al.* 2016
Trachymyrmex zeteki	294.8	268.0	4,623	1,333,945	15,530	6240	97.90%	24.90%	Nygaard *et al.* 2016	Nygaard *et al.* 2016
Vollenhovia emeryi		287.9	13,258	1,346,088	14,870	6283	98.60%		Mikheyev & Linksvayer 2015	
Wasmannia auropunctata	324.12		77,788	1,175,369	15,458	6241	97.90%			

CHAPTER 4

Divergent adaptational trajectories despite similar selection regimes among sister species

Juliane Hartke, Ann-Marie Waldvogel, Philipp P. Sprenger, Thomas Schmitt, Florian Menzel, Markus Pfenninger, Barbara Feldmeyer

Submitted

ABSTRACT

Species living in sympatry and sharing a similar niche often realise parallel phenotypes as a response to the same selection pressures. Whether this phenotypic parallelism is also reflected at underlying levels of biological organization, i.e. nucleotides to functions, is widely debated. We use multi-dimensional genomic associations to assess the basis of local and climate adaptation in two sympatric, morphologically cryptic ant species along a tropical climate gradient. In addition, we investigated the genomic basis of chemical communication in both species. Communication in insects proceeds mainly mediated by cuticular hydrocarbons (CHCs), which additionally protect against water loss and can change plastically in response to environmental cues such as climatic conditions. The combination of environmental and chemical association analyses based on genome-wide Pool-Seq data allowed us to identify SNPs associated with climate as well as differences in CHC profiles. Within species, changes in CHC profiles seem to be driven by phenotypic plasticity, since there is no overlap between climate and CHC associated SNPs. The only exception is the odorant receptor OR22c, which may be a prime candidate for population specific CHC recognition in one of the species. Both species show significant association to climate in their CHC profiles as well as on the level of population differentiation, but a between species comparison revealed highly divergent genomic adaptation patterns. The associated candidates are species-specific and there is no evidence for parallelism neither on nucleotide nor functional level. This highlights that closely related species may follow divergent evolutionary trajectories when realizing similar adaptive phenotypes.

Keywords: EAA, population divergence, BayPass, mutualism, Formicidae, parallel evolution

INTRODUCTION

The degree to which evolution is predictable has been widely debated among evolutionary biologists (Gould 1989; Beatty 2006; Morris 2010). If the outcome of ecological selection pressures was predictable, we would repeatedly witness the same phenotypical and genomic patterns in nature. Indeed, many cases of parallel evolution of phenotypical traits are well represented in the literature (e.g. Hume & Martill 2019; Oke *et al.* 2017; Therkildsen *et al.* 2019). Whether this is an accurate representation of natural systems or rather the exception to the rule is hard to determine due to potential reporting bias. In general, the likelihood of parallel evolution largely depends on the level of biological hierarchy, decreasing from overall phenotype to organ structure, pathways, genes down to the nucleotide level (Bailey *et al.* 2015). Parallel evolution seems to be more likely when the compared groups are subject to exactly the same selection regime and the more these groups approach their fitness optimum, leaving fewer possible beneficial mutations (Orr 2005; Bailey *et al.* 2015). The likelihood for parallel evolution furthermore seems to increase with the strength of natural selection and for genes that have large phenotypic effects (MacPherson & Nuismer 2017). Intuitively, parallel evolution should be most likely to occur under completely identical natural selection (see Thompson *et al.* 2019) and under the premise of shared evolutionary histories. Studies from natural systems so far showed several instances, in which species or populations showed a substantial overlap in patterns of genomic changes (Chaturvedi *et al.* 2018; Haenel *et al.* 2019), while in several other instances there was evidence for largely idiosyncratic evolution (Wittkopp *et al.* 2003; Kaeuffer *et al.* 2011; Pfenninger *et al.* 2015; Feldmeyer *et al*. 2017). However, these studies suffer from one or both of the following shortcomings: firstly, if the studied species lacked a common direct ancestor, the potential for parallel evolution was diminished due to different evolutionary histories. Furthermore, if the focal species or populations inhabited different locations, potential site-specific selection regimes could further constrain parallel evolution. Predicting the outcome of selection pressures acting on species has never been of greater importance than in the light of anthropogenic habitat destructions and climate warming and will greatly aid in conservation efforts (Waldvogel *et al.* 2019). One promising approach is to compare phenotypic responses of closely related sympatric species to truly parallel selection pressures and analyse the genomic underpinnings of resulting phenotypes.

Amongst the most important local selection pressures are temperature and precipitation that ultimately shape virtually all abiotic and biotic conditions (Scheffers *et al.* 2016). Both factors can vary within short geographical distances leading to pronounced climatic clines. Species therefore often encounter gradually differing climatic selection pressures across their distribution ranges, which drives climate adaptation of local populations, maximising their fitness under local climate conditions. The importance of local climate adaptation has

been recognized in the past years and their genomic underpinnings has been studied in plants (Turner *et al.* 2010; Rúa *et al.* 2016), especially crops (Mercer *et al*. 2008; Pyhäjärvi *et al.* 2013), and in animals as for example some birds (Slabbekoorn & Peet 2003; Dreiss *et al.* 2012; Charmantier *et al.* 2016) and in several insect species (e.g. Pringle *et al.* 2012; Bergland *et al*. 2016; Waldvogel *et al*. 2018). Especially the latter, being ectotherms, are particularly vulnerable to variation in ambient temperature, which directly influences essentially all functions from locomotion to metabolism. In insects, an important part of the acclimatory response to climate variation concerns changes in the cuticular hydrocarbons (CHCs) profile. CHCs make up the main part of the epicuticular layer of terrestrial arthropods. It is hypothesized that they originally evolved as waterproofing agents and they were shown to be plastic within populations in response to different temperature and humidity levels (Menzel *et al.* 2018; Otte *et al*. 2018; Sprenger *et al*. 2018). Heritable differences were also shown between natural subpopulations as a response to parameters that covary with latitude and season (Frentiu & Chenoweth 2010; Rajpurohit *et al*. 2017). However, CHCs in insects not only serve in desiccation resistance, but are also used in communication and are highly important in species recognition (Blomquist & Bagnères 2010). Some species can plastically change their CHC profiles in response to social cues, e.g. acoustic stimuli or rivalry (Thomas & Simmons 2011), and courtship (Petfield *et al*. 2005). This dual function of CHCs, as a short-term and long-term adaptive response to ecological selection pressures, as well as communication signals, makes it challenging to precisely determine the cause for differences in CHC profiles and untangle possible environmental adaptation from their role in communication and species recognition (Sprenger & Menzel 2020). This is especially true in social insects, in which communication via CHCs is highly elaborate and allows the discrimination between species, colonies and castes (Pamminger *et al.* 2014; Leonhardt *et al.* 2016), facilitates task allocation (Greene & Gordon 2003) and thus enables their sophisticated social structure.

We present the neotropical cryptic ant species *Crematogaster levior* A and *C. levior* B as a model system to test parallelism in response to local climate selection (Hartke *et al.* 2019a). Colonies of these cryptic sister species occur in sympatry (often within a few metres) over a large part of their distribution range and are thus subject to the same climatic selection pressures in these areas (Hartke *et al.* 2019a; Sprenger *et al.* 2019). The two species live mutualistically with either one of two cryptic species of *Camponotus femoratus* (*C. femoratus* PAT and *C. femoratus* PS), without any obvious partner preference (Hartke *et al*. 2019a). Each of the two cryptic species pairs shows striking differences in their cuticular hydrocarbon profiles (Menzel *et al.* 2017; Hartke *et al.* 2019a; Sprenger *et al.* 2019), which might be indicative for differential adaptation to climate (Gibbs *et al.* 1997), or differential selection on species recognition e.g. mate-choice (Thomas & Simmons 2009). In French Guiana all four species largely share the same habitat. This system is therefore the ideal

model to test parallelism in response to local selection pressures in two closely related species.

We conduct population genomic analyses (Pool-Seq) with *C. levior* A and B and (i) use a F_{ST} outlier approach to identify signatures of local adaptation among conspecific populations; (ii) conduct an environmental association analysis (EAA) to investigate the influence of varying selection pressures, such as temperature, precipitation and vegetation cover, on population differentiation; and (iii) conduct a chemical association analysis (CAA) to find patterns of genomic differentiation that correlate with differences in the populations' cuticular hydrocarbon profiles. Finally, we compare these results between the two species to identify differences and similarities in adaptive response and ultimately to determine whether local adaptation follows a parallel pattern or species-specific trajectories.

MATERIAL AND METHODS

Sampling and Sequencing

Individuals used for pooled sequencing were collected in French Guiana during September and October 2016. For both *C. levior* A and C. *levior* B, 15 colonies of five populations each were chosen (Figure 4.1A). Within each population, all colonies were at least 30m apart (for detailed locations see Supplement Table S4.1). Per population, the heads of 105 individuals were pooled and sequenced using 2x150 bp paired-end Illumina sequencing. The resulting data were quality processed and mapped on the *C. levior* A reference genome (Hartke *et al*. 2019b). The mapped reads were filtered for duplicates and all bases below a minimum quality of 20 were discarded (for more details on DNA isolation, sequencing and downstream processing: Supplement 1.1).

Population differentiation and local adaptation

We followed the PoPoolation2 v1201 pipeline for further downstream processing (Kofler *et al*. 2011; Supplement 1.1). Pairwise F_{ST} was calculated in non-overlapping 1kb windows between all populations as a measure of population differentiation using the *fst-sliding.pl* script integrated in PoPoolation2. We conducted Mantel tests with 9999 permutations between genome-wide mean F_{ST} values and geographical distance to test for genome-wide isolation by distance patterns. Based on 10.000 randomly drawn SNPs, we performed a principal component analysis (PCA) with pcadapt v4.0.0 (Luu *et al.* 2017) to visualize the population structure of *C. levior* A and B. To this end, observed allele frequencies at each locus were used to simulate 20 individuals, following the BAYENV approach (Günther & Coop 2013).

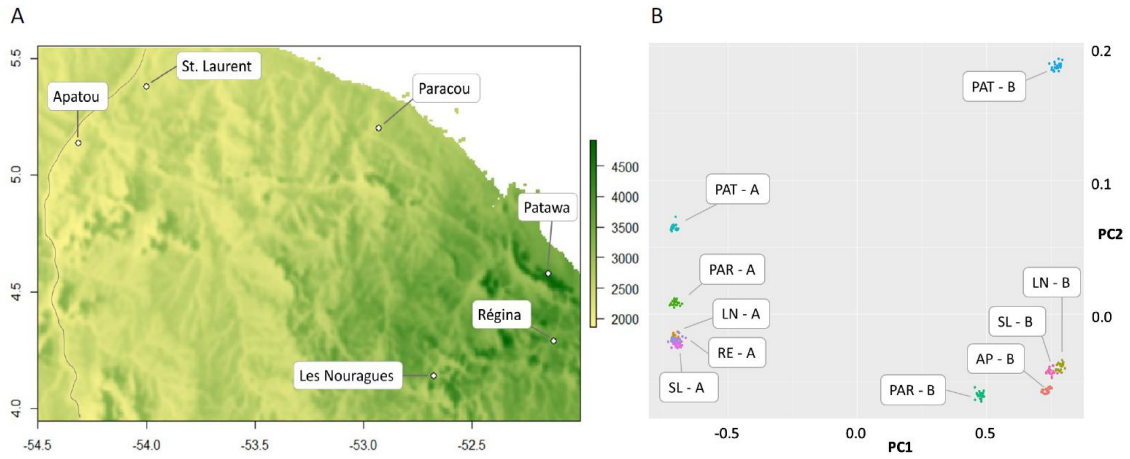


Figure 4.1: **(A)** Overview of sampling sites. Map coloration refers to mean annual precipitation (CHELSA Bioclim 1979-2013). **(B)** Allele frequency PCA computed from 1,000,000 randomly sampled genome-wide SNPs between all sequenced populations. The letters after the location codes refer to the species *C. levior* A and B respectively.

BayPass was used to detect significant outliers in allele frequencies among our sampled populations within each of the two species. BayPass allows for a robust identification of differentiated markers across populations, while correcting for demographic effects, which proved advantageous compared to similar tools based on a Bayesian framework (Gautier 2015). BayPass first computes the allele frequency covariance matrix Ω that gives information on the shared demographic histories of the populations. Following the suggestion of Gautier (2015) we used the R function *simulate.baypass()* to simulate a pseudo-observed dataset consisting of 250.000 SNPs based on the covariance matrix. BayPass calculated the F_{ST}-like parameter XtX that accounts for shared population history (Günther & Coop 2013). The upper 1% quantile threshold of the simulated dataset (*C. levior* A: XtX > 10.19; *C. levior* B: XtX > 11.01) was used to define outliers in our empirical datasets which may reflect signatures of local adaptation.

Environmental data for association analysis

We calculated a PCA with *factoextra* v1.0.5.999 (Kassambara & Mundt 2016) for each species using the CHELSA Bioclim variables (monthly climate data for the years 1979 - 2013; Karger *et al.* 2017) in combination with the MODIS datasets of the year 2016 for Evapotranspiration (Running *et al.* 2017), Thermal Anomalies (Giglio & Justice 2015) and Vegetation Indices (Didan 2015). The first three PC axes that explain >90% of the variance were used as input for the environmental association analysis (EAA) with BayPass v2.1 (Gautier 2015). As additional input the MODIS dataset Vegetation Continuous Fields was used (Dimiceli *et al.* 2015; for detailed description of environmental data see Supplement 1.2).

Chemical data for association analysis

Extraction and GC-MS analysis of cuticular hydrocarbons of the two *Crematogaster* species is explained in detail in Hartke *et al*. (2019a). For the chemical association analysis, we calculated the mean relative abundance of each hydrocarbon for the 15 colonies used per pool. We then again calculated a PCA for each species on this dataset and used the first four PC axes that explain 100% of the variance as input for BayPass. We additionally computed a PCA for the chemical dataset that included both *C. levior* A and B, to visualize the highly differentiated cuticular hydrocarbon profiles of the two species. In this PCA only those substances that were present in both species were included. We furthermore tested whether population specific CHC profiles correlated with the PC axes of the environmental PCA with a PERMANOVA (function: *adonis*, package: *vegan*, 999 permutations).

Genome-wide association analysis

BayPass was used according to standard protocol (Gautier 2015) to identify SNPs whose allele frequencies were significantly associated with population specific parameters based on the SNP frequency differences inferred with PoPoolation2. As population specific parameters, we used the environmental dataset and for a second analysis, the chemical dataset. As a threshold to identify significant associations, we followed Jeffreys' rule (Jeffreys 1961) and only considered SNPs with BF > 20 as decisive.

Downstream analysis

For all significantly associated SNPs and for XtX outliers that were found in or nearby (< 2000 bp) genes, we ran BLASTp and InterProScan (Quevillon *et al.* 2005) on the genes' protein sequence. We furthermore checked for gene function using UniProt based on gene names found with BLAST. To gain a more comprehensive overview of outlier functions, we summarized the GO terms of outliers within and nearby genes with REVIGO (Supek *et al.* 2011), using the *Drosophila* database. We also performed a gene enrichment analysis with the R package topGO v2.34.0 (Alexa & Rahnenführer 2016) using the GO terms of all *Cr. levior* genes as a reference. GO terms were retained with a significance level cut-off of $p < 0.05$.

RESULTS

Population Differentiation and outlier analysis

We generally found very low F_{ST} among all population-pairs within *C. levior* A and B respectively. In *C. levior* A, mean pairwise F_{ST} ranged from 0.043 to 0.057 (mean = 0.047; Table 1, Supplement S1A). In *C. levior* B, albeit being higher than in A, pairwise F_{ST} were

still low with values between 0.057 and 0.082 (mean = 0.066; Table 4.1, Supplement S4.1B), indicating high connectivity between populations within French Guiana and very low levels of population differentiation. This is also reflected in the PCA based on a randomly drawn subset of SNPs (Figure 4.1B). *C. levior* A is generally more uniform than *C. levior* B (see also Hartke *et al.* 2019a). The two species are clearly separated by the first PC, while intraspecific allele frequency variation is mainly reflected on PC2. Actual geographical structure of both species is not mirrored within the PCA, and also the Mantel tests revealed no sign for isolation by distance in *C. levior* A ($p = 0.258$, $r = 0.24$), while in *C. levior* B we found a weak trend ($p = 0.083$, $r = 0.35$). This suggests a stronger population structure within *C. levior* B, which is also reflected in slightly higher pairwise F_{ST} values.

Table 1: Mean pairwise F_{ST} between populations, calculated in 1 kb windows. Values of pairwise comparisons of *C. levior* A populations are shown in the lower (bold) part, values for *C. levior* B populations in the upper part (italic).

	LN	*AP*	*PAR*	*PAT*	*SL*	*Cr-B*
LN	-	*0.062*	*0.064*	*0.063*	*0.056*	*LN*
RE	**0.041**	-	*0.074*	*0.076*	*0.054*	*AP*
PAR	**0.045**	**0.045**	-	*0.077*	*0.067*	*PAR*
PAT	**0.051**	**0.051**	**0.050**	-	*0.07*	*PAT*
SL	**0.046**	**0.045**	**0.047**	**0.056**	-	*SL*
Cr-A	**LN**	**RE**	**PAR**	**PAT**	**SL**	

BayPass calculates XtX as measure of allele frequency differentiation between populations, which one can use to discriminate between selection and neutrality and thus deduct significant outliers that indicate signatures of local adaptation (Gautier 2015). We found 114 outlier SNPs in *C. levior* A and 34 genes, in or nearby which outliers were located. Outlier analysis in *C. levior* B yielded more SNPs (N = 399) and more genes (N = 168). One of the genes was found in both species (*reverse transcriptase*). GO term analysis of outliers found between populations and located in or nearby genes, yielded 10 biological processes in *C. levior* A and 56 in B, respectively (Supplement Table S4.2). Of those, 5 and 18 were significantly enriched in *C. levior* A (e.g. sensory perception of sound and light stimulus) and B (e.g. humoral immune response and glycerol ether metabolic process) respectively. None of the terms were found in both species (note that one gene was shared between species, however, this gene lacks GO information).

Principal Component Analyses of population specific parameters

In *C. levior* A and B, the PCAs with environmental factors resulted in three PC axes that together explained 95.3% and 93.8% of the variance respectively. The first PC axis in both cases correlated positively to precipitation, and negatively to temperature (Figure 4.2A, B).

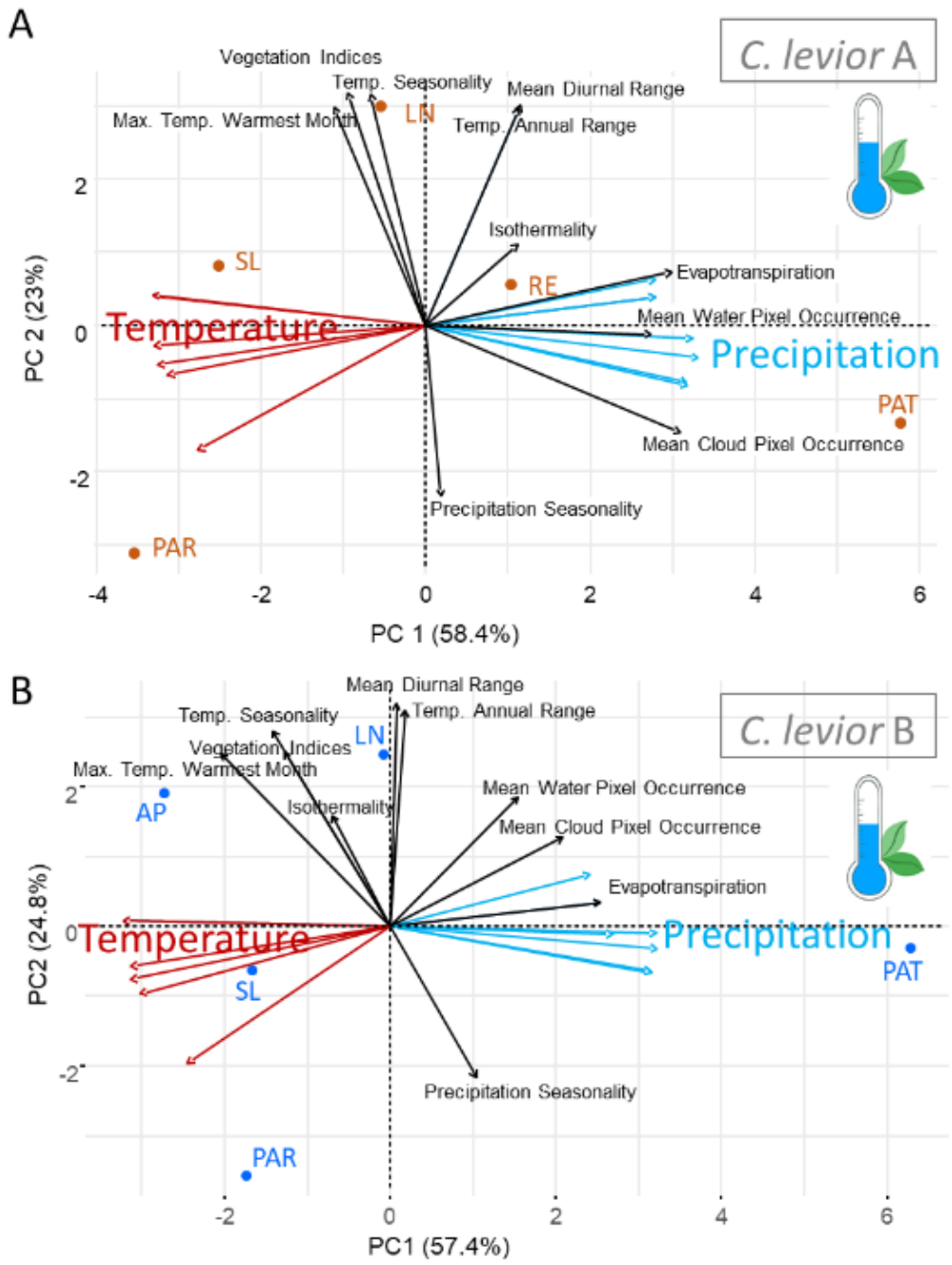


Figure 4.2: Principal component analysis of environmental and climate parameters in 5 populations of **(A)** *C. levior* A and **(B)** *C. levior* B. Not shown is PC3 that explained 13.9% and 11.6% of the variance in *C. levior* A and B respectively. For clarity reasons overlapping temperature and precipation arrows are indicated in red and blue respectively without detailed description.

The colonies from Patawa were separated from the other populations by higher precipitation and lower temperature, and Paracou was separated from the other populations by a higher precipitation seasonality, both of which is also visible when looking at an overview of the populations' temperature and precipitation parameters (Fig. S4.2). The

PCAs based on the cuticular hydrocarbon profiles of colonies resulted in four PCs that explained 100% of the variance in both species (Fig. S4.3). A combined PCA based on CHC profiles for both *C. levior* A and B (Figure 4.3) shows a clear separation of the species (see also Hartke *et al*. 2019a) and reflects the highly differentiated cuticular hydrocarbon profiles (Sprenger *et al*. 2019). The first principal component explains 59.6% of the variation and separates *C. levior* A from *C. levior* B. The hydrocarbons associated with this axis largely consist of long-chain, unsaturated CHCs typical for *C. levior* A and the shorter, mostly saturated ones typical for *C. levior* B. While the populations belonging to *C. levior* A for the most part cluster together, *C. levior* B shows more diversity (F-test, p = 0.046, F = 0.099). Paracou-B is clearly separated from the other populations. This is visible by a separation on the second PC axis, which is correlated with the abundance of a C29-alkadiene (Fig. 4.3). Climate-wise, the location Paracou is shaped by rather high temperatures throughout the year and high precipitation seasonality, with almost no precipitation during the dry season (Fig. S4.2).

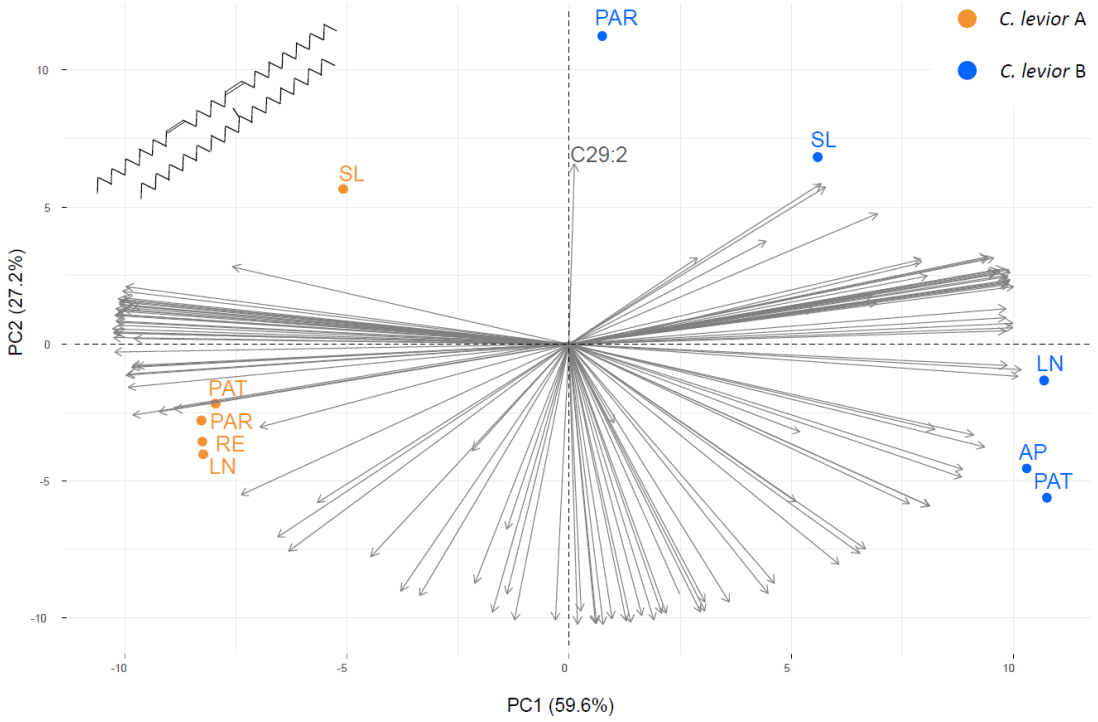


Figure 4.3: Principal Component Analysis of population-specific substances in cuticular hydrocarbon profiles of *C. levior* A and B. Note that the arrow representing C29:2 (nonacosadiene) is the strongest one positively associated with the Paracou-B population. Grey arrows indicate the axis association of CHC compounds.

Environmental Association Analysis (EAA)

We found 122 SNPs that were significantly associated (BF > 20) with environmental parameters in *C. levior* A (Figure 4.4A; Fig. S4.4). Of those, 43 were found in genes (Table 4.2). 13 SNPs are privately associated with the first environmental PC axis, which is mostly

explained by an inverse correlation of precipitation and temperature (Figure 4.5A). PC2 is correlated with temperature seasonality and annual range and precipitation seasonality and associated with 14 private SNPs. Most of the associated SNPs are found to be correlated with PC3 with 88 private SNPs. The parameters that make up most of the variation of this axis are isothermality inversely correlated with precipitation and temperature seasonality. We furthermore identified 2 SNPs that were solely associated with the percentage of non-

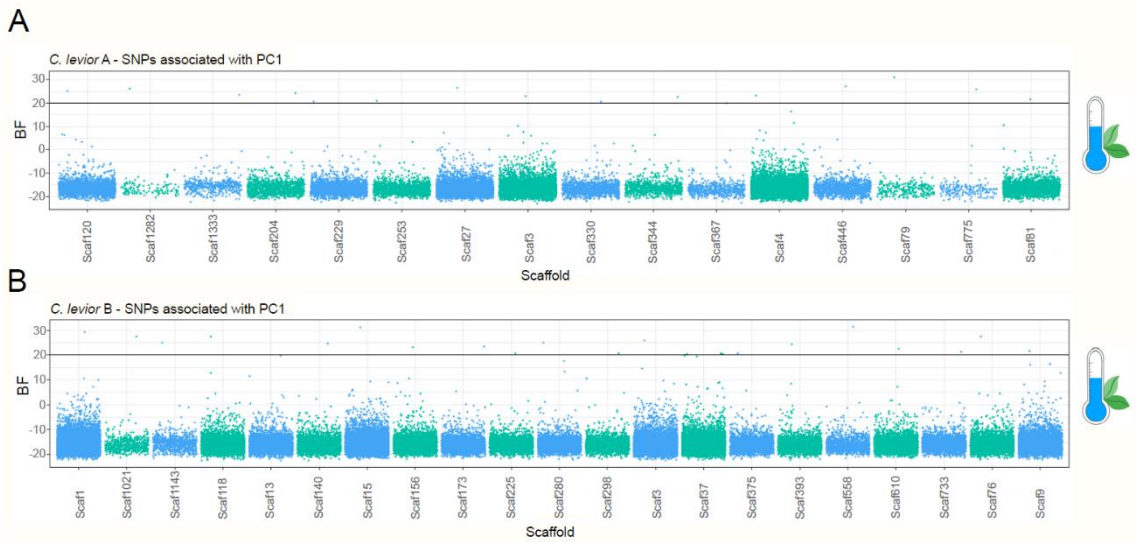


Figure 4.4: Overview of SNPs that are significantly correlated (Bayes Factor >20) with PC1 (see Fig. 4.2) in the environmental association studies in **(A)** *C. levior* A and **(B)** *C. levior* B. Only scaffolds that contain significantly associated SNPs are shown. X-axis does not represent sequence length.

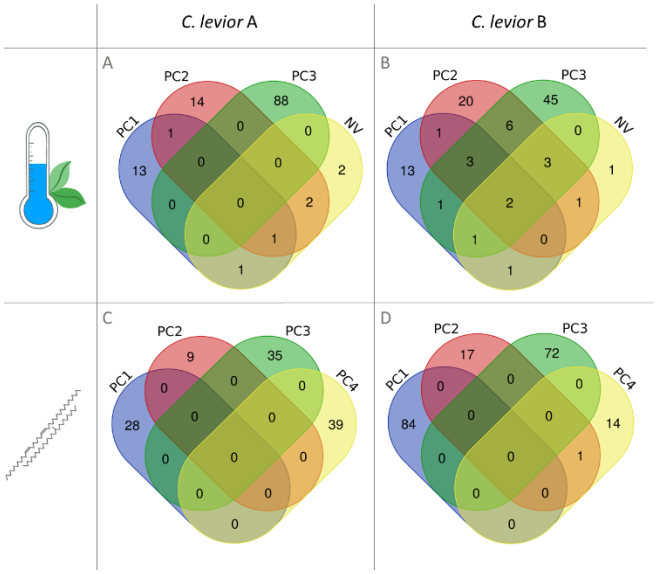


Figure 4.5: Number of significantly associated SNPs of (**A,B**) environmental and (**C,D**) chemical association analyses within (**A,C**) *C. levior* A and (**B,D**) *C. levior* B. For both analyses, PC1-PC4 refer to PC axes from Principal Component Analyses. NV refers to the percentage of non-vegetated landscape and was not

vegetated landcover (Figure 4.5A). We found no association with the percentage of tree cover or non-tree cover. Most of the SNPs were private to one of the environmental parameters, and only few were shared (117 SNPs vs. 5).

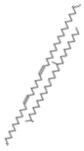
In *C. levior* B, we found fewer SNPs (N = 98, of those 52 in genes (Table 4.2) (Figure 4.4B, Fig. S4.5) associated with environmental parameters compared to *C. levior* A. The PC axes can be described similarly as above. PC1 is privately associated with 13 SNPs (Figure 4.5B), PC2 with 20 and PC3 with 45. The percentage of non-vegetated landcover is privately associated 1 SNP (Figure 4.5B).

Like in *C. levior* A, we found no association with the percentage of tree cover or non-tree cover. We performed one GO term analysis for all significantly associated SNPs that were found in genes for each species. Six GO terms were significantly enriched for *C. levior* A and five for *C. levior* B (Table 4.3). Functions significantly enriched in *C. levior* A included "DNA replication initiation" and "histone H3-K27 demethylation" and in *C. levior* B "chitin metabolic process".

Chemical Association Analysis (CAA)

The chemical association analysis in general yielded less overlap between PC axes in both species compared to the environmental association analysis (Figure 4.5C, D). In *C. levior* A we found 111 SNPs, 51 of which were located in genes. PC1 was privately associated with 28 SNPs, PC2 with 9, PC3 with 35 and PC4 with 39 SNPs (Figure 4.5C, Fig. S4.6). In *C. levior* B, we found 188 SNPs (72 in genes), 84 of which were privately associated with PC1, 17 with PC2, 72 with PC3 and 14 with PC4 (Figure 4.5D, Fig. S4.7).

Table 4.2: Excerpt of genes that are significantly associated with either environment or cuticular hydrocarbon profile in *C. levior* A and *C. levior* B.

Analysis	Species	Gene	Function
 *Environmental association*	*C. levior* A	Unknown Odorant receptor	Olfactory perception
		Odorant receptor OR22c	Olfactory perception
		Dunce	Olfactory learning
		Apolipophorins	Fatty acid transport
	C. levior B	Nuclear pore complex	Chitin metabolic process
		Scavenger receptor class B	Chitin metabolic process
		Lamin Dm0	Adult locomotory behavior, cell aging
 *Chemical association*	*C. levior* A	Odorant receptor OR22c	Sensory perception of smell
		Unknown Odorant receptor	Sensory perception of smell
		Gld	Cuticular modifications during development
	C. levior B	Tricorner	Antennal development
		Fatty acyl-CoA reductase	Wax biosynthetic process
		Futsch	Olfactory learning
		Odorant receptor OR67c	Sensory perception of smell

A GO term analysis encompassing all SNPs in genes yielded five significantly enriched terms in *C. levior* A (e.g. sensory perception of smell) and six terms in *C. levior* B (e.g. glycogen related processes) (Table 4.4). Among the genes that contained significantly associated SNPs in *C. levior* A (Table 4.2), two are odorant receptors and are involved in the olfactory perception of cuticular hydrocarbons. Interestingly, one of the odorant receptors (OR22c) was also significantly correlated with PC3 in the EAA, although here, a different SNP within the gene was associated.

Table 4.3: GO terms that were significantly enriched in the environmental association analysis in *C. levior* A and *C. levior* B. For complete list of GO terms implicated in environmental association see Supplement Table S4.2.

***C. levior* A**	
PC1	DNA replication initiation histone H3-K27 demethylation
PC3	ammonium transmembrane transport phosphatidylinositol-mediated signaling vesicle docking involved in exocytosis phosphatidylinositol phosphorylation
***C. levior* B**	
PC1	chitin metabolic process
PC2	homophilic cell adhesion via plasma membrane adhesion molecules response to metal ion transmembrane receptor protein tyrosine kinase signaling pathway
PC3	ATP hydrolysis coupled proton transport

In *C. levior* B (Table 4.2), we found two genes involved in the synthesis of CHCs and the insect cuticle (*serine/threonine-protein kinase tricorner*, *fatty acyl-CoA reductase*), and two genes putatively involved in the perception of CHCs (*futsch*, *odorant receptor 67c*). The significant association of the SNP within *fatty acyl-CoA reductase* is explained by its divergent allele frequency within the Paracou population compared to a fixed allele in all other populations. Interestingly, in the combined chemical PCA Paracou is also distinctly diverged from the other populations, which is explained by a single chemical component, C29-diene. We checked whether the SNP we found within Paracou represents a non-synonymous substitution, and indeed this SNP changes isoleucine to valine. Also the SNP within *tricorner*, a gene that is associated with antennal morphology, only occurs within the population of Paracou.

Table 4.4: GO terms that were significantly enriched in the chemical association analysis in *C. levior* A and *C. levior* B. For complete list of GO terms implicated in chemical association see Supplement Table S4.3.

***C. levior* A**	
PC1	sensory perception of smell ATP hydrolysis coupled proton transport
PC3	neurotransmitter transport deoxyribonucleotide biosynthetic process phosphatidylinositol dephosphorylation
***C. levior* B**	
PC1	glycogen metabolic process mitochondrial translation posttranslational protein targeting to membrane, translocation
PC3	glycogen catabolic process glycogen biosynthetic process dsRNA transport

Multi-dimensional genomic association and species comparison

The EAA revealed genes and associated GO terms that are significantly correlated to environmental/climate variation across the distribution range of the two species. Nevertheless, signatures of local climate adaptation did not overlap between the two species since we neither found shared significantly associated SNPs, nor genes nor GO terms. This is consistent with the result of the CAA that revealed the species-specific association of genes and GO terms to CHC profiles without overlap between species. Nevertheless, we found the variability of CHC profiles to be significantly correlated to environmental variation across the geographic distance of the sampling area both within *C. levior* A (PERMANOVA; PC1: F = 2.776, p = 0.019; PC2: F = 0.0697, p = 0.0589; PC3: F = 0.584, p = 0.692; PC4: F = 4.068, p = 0.005) and within *C. levior* B (PERMANOVA; PC1: F = 2.241, p = 0.039; PC2: F = 3.179, p = 0.011; PC3: F = 3.051, p = 0.01; PC4: F = 2.093, p = 0.068). Multi-dimensional genomic association finally revealed the odorant receptor (OR22c) as prime candidate to be significantly associated to environmental variation as well as variation in CHC profiles in *C. levior* A.

DISCUSSION

In this study we conducted a population genomic analysis with two closely related parabiotic ant species *C. levior* A and B. The two species are morphologically very similar, share the same habitat and show no difference in the choice of their mutualistic partners (Hartke *et al.* 2019a), but show striking differences in their CHC profiles (Menzel *et al.* 2014; Menzel *et al.* 2017; Sprenger *et al.* 2019). We correlated different environmental parameters and the populations' specific CHC profiles to allele frequency differences within each species to untangle the genomic basis of local and climate adaptation and identify potential signatures of parallel evolution.

Population Structure and Local adaptation

Each of the two species shows rather high levels of connectivity and gene flow among populations, reflected in low pairwise F_{ST} and no indication for isolation by distance, which are ideal settings to identify signatures of local adaptation without confounding demographic effects (Hoban *et al*. 2016). French Guiana is crossed by several large rivers that do not seem to pose any kind of natural barrier for the species, as population differentiation is rather homogeneous across the sampling range. Interestingly, in South East Asia, the opposite pattern was found. There, *Crematogaster* species show substantial population divergence even on a very small spatial scale (Feldhaar *et al.* 2010), which could

be indicative for species-specific dispersal strategies that differ between South East Asian and South American *Crematogaster*.

Testing for local adaptation based on allele frequency differences among populations, showed much higher differentiated SNPs among *C. levior* B populations compared to *C. levior* A populations. Most of the SNPs, genes and also GO functions with a role in local adaptation were specific to each species – with the exception of the gene *reverse transcriptase* – which shows that even though the two species adapt to the same local selection regime, the genomic footprint of adaptation involves species-specific genes and pathways.

Climate Structure and Environmental Association

The east-west transect across French Guiana coincides with a temperature and precipitation gradient. This was reflected in the climate PCAs, with the populations of Saint Laurent and Patawa at different ends of the gradient (Fig. 4.2). Saint Laurent is characterized by higher temperatures and low precipitation, while Patawa represents the coldest and wettest of our sampling locations. Interestingly, in the PCA based on randomly drawn SNPs, Patawa is also distinctly separated from the other populations in *C. levior* B, and also in *C. levior* A, although to a lesser degree. This might indicate population-specific adaptation to a cooler and wetter climate that is reflected within the genome of both species. The fact that this is visible on a genomic level is not surprising in the light of a recent study that found roughly 1% of annotated genes to be involved in clinal climate adaptation, and another 8% in local adaptation (Waldvogel *et al.* 2018).

In *C. levior* A, we found several genes that seem to have a function in the detection of CHCs to be implicated with climate variation. Two of those genes are odorant receptors, and one gene, *dunce*, is responsible for olfactory learning in *Drosophila melanogaster* (Walkinshaw *et al*. 2015). The effects of climate variation on olfaction seem to be twofold. On the one hand, ambient temperature increases the concentration of volatiles (including cuticular hydrocarbons, Menzel *et al.* 2019) in the air, but also influences an organism's sensitivity towards those substances, with cooler temperatures allowing a better perception of olfactory cues (Riveron *et al.* 2009). On the other hand, environmental factors can cause a change in the CHC profile (Menzel *et al.* 2017b; Sprenger *et al.* 2018) that in turn can lead to a need for changes in olfaction. Interestingly, we found one of those odorant receptors (*OR22c*) to also be associated with population specific variance of CHC profiles. In the case of this gene, the second scenario thus seems to be more plausible. Climate variation caused a shift in CHC profiles that in turn caused selection on olfactory perception. Indeed, odorant receptors seem to be highly specialized in ants and are able to discriminate even between intra-colonial quantitative differences in CHC profiles (Pask *et al.* 2017; Slone *et al*. 2017).

Deeper investigations will be necessary to assess the actual role in climate adaptation of this prime candidate.

In *C. levior B*, we found two genes associated with intraspecific environmental variation that have functions in chitin metabolism. This pathway is associated with temperature adaptation in *D. melanogaster* (MacMillan *et al.* 2016) and the parasitoid wasp *Aphidius colemani* (Clark & Worland 2008), and might thus also play a role in temperature adaptation in *C. levior* B.

Chemical Association

Cuticular hydrocarbons are the foundation to the extraordinary social structure of social insects and furthermore facilitate acclimation and adaptation to temperature and humidity (Frentiu & Chenoweth 2010; Sprenger *et al.* 2018). In line with the importance of CHCs in social insects, hymenopterans as such and ants in particular, possess a high number of CHC synthesis genes (Hartke *et al.* 2019b). It is thus of great interest to identify the genomic underpinnings of CHC evolution with a focus on intraspecific variation within profiles. We show here the first genome-wide cuticular hydrocarbon association analysis and show that this method is indeed suited to find candidate genes underlying CHC diversity in social insects. In *C. levior* A, we found two odorant receptors significantly associated with population-specific variation in CHC profiles. One of them was also associated with environmental parameters. As discussed above, a scenario in which climate parameters caused a shift in CHC profiles that in turn caused selection on olfactory perception seems the likely explanation for this pattern. Another gene that was significantly associated with CHC diversity, *Gld*, serves in cuticular modifications during development (Cavener *et al*. 1986).

In *C. levior* B, a SNP within the gene *fatty acyl-CoA reductase* was significantly associated with population-specific differences in CHC profiles. This can be explained by the introduction of a mutation in the population of Paracou, which is separated from the other populations in the chemical PCA with both species by a higher abundance of C29-diene. Fittingly, in *D. melanogaster*, *fatty acyl-CoA reductases* are responsible for intraspecific CHC variation and have furthermore been shown to influence the synthesis of C29-diene (Dembeck *et al.* 2015). We may therefore cautiously conclude that also in *C. levior* B, this specific *fatty acyl-CoA reductase* may be involved in the synthesis of C29-diene and that a mutation in this gene caused a shift in the cuticular hydrocarbon profile within the *C. levior* B population of Paracou. Due to the dual role of cuticular hydrocarbons, it is hard to say whether this shift in allele-frequency was the result of environmental adaptation to a site-specific factor, or the product of e.g. sexual selection that is confined to Paracou. Also

within Paracou, we found a shift in allele frequency in the gene *tricorner* that is responsible for antennal morphology (He & Adler 2002). As antennae are the structures containing olfactory receptors that are important for CHC detection (Schneider 1964; d'Ettorre *et al.* 2004; Lihoreau & Rivault 2009), this finding might imply that changes in the CHC profile may have caused a change in the olfactory perception, as in *C. levior* A. In *C. levior* B, we furthermore found three different GO terms that are involved in the breakdown and synthesis of glycogen to be significantly enriched in the chemical association analysis (CAA). Glycogen is the storage form of glucose, which is vital for several biological functions, for example for the synthesis of the chitin layer. Increased glycogen storage is also used as an adaptation to heat and drought (Marron *et al*. 2003; Parkash *et al.* 2013, 2014; Reidenbach *et al.* 2014). It could thus be, that the same selection pressure, e.g. high temperature, that led to a change in hydrocarbon profiles, also led to changes in the synthesis and breakdown in glycogen to increase resistance against water loss.

Identification of parallel evolution and climate associated CHCs

To identify patterns of parallel evolution as well as putative CHC candidates involved in climate adaptation, we checked whether any overlap of SNPs or genes could be found between species or analyses. We found one gene, a reverse transcriptase, to be significantly differentiated among populations of both species. Whether this indicates parallel local adaptation is hard to tell, because in eukaryotic genomes, reverse transcriptases belong to the enzymatic machinery of retrotransposons as well as retroviruses (Baltimore 1970; Temin & Mizutani 1970). The overlapping signal of significant differentiation in this reverse transcriptase could thus result from a shared (local) retrovirus infection in both species (Svarovskaia *et al.* 2003), ancestral polymorphisms (Han *et al.* 2017) or, and least plausible, a signature of parallel local adaptation. Nevertheless, in other ants, mobile elements have been implicated in adaptation to novel environments (Schrader *et al*. 2014).

We found no overlap in SNPs or genes between the two species in either association analysis. Also, a GO enrichment analysis yielded no overlap between the two species in significantly enriched GO terms within both chemical and environmental association analyses. One of the GO terms that we found most often was 'sensory perception of smell'. Even though this term was only significantly enriched in the CAA in *C. levior* A, we also found genes associated to this term in the CAA in *C. levior* B, and in the EAA in *C. levior* A, which is further underlining the importance of CHCs in climate adaptation and the following changes in CHC perception.

Studies that modelled the probability of genome-wide parallel evolutionary changes predicted high likelihoods under completely parallel natural selection (Thompson *et al.*

2019). Similar selection pressures were mirrored in parallel genomic changes in other systems such as stickle backs (e.g. Colosimo *et al.* 2005; Haenel *et al.* 2019), *Tinema* stick insects (Rivas *et al.* 2018), *Lycaeides* butterflies (Chaturvedi *et al.* 2018) and rails (Hume & Martill 2019). *C. levior* A and *C. levior* B co-occur throughout their distribution range in French Guiana and are therefore subject to the same local and climate selection pressures. They furthermore share the same direct ancestor and thus the same evolutionary background. Still, we found no overlap between species neither in SNPs, genes nor GO terms that are implicated in local or climate adaptation. This is in line with other studies which did not find evidence for parallelism but species-specific trajectories (Dennis *et al.* 2015; Pfenninger *et al.* 2015; Feldmeyer *et al.* 2017). In our system, adaptation to the same ecological and climate factors seem to be facilitated by different genes and pathways that lead to the same adaptive value. A pattern, that has also been found in *Drosophila* as a response to temperature (Barghi *et al.* 2019) or high altitude (Bastide *et al*. 2016). Indeed, parallel evolution, i.e. the same mutational changes on SNP level, are only to be expected under certain evolutionary constraints and when only few possible beneficial mutations are available (Orr 2005; Bailey *et al.* 2015).

Within species, we found a single gene overlapping between both environment and chemical association analyses in *C. levior* A, the *odorant receptor 22c*. In *Drosophila melanogaster*, this receptor is specifically expressed in larvae (Störtkuhl & Fiala 2011). In social insects, odorant receptors are generally responsible for the perception of cuticular hydrocarbons (Brand *et al*. 2015; Pask *et al.* 2017; Slone *et al.* 2017). This might therefore indicate that populations within *C. levior* A experience different environmental selection pressures, that affect the CHC profile as a response to climate adaptation, and that in turn triggered a change in the genes responsible for the perception of CHCs. Even though colony specific CHC profiles correlate with environment, the genomic footprints underlying CHCs and environmental adaptation do not overlap except for this one odorant receptor. Reasons for this might be either that CHC profiles are plastic and thus possess no obvious population genomic signatures or that the overlapping signatures are lost in functional redundancy.

CONCLUSION

This is the first genome-wide association study in ants and, to our knowledge, the first study linking cuticular hydrocarbon diversity to allele frequency differences. In a multi-dimensional genomic association approach, we identified genomic signatures of local and climate adaptation, as well as potential candidate genes underlying CHC synthesis and detection in the two cryptic ant species *C. levior* A and *C. levior* B. Many of the genes that are

associated with variation in CHC profiles have already been implicated in serving functions in the detection or synthesis of CHCs and can thus be taken as proof of principle that CHC association analyses can indeed yield interesting candidate genes underlying chemical communication. Even though both species co-occur at most sites across the study area, we neither found the same patterns of environmental selection, nor did we find the same genes implicated with population specific shifts in cuticular hydrocarbon profiles. *C. levior* A and *C. levior* B are sympatric sister species and thus subject to the same environmental and climatic selection pressures. While parallel genomic changes as a response to the same selective pressures have been found in other systems, here, they are completely absent. Parallel adaptation in the two species thus seems to be facilitated by different genes that in the end serve the same adaptive function. Using a multi-dimensional approach, this study contributes to the growing evidence for an often rather idiosyncratically acting evolution (Kaeuffer *et al.* 2011; Pfenninger *et al*. 2015; Feldmeyer *et al.* 2017; Goedbloed *et al*. 2017).

AUTHOR'S CONTRIBUTIONS

JH, BF, FM, AMW and MP conceived the study. JH generated and analysed the data. JH, BF, AMW, PS, FM, TS, MP drafted the manuscript.

ACKNOWLEDGEMENTS

We thank [redacted] for his helpful advice regarding climate data analysis. This work was supported by the German Science Foundation (DFG) as a grant to BF. (FE 1333/7-1), TS. (SCHM2645/7-1), and FM. (ME 3842/5-1). This work was additionally sup- ported by the LOEWE Centre Translational Biodiversity Genomics (TBG), funded by the state of Hesse, Germany. This study is compliant with the Nagoya protocol (permission number: TREL1734890A/13).

SUPPLEMENTARY MATERIAL

EXTENDED METHODS

Sequencing and downstream processing

From 15 colonies the heads of seven individuals each were taken for pooled DNA Isolation using the Qiagen DNeasy Blood and Tissue kit (QIAGEN, Hilden, Germany) according to spin column protocol. Final elution was done in 80µl of Buffer. Pooled 2x150 bp paired-end sequencing was conducted by Genewiz (Genewiz, UK Ltd, England) on an Illumina HiSeq with an insert size of 300 bp. The resulting reads were trimmed with Trimmomatic (Bolger *et al.* 2014) used with the autotrim script (Waldvogel *et al.* 2018) and mapped to the reference genome using the *bwa mem* algorithm (v. 0.7.15; Li & Durbin 2009) with a seed length of 30. Duplicates were filtered with picard tools (https://github.com/broadinstitute/picard) and the minimum quality limit for bases was set to 20 (samtools). The resulting filtered *bam* files were then summarized to a single *sync* file for each of the species, containing 5 Pools, with samtools mpileup (Li *et al*. 2009) and the *mpileup2sync* script from PoPoolation2 (Kofler *et al.* 2011). Further processing steps included filtering of indel regions (*filter-sync-by-gtf.pl*) and subsampling to a uniform coverage of 20x (*subsample-syncronized.pl*), before SNPs were identified between the 5 Pools of each species (*snp-frequency-diff.pl*).

Environmental data - PCA

One part of the input for BayPass (Gautier 2015) consisted of the CHELSA Bioclim (Karger *et al*. 2017) variables that were obtained for our specific sample locations. We furthermore used MODIS datasets for Evapotranspiration (MOD16A2; Running *et al.* 2017), Thermal Anomalies (MOD14A2; Giglio & Justice 2015) and Vegetation Indices (MOD13A1; Didan 2015) of the year 2016. For Evapotranspiration, we calculated the mean for the year 2016 of all available datasets of this period. For Vegetation Indices we used datasets for four time points in 2016 (01.01.2016, 01.04.2016, 01.08.2016, 01.12.2016) and then calculated the mean. For Vegetation Continuous Fields (MOD44B; Dimiceli *et al*. 2015), we obtained a single value for each location split into "Tree Cover", "Non-Tree Cover" and "Non-vegetated". Except for this dataset all other CHELSA and MODIS variables were combined into a single dataset and a PCA was calculated with factoextra v1.0.5.999 (Kassambara & Mundt 2016) of which we used the scores of the first three axes that explained 95.3% and 93.8% of the variance in *C. levior* A and B respectively as input for the following outlier association with BayPass.

Table S4.1: Exact locations of colonies used for Sequencing. Last letters in Colony name refer to sampling locations. SL = Saint Laurent; PAR = Paracou; LN = Les Nouragues; PAT = Patawa; RE = Régina; AP = Apatou

Crematogaster levior **A**			*Crematogaster levior* **B**		
Colony	**Latitude**	**Longitude**	**Colony**	**Latitude**	**Longitude**
20_SL	5.456567	-54.010867	16_SL	5.448417	-53.994250
21_SL	5.456217	-54.011200	17_SL	5.447067	-53.992600
320_SL	5.452283	-54.013617	22_SL	5.455700	-54.011633
323_SL	5.448833	-54.014083	23_SL	5.455417	-54.012217
325_SL	5.484800	-54.001550	319_SL	5.454033	-54.012883
327_SL	5.483883	-54.001483	322_SL	5.449083	-54.013633
328_SL	5.483400	-54.001517	326_SL	5.484417	-54.001517
332_SL	5.482450	-53.995900	329_SL	5.482267	-54.002083
335_SL	5.484367	-53.995667	330_SL	5.481133	-54.001867
336_SL	5.484450	-53.995283	331_SL	5.481083	-54.000983
337_SL	5.484750	-53.990867	333_SL	5.483150	-53.995500
338_SL	5.484733	-53.990283	334_SL	5.484033	-53.996133
339_SL	5.485000	-53.990150	344_SL	5.482117	-53.998750
340_SL	5.483117	-53.987167	346_SL	5.478500	-53.998283
342_SL	5.480450	-53.987317	347_SL	5.476650	-53.997800
46_PAR	5.276367	-52.926433	47_PAR	5.275683	-52.929000
48_PAR	5.272250	-52.932167	57_PAR	5.258017	-52.927333
49_PAR	5.270850	-52.929517	59_PAR	5.255733	-52.926950
51_PAR	5.268283	-52.926850	61_PAR	5.255917	-52.925850
54_PAR	5.274767	-52.924050	62_PAR	5.260517	-52.929050
55_PAR	5.261700	-52.927933	68_PAR	5.258817	-52.932267
73_PAR	5.258783	-52.936317	69_PAR	5.258383	-52.933033
368_PAR	5.279650	-52.923783	72_PAR	5.258233	-52.935967
369_PAR	5.280550	-52.922067	76_PAR	5.261583	-52.932433
370_PAR	5.279600	-52.922783	380_PAR	5.267267	-52.935167
371_PAR	5.278133	-52.924433	383_PAR	5.262100	-52.929767
373_PAR	5.270567	-52.933100	385_PAR	5.261850	-52.930067
381_PAR	5.269667	-52.935550	120_LN	4.042267	-52.674900
387_PAR	5.259433	-52.927000	122_LN	4.040350	-52.674867
388_PAR	5.259033	-52.926917	124_LN	4.039467	-52.674517
136_LN	4.049083	-52.663800	126_LN	4.040517	-52.675567
143_LN	4.041600	-52.660900	129_LN	4.041683	-52.676467
145_LN	4.040683	-52.661850	130_LN	4.048200	-52.662650
148_LN	4.039833	-52.664433	131_LN	4.051100	-52.662650
149_LN	4.039550	-52.665183	132_LN	4.051033	-52.663317
158_LN	4.038283	-52.670550	133_LN	4.050633	-52.663250
168_LN	4.030717	-52.680700	140_LN	4.048267	-52.662283
169_LN	4.033383	-52.680017	152_LN	4.039050	-52.667200
178_LN	4.066533	-52.685517	156_LN	4.039000	-52.668100
179_LN	4.066633	-52.685717	177_LN	4.063900	-52.683467
180_LN	4.067083	-52.686050	182_LN	4.063850	-52.681450

192_LN	4.043767	-52.684217	190_LN	4.042367	-52.683200
193_LN	4.044133	-52.686267	218_PAT	4.545050	-52.133183
194_LN	4.043733	-52.687250	219_PAT	4.545033	-52.132433
198_LN	4.037267	-52.672200	221_PAT	4.544883	-52.131550
217_PAT	4.545767	-52.134733	222_PAT	4.544783	-52.131200
233_PAT	4.550083	-52.147500	223_PAT	4.544917	-52.130517
234_PAT	4.550483	-52.147350	226_PAT	4.545733	-52.130717
235_PAT	4.550900	-52.148017	227_PAT	4.546067	-52.130483
236_PAT	4.552467	-52.150333	228_PAT	4.545917	-52.130000
237_PAT	4.552767	-52.150417	230_PAT	4.526200	-52.118450
238_PAT	4.553383	-52.150550	245_PAT	4.555333	-52.152750
240_PAT	4.554017	-52.150667	348_PAT	4.557150	-52.154783
242_PAT	4.554683	-52.151083	359_PAT	4.549550	-52.130617
244_PAT	4.555300	-52.152083	360_PAT	4.558317	-52.185767
352_PAT	4.560000	-52.155267	361_PAT	4.558733	-52.186417
354_PAT	4.541667	-52.135400	366_PAT	4.553333	-52.143867
362_PAT	4.548483	-52.145517	27_AP	5.164933	-54.328917
363_PAT	4.549083	-52.145233	28_AP	5.164433	-54.327183
367_PAT	4.555400	-52.144900	33_AP	5.139133	-54.325917
200_RE	4.226833	-52.122033	34_AP	5.138950	-54.325533
201_RE	4.222267	-52.121633	37_AP	5.202183	-54.311350
202_RE	4.222750	-52.121450	38_AP	5.209550	-54.308933
203_RE	4.223000	-52.121133	40_AP	5.226633	-54.299533
204_RE	4.223167	-52.120683	41_AP	5.226367	-54.298983
205_RE	4.226550	-52.117200	42_AP	5.226100	-54.298650
206_RE	4.226833	-52.117217	43_AP	5.226900	-54.298883
207_RE	4.230400	-52.117400	44_AP	5.239200	-54.284217
208_RE	4.231950	-52.116650			
209_RE	4.213683	-52.120117			
210_RE	4.213833	-52.119583			
211_RE	4.214683	-52.119117			
213_RE	4.214600	-52.117833			
214_RE	4.184567	-52.135467			
215_RE	4.167233	-52.126100			

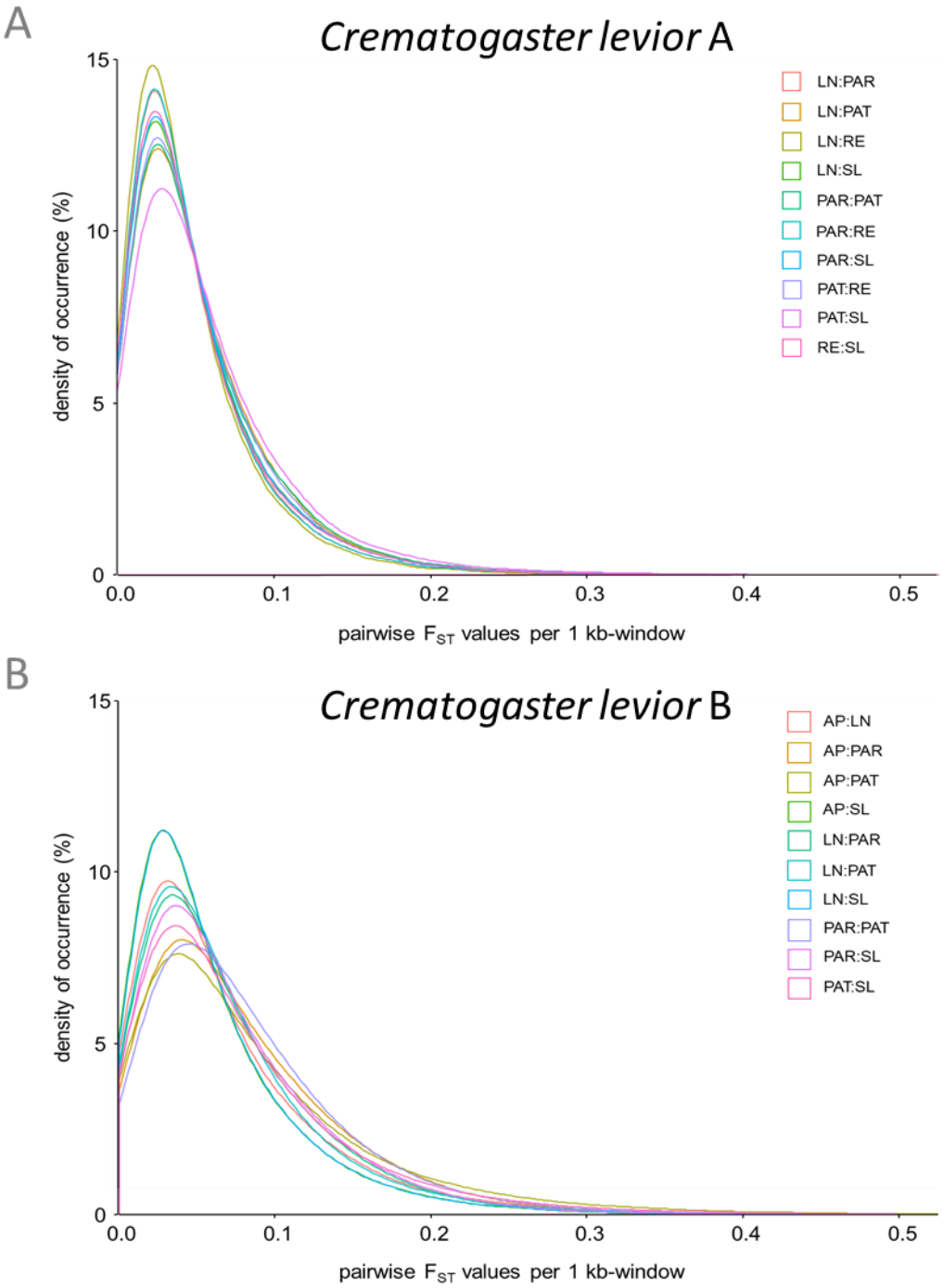


Figure S4.1: Comparison of pairwise FST values between **(A)** *C. levior* A and **(B)** *C. levior* B populations, calculated in 1 kb sliding windows

Table S4.2: Summary of GO functions of genes implicated in local adaptation in *C. levior* A and B. Significantly enriched GO terms indicated in bold.

***C. levior* A**	***C. levior* B**
N-acetylglucosamine metabolic process	**humoral immune response**
sensory perception of light stimulus	**establishment of planar polarity of embryonic epithelium**
translational initiation	**clathrin coat assembly**
exocytosis	**microtubule severing**
sensory perception of sound	**glycerol ether metabolic process**
intracellular protein transport	**carbohydrate phosphorylation**
G-protein coupled receptor signaling pathway	**lipoprotein metabolic process**
protein dephosphorylation	**protein sumoylation**
vesicle-mediated transport	**establishment or maintenance of actin cytoskeleton polarity**
signal transduction	**tRNA thio-modification**
	glycerol-3-phosphate catabolic process
	transmembrane receptor protein tyrosine kinase signaling pathway
	galactose metabolic process
	threonyl-tRNA aminoacylation
	protein phosphorylation
	protein localization involved in establishment of planar polarity
	negative regulation of Wnt signaling pathway
	ATP hydrolysis coupled proton transport
	carbohydrate metabolic process
	protein folding
	homophilic cell adhesion via plasma membrane adhesion molecules
	sensory perception of smell
	metabolic process
	protein transport
	regulation of GTPase activity
	cell redox homeostasis
	gene silencing by RNA
	DNA replication
	regulation of Rho protein signal transduction
	multicellular organism development
	tRNA wobble uridine modification
	cellular lipid metabolic process
	glycerol-3-phosphate metabolic process
	calcium ion transmembrane transport
	potassium ion transmembrane transport
	translation
	proteolysis
	oxidation-reduction process
	positive regulation of transcription, DNA-templated
	regulation of transcription, DNA-templated
	signal transduction
	transmembrane transport
	microtubule-based movement
	tRNA aminoacylation for protein translation
	tRNA aminoacylation
	protein peptidyl-prolyl isomerization
	regulation of actin filament polymerization
	Arp2/3 complex-mediated actin nucleation
	actin filament polymerization
	intracellular signal transduction
	Wnt signaling pathway
	lipid transport
	rRNA processing
	GPI anchor biosynthetic process
	ATP synthesis coupled proton transport
	protein glycosylation

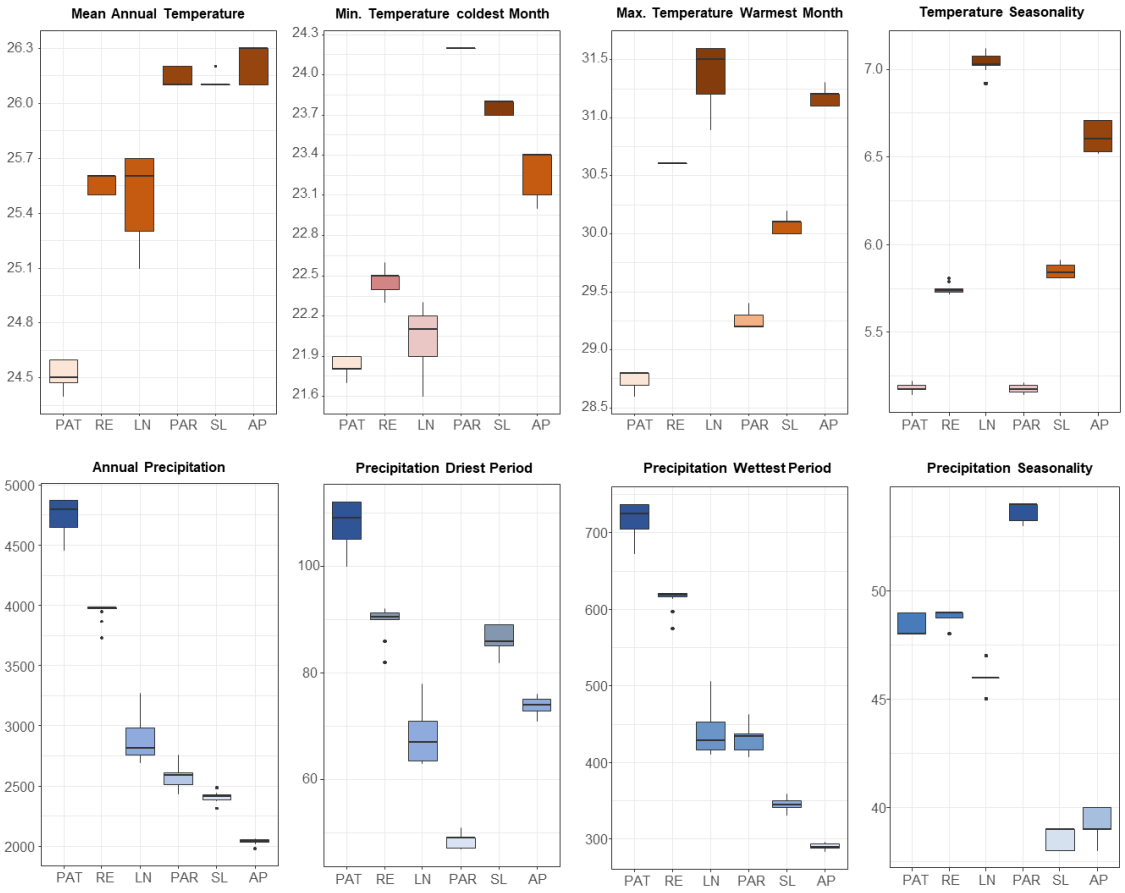


Figure S4.2: Climate parameters for the different sampling locations during the year 2016.

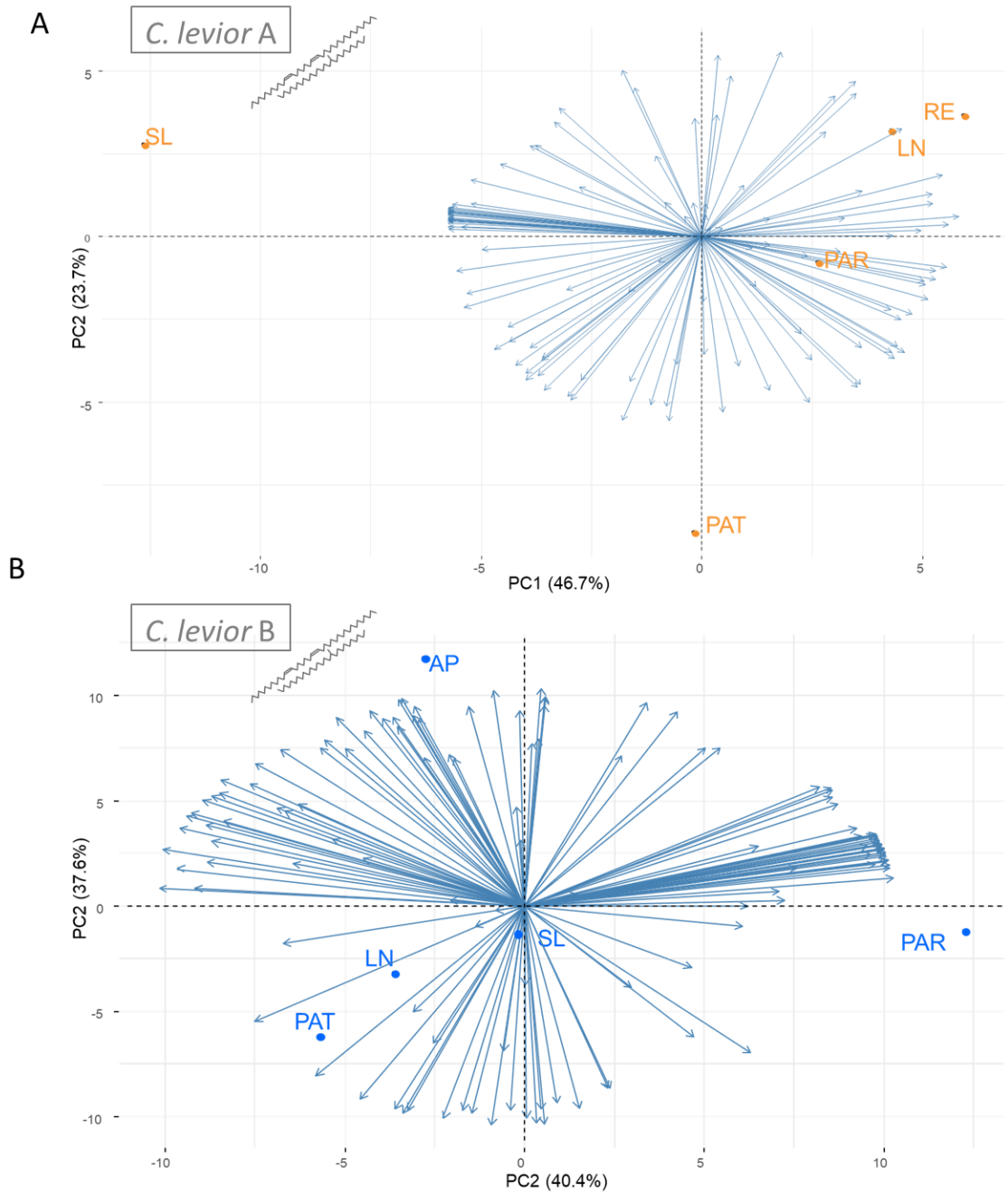


Figure S4.3: Principal Component Analysis of population-specific cuticular hydrocarbon profiles of **(A)** *C. levior* A and **(B)** *C. levior* B. PC3 and PC4 are not shown. They explain 22% and 7% in *C. levior* A, and 12.7% and 9.1% in *C. levior* B.

Figure S4.4: Overview of SNPs identified with BayPass that are significantly correlated (Bayes Factor >20) with (**A**) PC2, (**B**) PC3 and (**C**) the percentage of non-vegetated landcover in the environmental association studies in *C. levior* A.

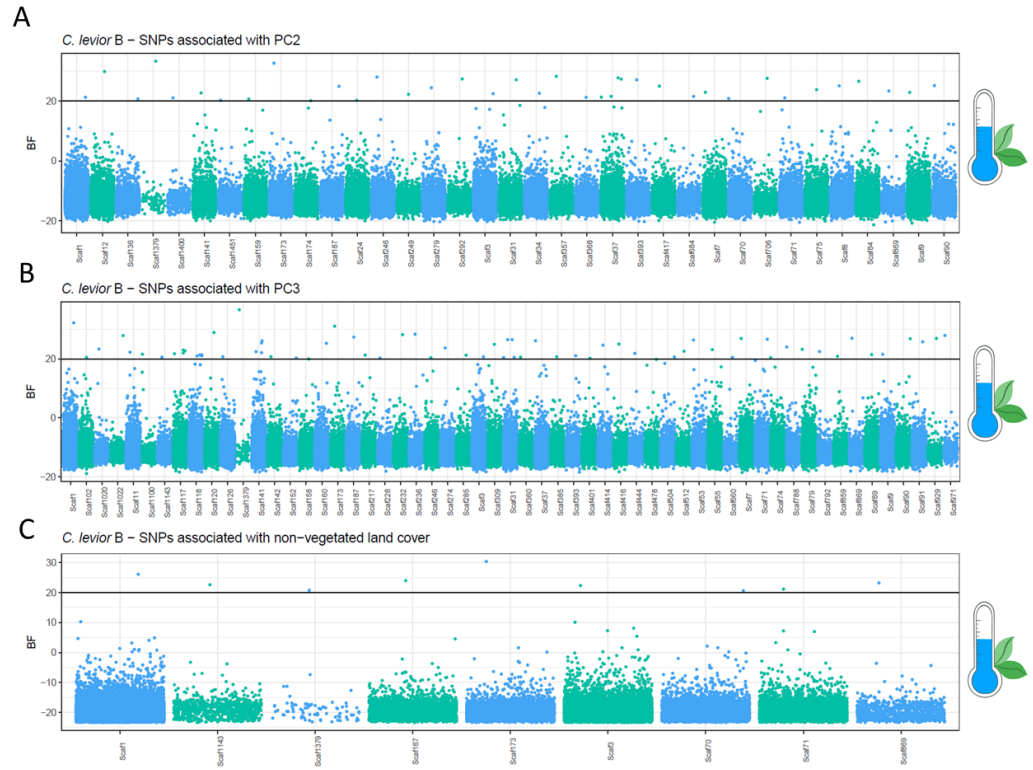


Figure S4.5: Overview of SNPs identified with BayPass that are significantly correlated (Bayes Factor >20) with **(A)** PC2, **(B)** PC3 and **(C)** the percentage of non-vegetated landcover in the environmental association study in *C. levior* B.

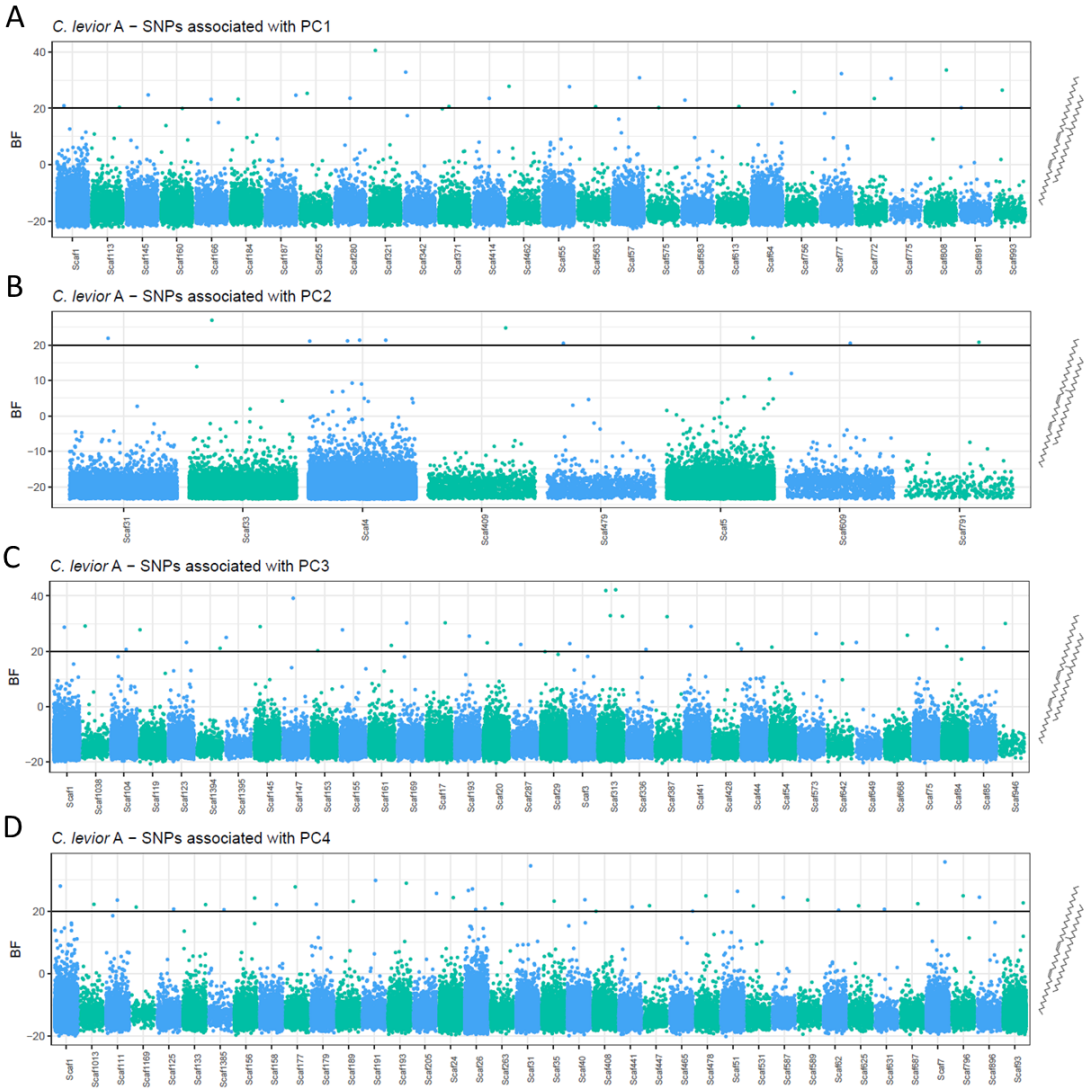


Figure S4.6: Overview of SNPs identified with BayPass that are significantly correlated (Bayes Factor >20) with **(A)** PC1, **(B)** PC2 **(C)** PC3, and **(D)** PC4 in the chemical association study in *C. levior* A.

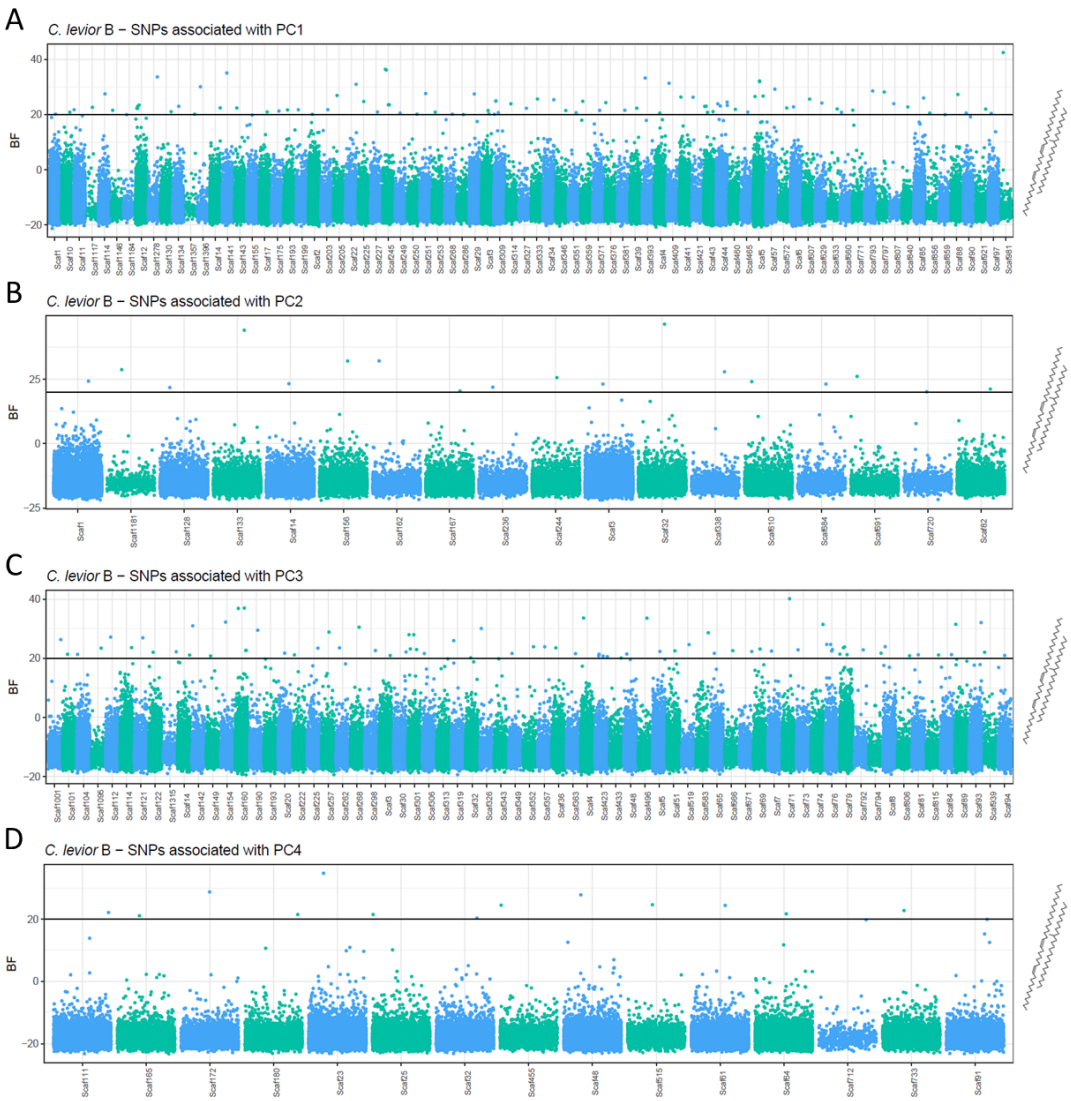


Figure S4.7: Overview of SNPs identified with BayPass that are significantly correlated (Bayes Factor >20) with **(A)** PC1, **(B)** PC2 **(C)** PC3, and **(D)** PC4 in the chemical association study in *C. levior* B.

Table S4.2: GO terms that were found in the environmental association analysis in *C. levior* A and *C. levior* B. Significantly enriched processes are indicated in bold type.

***C. levior* A**	
PC1	**histone H3-K27 demethylation** **DNA replication initiation** sensory perception of smell regulation of gene expression protein folding
PC2	chitin metabolic process ion transmembrane transport
PC3	**ammonium transmembrane transport** **phosphatidylinositol-mediated signaling** **vesicle docking involved in exocytosis** **phosphatidylinositol phosphorylation** phosphatidylinositol metabolic process lipid transport microtubule-based movement protein phosphorylation DNA integration transmembrane transport
NV	chemical synaptic transmission intracellular signal transduction
***C. levior* B**	
PC1	**chitin metabolic process** cellular amino acid metabolic process proteolysis
PC2	**transmembrane receptor protein tyrosine kinase signaling pathway** **homophilic cell adhesion via plasma membrane adhesion molecules** protein phosphorylation **response to metal ion**
PC3	**ATP hydrolysis coupled proton transport** microtubule-based movement carboxylic acid metabolic process translation transmembrane transport

Table S4.3: GO terms that were found in the chemical association analysis in *C. levior* A and *C. levior* B. Significantly enriched processes are indicated in bold type.

***C. levior* A**	
PC1	**sensory perception of smell** ribosome biogenesis **ATP hydrolysis coupled proton transport**
PC3	**neurotransmitter transport** apoptotic process regulation of transcription, DNA-templated phosphorus metabolic process **deoxyribonucleotide biosynthetic process** **phosphatidylinositol dephosphorylation**
PC4	oxidation-reduction process

C. levior **B**	
PC1	**glycogen metabolic process** **mitochondrial translation** immune response potassium ion transmembrane transport protein phosphorylation lipoprotein metabolic process regulation of transcription, DNA-templated protein glycosylation lipid transport intracellular signal transduction DNA repair signal transduction **posttranslational protein targeting to membrane, translocation**
PC2	response to stress cell surface receptor signaling pathway G protein-coupled receptor signaling pathway cilium organization
PC3	**glycogen catabolic process** **dsRNA transport** **glycogen biosynthetic process** ion transmembrane transport sensory perception of smell
PC4	regulation of small GTPase mediated signaling

General Discussion

Juliane Hartke

Chemical communication via cuticular hydrocarbons (CHCs) is at the very basis of the success of social insects. It enables the functioning of their complex communities by serving in foraging (as footprints), task-allocation, nest-mate recognition, mate-signalling and identification of features such as caste, sex and fertility (Thomas & Simmons 2008; van Zweden & d'Ettorre 2010). While gene families that are involved in CHC synthesis have been identified (reviewed in Chung & Carroll 2015) determining the genomic basis of specific changes in CHC profiles is challenging, as those gene families are also involved in fatty acid synthesis (but see Dembeck *et al.* 2015). Aside from being social signals, CHCs furthermore prevent desiccation and are thus additionally involved in ecological adaptation (Menzel *et al*. 2017; Sprenger *et al*. 2018). Often, traits with dual functions in both ecological adaptation and mate-signalling are proposed as so-called 'magic traits', i.e. traits with the potential to rapidly drive species divergence (Smadja & Butlin 2009; Chung *et al*. 2014; Chung & Carroll 2015). While this dual function makes CHCs a particularly interesting trait to study, at the same time it leads to difficulties in pinpointing the cause and effect of changes in CHC profiles.

In this thesis, I attempted to identify the genomic basis of CHC synthesis in the ant *Crematogaster levior* and furthermore tried to untangle the genomic patterns of population differentiation as well as environmental selection pressures underlying CHC profile variation. In collaboration with several project partners, I identified cryptic species within both *Cr. levior* and its mutualistic partner *Ca. femoratus*, using an integrative approach (*chapter 1*). I analysed the population structure of *Cr. levior* in detail and in correlation to several environmental parameters, showing that population substructure is weak and independent of climate. I sequenced the genome of *Cr. levior* A by using a combination of the three most commonly used sequencing techniques, Illumina, PacBio and Oxford nanopore MinION (*chapter 3*). This led to a highly contiguous assembly, which enabled the identification of potential candidate genes and the analysis of whole-genome population sequencing (Pool-Seq) data (*chapter 4*). Using a new approach in combining CHC profile information (*chapter 2*) and Pool-Seq data, I identified several candidate genes putatively involved in synthesis as well as detection of CHCs. In addition, I identified candidate genes under environmental selection and compared underlying genomic variation between cryptic species and found that patterns of selection were species-specific.

Identification of cryptic species

Tentative estimations suggest that the actual worldwide species diversity might, in fact, be much higher than what has been observed so far (Pfenninger & Schwenk 2007; Vodă *et al.* 2015a; Fišer *et al*. 2018; Struck *et al.* 2018). A large barcoding study in the Neotropics, for

example, confirmed this notion by showing that lepidopteran biodiversity has been underestimated by about 10-20% (Janzen *et al.* 2017). The reason for underestimations is that often distinct species are not recognized as such due to highly similar or identical morphological features (Bickford *et al.* 2007). In the case of neotropical lepidopterans, individuals that were genetically identified as cryptic species showed the same wing patterns and thus were previously classified as the same species (Janzen *et al.* 2017). Those so-called cryptic species are often found accidentally as a by-product of genetic analyses. It has been debated why cryptic species lack obvious morphological differences and proposed reasons are, amongst others, a lack of time for the evolution of distinct morphological features (Gustafsson *et al.* 2014), functional constraints (Smith *et al.* 2011c) or selection on morphological stasis (Bickford *et al.* 2007; Struck *et al.* 2018). However, it should be noted that the need for visual differentiation is entirely anthropocentric and there are obviously mechanisms in place by which cryptic species can distinguish between conspecifics and non-conspecifics. In social insects, for example, recognition is largely based on cuticular hydrocarbons, diminishing the need for visual cues (Hudson & Price 2014). Therefore, the fact that two distinct chemical morphs (chemotypes) were found in both *Cr. levior* (Emery & Tsutsui 2013) as well as *Ca. femoratus* (Menzel *et al.* 2014) called for a closer look to establish whether this was indicative for previously unidentified cryptic diversity within both genera (question (**i**)).

Clarification of species status was done using an integrative approach. Chemical, genetic as well as morphological analyses all showed clear results pointing at the existence of two cryptic species pairs, hereafter referred to as *Cr. levior* A, *Cr. levior* B, *Ca. femoratus* PAT and *Ca. femoratus* PS, after the previously characterized chemotypes (*chapter 1*). Notably, genetic species divergence on a level of mitochondrial DNA was much higher among *Cr. levior* A and B compared to *Ca. femoratus* PAT and PS, which was why for the latter additional nuclear markers were sequenced (*chapter 1*). However, results were inconsistent among markers, and only when combined show divergent lineages within *Ca. femoratus*. The simplest explanation for this is incomplete lineage sorting due to a rather recent divergence in the two *Ca. femoratus* species (Suh *et al*. 2015). I found higher genomic diversity within *Cr. levior* B compared to its sister species (*chapter 1 & 4*), and a higher number of loci that were under selection (*chapter 4*). Generally, adaptive potential should be positively correlated to standing genetic variation. Tropical species are postulated to be especially vulnerable to rising temperatures following climate change, as those species, compared to temperate species, are living closer to their thermal maximum (Deutsch *et al*. 2008; Tewksbury *et al.* 2008; Somero 2010). Thus, even a modest increase in temperature could have potentially devastating effects. It seems as if *Cr. levior* B, with a higher standing genetic variation may be better prepared to adapt to rising temperatures than *Cr. levior* A, although

of course, this is highly speculative, and the actual adaptive potential to a specific parameter depends on many additional factors.

Bearing the general assumption in mind that cryptic species are more prevalent than previously assumed (Pfenninger & Schwenk 2007; Vodă *et al.* 2015a; Fišer *et al*. 2018; Struck *et al.* 2018) and that they are now more readily being discovered due to the more frequent use of genetic methods, my findings are not too surprising. Especially species that do not depend on visual cues seem to harbour large cryptic diversity with many cryptic insect species having been discovered (e.g. Schultz *et al.* 2002; Fournier *et al.* 2012; Wilson *et al.* 2012; Darwell & Cook 2017a; Minard *et al.* 2017). However, the identification of cryptic species is of high importance, as their existence can have immense implications. In conservation biology, unidentified cryptic diversity can lead to misguided management efforts due to overestimated population sizes, distribution ranges and ecological flexibility (Fennessy *et al*. 2016; Delić *et al*. 2017). And in ecological studies, the discovery of cryptic species may, for example, change the interpretation of species interactions (Bickford *et al.* 2007).

Speciation

Inferring past events from present observations comes almost automatically for evolutionary biologists. Naturally, I was interested in why and how cryptic speciation happened in both genera and which processes drove divergence. Especially the pattern of sympatric occurrence across French Guiana in such closely related species that on top of it all also share the same mutualistic partner was puzzling. The divergence of species is a process shaped by multiple mechanisms, that may act at the same time. Depending on whether the speciation event was driven by ecological selection or by stochastic processes, the divergence can be classified as either ecological speciation (see 1.1.1) or non-ecological speciation (1.1.2). However, since none of those processes are mutually exclusive and one process can influence the strength and effect of the other ones, the underlying cause for a single speciation event is hard to determine. In addition, speciation mostly is a gradual process happening over a lengthy period of time that cannot be monitored constantly and directly. Integrating genome-wide sequencing data can help to make inferences on targets and agents of selection (reverse ecology) and thus deduct possible speciation scenarios (Li *et al*. 2008). However, evidence for either mode of speciation can, by its own nature, only be indirect (Shafer & Wolf 2013). Many (largely futile) attempts were taken to show for a certain system how speciation took place and which mechanisms played the leading role. I will humour this debate and, in the following, discuss which mechanisms of species divergence were plausible in having driven the separation in both species pairs, but in the end show, that also here in this system, this question cannot be answered with any certainty.

Ecology as a potential driver of speciation

At the moment, ecological speciation seems to be at the pole position in the debates trying to determine the most likely driver of speciation with more and more indirect evidence accumulating for a variety of systems (Schluter 2001, 2009; Rundle & Nosil 2005; Funk *et al.* 2006; Shafer & Wolf 2013). Ecological selection pressures can be exerted on individuals by any biotic or abiotic factor within their habitat, including species interactions and climate (details in 1.1.1). After identifying the cryptic species pairs, I examined several parameters for their likelihood of having contributed to species divergence (question (**iv**)). The climate in French Guiana, for instance, follows a gradient from a wetter and cooler east to a drier and warmer west. If speciation was driven by allopatric adaptation to different climates (not necessarily within French Guiana), this should, in theory, be visible as a distinct distribution pattern across the heterogeneous sampling range (question (**ii**)). In principle, when two species largely share the same niche, which is the case in this system, the species that are better adapted to local conditions should outcompete the less well-adapted species (Urban *et al*. 2012; Bocedi *et al*. 2013). *Cr. levior* A and B are occurring in sympatry across the sampling range in French Guiana and their abundance at different sampling sites is independent of any climate factors (*chapter 1*). This makes it rather unlikely that the speciation process was driven by adaptations to different climates unless both species are phenotypically plastic enough to compensate for non-optimal adaptation to their current distribution range. In *Ca. femoratus* on the other hand, the two cryptic species show a distinct difference in their distribution that correlates with the local climate. *Ca. femoratus* PS is more abundant in the drier and warmer regions, while *Ca. femoratus* PAT occurs more frequently in the cooler and wetter parts of French Guiana. Here, a scenario of allopatric speciation in two climatically distinct areas and a subsequent range expansion into French Guiana by either one or both species seems plausible. Classically, allopatric speciation follows a certain pattern of geographical separation of populations that then accumulate genetic and often phenotypical differences, either randomly or as an adaptive response to new selection pressures. This might then lead to either partial or complete reproductive isolation. Following a secondary contact, selection against hybrids is taking place and premating isolation evolves by reinforcement (Schluter 2001; Nosil 2012) Depending on which mitochondrial molecular clock estimation is applied (2.3 - 4% per million years in insects; Norman *et al.* 2016), the split between the two *Ca. femoratus* species occurred sometime between 96 kya and 168 kya (based on 0.3% sequence divergence in the mitochondrial marker COI; *chapter 1*). The Guiana Shield, which spans a coastal region in southern America from Venezuela into Brazil, harbours about a quarter of the world's tropical evergreen rainforest (Higgins 2007) and is known for its huge biodiversity (Melo-Sampaio *et al.* 2019). Fluctuations in temperature, precipitation and atmospheric CO2 that are associated with glacial and interglacial periods in the mid to late Pleistocene are thought

to have driven diversification in this region according to the **Disturbance-Vicariance Hypothesis** (Bush 1994; Noonan & Gaucher 2005). While it has been shown that the rainforest itself more or less stayed intact even during the Last Glacial Maximum (Colinvaux & Oliveira 2001; Kleidon & Lorenz 2001; Bush & Oliveira 2006; Wang *et al.* 2017), it seems that parts of the now evergreen rainforest might have given rise to more drought-resistant deciduous forests during glaciation periods in the Andes and a resulting drop in temperature and precipitation in the Amazonian low lands (Mayle *et al*. 2009). Especially in ectothermic organisms such as ants, where temperature is one of the most important selection pressures (Deutsch *et al.* 2008), fluctuating temperatures could very plausibly have led to diversification (Noonan & Gaucher 2005). Thus, for the two *Camponotus femoratus* species, French Guiana might constitute the easternmost and westernmost part of their respective distribution ranges and could be a secondary contact zone. However, this is highly speculative, as data for outside of French Guiana is sparse. Future work should thus expand the sampling ranges to the south-east into Brazil and to the west into Suriname and Guiana. Additional data on species abundance from these regions might give clues about the species migration histories and possibly show whether ecological selection might have played a role in species divergence.

Mutualism as a driver of speciation

Cr. levior A and B seem to have diverged earlier than the two *Camponotus* species about 0.95 mya to 1.65 mya (based on 3.78% sequence divergence in the mitochondrial marker COI; *chapter 1*). Since the speciation events among cryptic *Cr. levior* and *Ca. femoratus* species happened at different time points, a co-speciation scenario, in which species divergence of *Ca. femoratus* followed the split of the two *Cr. levior* lineages due to differential selection pressures exerted by different mutualistic partner species is possible (question (**iv**)). The most pronounced dissimilarity between *Cr. levior* A and B are their cuticular hydrocarbon profiles. This might have driven recognition-based divergence in their mutualistic partner. A similar mechanism has been proposed to explain the matching phylogenies of pollinating and non-pollinating fig wasps that seem to be based on the recognition of the same chemicals (Machado *et al.* 1996). If co-speciation between *Cr. levior* species and *Ca. femoratus* species would have taken place, specialization to a specific mutualistic partner should be observable. However, the pairing of species within the nests was random and seemed to be a question of availability rather than preference (*chapter 1*), which makes a co-speciation scenario less likely. While I did not find evidence for co-speciation, *Cr. levior* and *Ca. femoratus* do show signs of coevolution. In *chapter 2* we demonstrate that cuticular hydrocarbon profiles are influenced by the parabiotic partner, which means that the species are either actively acquiring the substances from their partner (i.e. through grooming or trophallaxis) or selectively synthesising them dependent of the partner species. Whichever

the mechanism, this pattern usually indicates a form of chemical mimicry or camouflage (Bagnères & Lorenzi 2010).

Sexual selection as a potential driver of speciation

Since *Cr. levior* A and B share the same distribution range in French Guiana and thus do not seem to be adapted to different habitats or climates, sympatric speciation cannot be ruled out. For divergence to take place in sympatry, a strong driver is needed to prevent homogenizing gene flow between diverging entities. One such proposed driver is sexual selection (Higashi *et al*. 1999; Kirkpatrick & Ravigne 2012). As described in more detail in 1.1, sexual selection favours traits that are implicated in a higher mating success even though they might be detrimental to survival. Traits that are proposed as especially effective in driving divergence are so-called 'magic' traits, which at the same time provide ecological adaptive value and promote non-random mating (Servedio *et al.* 2011). In *Heliconius* butterflies, for example, colour patterns that are under selection due to mimicry, also cause non-random mating and as a result, led to reproductive isolation between two sister species (Jiggins *et al.* 2001). A prime candidate for such a trait in social insects is the cuticular hydrocarbon (CHC) profile, that is linked to many functions at once. CHCs provide ecological adaptive value by functioning in drought resistance, foraging (as footprints) and nest-mate discrimination and, at the same time, also function in mate choice and thus cause non-random mating (Blomquist & Bagnères 2010), as has been demonstrated in *Drosophila serrata* (Chung *et al.* 2014). *Cr. levior* A and B show vast differences in their CHC profiles as well as in the general abundance of substance classes (question (**iii**); *chapter 2*). A closer comparison including other species of the *Orthocrema* clade revealed that *Cr. levior* A was not only distinctly different from *Cr. levior* B but also from all other *Orthocrema* species. Since those species largely occur in the same region, and *Cr. levior* B occurs in exactly the same habitats (*chapter 1*), those profound changes in CHC profiles in *Cr. levior* A point to a very strong selection pressure that is most likely independent of climate. The changes could be implicated in ecological functioning, yet the divergence between the two cryptic species was very recent and for such pronounced chemical changes to have occurred without an external selection pressure, such as high temperature or precipitation, sexual selection is the more likely driver. Furthermore, even if the changes in the CHC profile occurred as an adaptation to a change in climate, such drastic changes would be highly inefficient, when simple tweaks in the abundance of substance classes would suffice. It is thus likely that the distinct differences in CHC profile in *Cr. levior* A evolved - at least in part - due to sexual selection. Whether this was the basis of species divergence or a measure of reinforcement after speciation due to another mechanism had already taken place, is hard to say in retrospect.

To make a conclusive inference about which mechanism drove species divergence in *Crematogaster* and *Camponotus*, data about the distribution and population structure outside of French Guiana is missing. However, from the data we have today, based on a lack of differential climatic adaptation and highly divergent CHC profiles in *Crematogaster*, I would think it more likely that in this case, divergence was driven by sexual selection. For *Camponotus*, where we found clear distribution patterns dependent on temperature and precipitation, I would speculate that the divergence took place in allopatry and under different climate regimes that are mirrored in the current distribution.

The genetic basis of cuticular hydrocarbons and climate adaptation – a synthesis

Functions of cuticular hydrocarbons have been studied extensively in several invertebrate groups, yet the genetic basis of single substances or changes in substance abundance have largely been neglected. Most studies on this topic have been conducted on *Drosophila* (Chung *et al*. 2014; Dembeck *et al.* 2015; Ivory-Church *et al.* 2015), whereas in social insects, in which CHCs fulfil functions that lie at the very basis of their success, surprisingly little is known. Hymenoptera and especially ants are known for the rapid evolution of specific gene families with roles in chemical communication, such as odorant receptors (McKenzie *et al*. 2016; Pask *et al.* 2017) and desaturases (Simola *et al.* 2013; Helmkampf *et al*. 2015). The expansion of these gene families is often associated with the evolution of eusociality (Simola *et al.* 2013; Helmkampf *et al.* 2015; Zhou *et al.* 2015).

Expansion of CHC gene families

In *chapter 3* we used the newly sequenced and annotated genome of *Cr. levior* A as a basis for a detailed search for elongases and desaturases, two gene families with major roles in CHC synthesis. Elongases are involved in elongation of carbon chains, and desaturases introduce double bonds and thus are involved in the synthesis of unsaturated cuticular hydrocarbons such as alkenes and alkadienes (Dallerac *et al.* 2000). I then compared the numbers found in *Cr. levior* A to those of all Hymenoptera with an available reference genome and show that ants have a higher number of elongases, and desaturases compared to the mean in all Hymenoptera. *Cr. levior* A had the highest number of desaturases among all compared species (together with *Pseudomyrmex gracilis*) and among the highest number of elongases, which is in line with the increase in chain elongation we were able to show in this species in *chapter 2*. The high number of desaturases in *Cr. levior* A might be reflective of the species mutualistic lifestyle. Desaturases introduce double bonds into the hydrocarbon chains, and *Cr. levior* A as well as *Ca. femoratus* PS and PAT have both been shown to produce an

unusually high number of dienes, indicating a large variation in double-bond positions (Menzel *et al.* 2014; Sprenger *et al*. submitted). In contrast, the CHC profile of *Cr. levior* B shows more saturated substances (Menzel *et al.* 2014; Sprenger *et al*. 2019). The number of desaturases in *Cr. levior* B is unknown, but the desaturase-related differences in the CHC profiles of the two sister species, that have been elucidated in *chapter 2*, could point to an expansion of the desaturase gene family in *Cr. levior* A. It has been speculated before that the expansion of desaturase genes in ants is related to their complex social structure and the resulting elaborate communication system (Helmkampf *et al*. 2015). In an ant species that due to its mutualistic lifestyle has an additional communication layer, an expansion of genes that facilitates the production of unsaturated compounds that provide sufficient diversity to act as recognition cues, intuitively makes sense. Indeed, unsaturated CHCs have for example been shown to facilitate recognition of the reproductive status of queens (Liebig *et al.* 2000) or caste differences (Cash 2016).

Phenotypic plasticity vs. heritability

CHC profiles can differ between species, colonies, castes and individuals, but also between different time points in an individual's life driven by life-history changes (Petfield *et al*. 2005; Thomas & Simmons 2011) or fluctuations in temperature and precipitation (Menzel *et al*. 2018; Sprenger *et al*. 2018). Differences between species are often qualitative, while acclimation-driven changes to the CHC profile are mostly quantitative (Sprenger & Menzel 2020). This indicates that differences in CHC profiles can be based both on heritable adaptation as well as plastic acclimation. Indeed, gene families with a known role in CHC synthesis, show distinct differences in expression patterns among *Cr. levior* and *Ca. femoratus* sister species (Sprenger *et al*. submitted). This shows that at least parts of species-specific profiles might be based on gene-expression differences, which has also been postulated for two closely related grasshopper species (Finck *et al.* 2016). In contrast to this, in *chapter 4* I show that even within a species, differences in CHC profiles can be based on heritable changes within the genome. *Cr. levior* A and B both showed distinct allele frequency differences between populations that were correlated with their specific CHC profiles. Taken together, this shows that there are apparently strong selection pressures on CHC profiles that result in genomic changes, but at the same time underlines the importance of flexible adjustments to individual profiles.

CHC profiles and climate adaptation

In *chapter 4*, it becomes clear that local climate is an important selection pressure on *Cr. levior* A and B, as both species show strong signals of selection exerted by differences in the local climate. In fact, even though both species show very high levels of genomic homogeneity and gene flow across French Guiana, signals of local adaptation to climate

were still present. In addition to this, we found in *chapter 2* and *chapter 4* that differences in individual CHC profiles and population-wide profiles respectively are correlated to climate. Thus, it seems plausible that climatic differences within the examined distribution range of *Cr. levior* have driven population-specific adaptations in CHC profiles. This pattern has been shown in *Drosophila melanogaster*, where CHC profiles vary along a climate gradient along the East coast of the United States (Rajpurohit *et al.* 2017). These strong effects of local climate are not too surprising, though, since insects are particularly sensitive to variations in temperature and precipitation. Consequently, also other studies were able to show genomic variation in response to climate clines. In the harlequin fly, *Chironomus riparius* about 1 % of genes show variation among populations across a climate gradient within Europe (Waldvogel *et al*. 2018), and in *Anopheles gambiae*, a precipitation gradient across Cameroon drove population-specific adaptations (Cheng *et al.* 2012). In contrast, the CHC profiles of *Ca. femoratus* PS and PAT are not correlated with climate parameters (*chapter 2*), which makes sense, as the two species already show a climate-dependent distribution. Following this line, their intraspecific CHC profiles are less diverse than the profiles of *Cr. levior* A and B (*chapter 2*), which probably reflects the fact that the two *Cr. levior* species each inhabit a larger variety of habitats and thus have to adapt their CHC profiles accordingly. Another possibility, however, is that CHC profiles between *Ca. femoratus* species are more distinct as a result of reinforcement of premating barriers in the absence of other incompatibilities.

Finally, the population genomic analyses (*chapter 4*) resulted in the identification of several candidate genes with putative roles in CHC synthesis and perception, as well as adaptation to climate (question (**v**)). Interestingly, between populations, selection often acted on genes that are implicated in the perception of CHCs: in both *Cr. levior* species, I found odorant receptors (ORs) that in their allele frequency were significantly correlated with the population-specific CHC profiles. In *Cr. levior* A, one of the ORs was also correlated with climate parameters, which makes it a prime candidate for the detection of climate-dependent recognition cues as well as population-specific recognition. To my knowledge, this is the first example of climate-dependent selection within an OR gene. Odorant receptors are required for the perception of smell. Interestingly, *Drosophila mojavensis* showed upregulations in the biological process *sensory perception of smell*, when put under desiccation stress (Rajpurohit *et al.* 2013). Similarly, in *Cr. levior*, climate factors that vary across the distribution range and lead to desiccation stress in parts of the habitat may have led to selection pressures on ORs. An explanation for this could be the fact that an individual's sensitivity towards CHCs decreases with increasing temperatures (Riveron *et al*. 2009) and thus, ORs might be under selection in warmer habitats to improve CHC perception. Another explanation for the association with climate could be more indirect. Changes in CHC profiles as a reaction to climate may have led to selection pressures on odorant receptors, although, there are no examples from other studies that demonstrate

genetic changes or differential expression in ORs as a response to changes in CHCs. Additionally, I found allele frequency variation within a fatty-acyl CoA reductase correlated with CHC profiles in *Cr. levior* B, which makes this gene a prime candidate for the origin of contrasting profiles between populations. As we found no correlation to climate of this gene, it seems that the changes it causes are not adaptive to distinct climate conditions, but rather function in communication. This fits a study, in which knockouts of fatty-acyl CoA reductases led to changes in the abundance of substances with a known role in communication in *D. melanogaster* (Dembeck *et al.* 2015).

At first glance, it might be surprising that I found almost no significant correlations between differences in CHC profiles and genes with a known function in their synthesis. However, a study on the transcriptomes of these species (Sprenger *et al*. submitted) underlined the importance of gene expression as a means to flexible alterations of CHC profiles. Hence, particularly within a single species changes to the CHC profiles are probably mainly facilitated by phenotypic plasticity and not heritable. On the other hand, the significant correlations to multiple odorant receptors could make sense if those structures are under selection, i.e. due to changed profiles, but cannot be plastically altered unlike the profiles themselves.

I furthermore showed in *chapter 4* that even though both *Cr. levior* species are subject to the same environmental selection pressures (*chapter 1*), the signatures of selection across their genomes were neither the same nor did they show any significant overlap in selected genes or pathways (question (**vi**)). This is either a strong indicator for genomic redundancy, with multiple separate genomic trajectories that lead to the same adaptational value, or a signal for genomic constraints that prohibit the same mutational outcome. The **mutation-order principle**, for example, proclaims that the same mutational changes in the same temporal order between two populations are highly unlikely, and thus by chance, different evolutionary trajectories are followed (Mani & Clarke 1990; Schluter 2009). However, which of the two possibilities is at work here, is hard to evaluate. Often, eco-evolutionary studies are driven by a wish to elucidate the capacity of a species to adapt to climate change and to find common evolutionary responses that can be used in prediction modelling (Alberto *et al.* 2013; Waldvogel *et al.* 2019). Yet, many past studies seem to support a mode of evolution that acts haphazardly and unpredictably (Pfenninger *et al.* 2015; Dennis *et al.* 2015; Feldmeyer *et al.* 2017). My results in *chapter 4* substantiate this notion among two closely related species with a largely congruent genomic basis and the same climatic selection pressures, which should, in theory, be an ideal scenario to witness parallel evolution (Bailey 2015). To predict a species potential to adapt to changes in their environment, it thus does not seem to be sufficient to make inferences from other species, even closely related ones, as they may be vastly different in adaptational trajectories or

potential. However, I only compared two species within a limited part of their distribution range, therefore caution in the interpretation is required.

CONCLUSION AND OUTLOOK

This thesis provides evidence for cryptic speciation within the two mutualistic ants *Cr. levior* and *Ca. femoratus*, following an increasing amount of studies that find cryptic insect species (e.g. Schultz *et al*. 2002; Fournier *et al.* 2012; Wilson *et al.* 2012; Darwell & Cook 2017; Minard *et al.* 2017). I took a deeper look into the population structure of the cryptic species and can deduct several plausible trajectories of species divergence from their distribution ranges and other traits. In my opinion, stark differences in cuticular hydrocarbon profiles point to sexual selection as the main driver of species divergence within *Cr. levior*, as they show no differences in their distribution ranges (within French Guiana) and furthermore show no signs of differential ecological adaptations. Sexual selection has gained more and more attention in the past as a mechanism that can rapidly drive speciation in sympatric species (Kirkpatrick & Ravigné 2002; Higashi *et al.* 1999), and the presented system could be an example for this scenario. Contrastingly, within *Ca. femoratus*, speciation, in my opinion, was more likely driven by allopatric adaptation to different habitats and possibly different climates, as those two species show distinct distribution patterns based on temperature and precipitation. Climate-driven reproductive isolation, so far, has never conclusively been shown; it is, however, argued that the complex changes following adaptation to climate should increase the likelihood of reproductive isolation (Qvarnström *et al.* 2016). Genome-wide association studies using Pool-Seq data have been demonstrated to be suited in the identification of candidate genes underlying quantitative traits within different species (Stinchcombe & Hoekstra 2008; Mallard *et al.* 2018; Waldvogel *et al.* 2018). In this thesis, I identified genes that are differentially selected in contrasting climates within both *Cr. levior* species. Most strikingly, those genes encompassed odorant receptors, for which selection due to climate has not been shown before, but that could be implicated in adaptation to warmer temperatures (Riveron *et al.* 2009; Rajpurohit *et al.* 2013). In addition to climate, ORs were also significantly associated with population-specific differences of CHC profiles. Generally, OR genes showed substantial expansions in ants compared to other Hymenoptera and are often implicated in facilitating their advanced level of eusociality due to their central role in communication (Zhou *et al.* 2015; Pask *et al.* 2017; Slone *et al*. 2017). Studies on the molecular basis of ORs and the physiological interaction between single CHC substances and ORs are only now emerging for ants and provide a basis for the study of the molecular underpinnings of sociality (McKenzie *et al*. 2016; Pask *et al.* 2017; Slone *et al.* 2017; McKenzie & Kronauer 2018). Taken together, my results substantiate claims of the immense importance of ORs in ants and provide indications for

additional selection pressures, i.e. climate. Overlap of targets of selection between *Cr. levior* A and B was minimal and limited to the level of biological pathways. Thus, it seems as if adaptation to the same environment is facilitated by different evolutionary trajectories, pointing to genomic redundancy or possibly evolutionary constraints.

The findings of this thesis provide a solid foundation for further studies on the genetic basis of cuticular hydrocarbon synthesis and adaptation. To functionally confirm the candidate genes I identified, as a next step, knockouts of genes and assessment of behavioural and chemical changes should be employed. First experiments show promising results for several genes, for which knockouts led to changes in CHC profiles and recognition (results not shown). However, this was only possible for *Cr. levior* A, as this is currently the only species for which we sequenced a reference genome. Thus, to enable further analyses for the other species within this mutualistic complex, reference genomes for *Cr. levior* B and *Ca. femoratus* PAT and PS are needed as well. Especially concerning the cryptic species, that were identified within this thesis, reference genomes for all four species would also help to gain a deeper insight into the mechanisms of speciation, which can be driven by structural changes to the genome (Feder *et al.* 2012; Potter *et al.* 2017). The distinctly different CHC profiles and the high number of desaturases in *Cr. levior* A could point to expansions of gene families, for example, due to large-scale duplications, as a very plausible player in the emergence of cryptic species. As differences between *Ca. femoratus* PAT and PS are rather small and speciation occurred only very recently, they are a highly interesting system to study ongoing divergence within a potential zone of secondary contact in French Guiana. Furthermore, as those two species possess vastly different CHC profiles and show differences in their distribution range dependent on temperature and precipitation, it would be highly interesting to compare genome-wide signatures of selection to those in *Cr. levior*. To clarify the species' evolutionary histories, sampling areas should be extended to outside of French Guiana. This would gain more information about the species' realized niches and could possibly clarify whether divergence might have been driven by differences in climate. If this is not the case, then the cryptic species here might be a highly interesting example for speciation through sexual selection on cuticular hydrocarbons.

REFERENCES

Abrams, P. A. 1987. On classifying interactions between populations. Oecologia 73:272–281.

Adler, P. B., J. HilleRisLambers, and J. M. Levine. 2007. A niche for neutrality. Ecol. Lett. 10:95–104.

Aidoo, M., D. J. Terlouw, M. S. Kolczak, P. D. McElroy, F. O. Ter Kuile, S. Kariuki, B. L. Nahlen, A. A. Lal, and V. Udhayakumar. 2002. Protective effects of the sickle cell gene against malaria morbidity and mortality. Lancet 359:1311–1312.

Aitchison, J. 1982. The Statistical Analysis of Compositional Data. J. R. Stat. Soc. B 44:139–177.

Alberto, F. J., S. N. Aitken, R. Alía, S. C. González-Martínez, H. Hänninen, A. Kremer, F. Lefèvre, T. Lenormand, S. Yeaman, R. Whetten, and O. Savolainen. 2013. Potential for evolutionary responses to climate change - evidence from tree populations. Glob. Chang. Biol. 19:1645–1661.

Alexa, A., and J. Rahnenführer. 2016. Gene set enrichment analysis with topGO. R Packag. version 2.24. 0.

Alleman, A., B. Feldmeyer, and S. Foitzik. 2018. Comparative analyses of co-evolving host-parasite associations reveal unique gene expression patterns underlying slavemaker raiding and host defensive phenotypes. Sci. Rep. 8:1–14. Springer US.

Alves-Bezerra, M., E. L. Klett, I. F. De Paula, I. B. Ramos, R. A. Coleman, and K. C. Gondim. 2016. Long-chain acyl-CoA synthetase 2 knockdown leads to decreased fatty acid oxidation in fat body and reduced reproductive capacity in the insect *Rhodnius prolixus*. Biochim Biophys Acta 1861:650–662.

Anderson, M. J. 2017. Permutational Multivariate Analysis of Variance (PERMANOVA). Wiley StatsRef Stat. Ref. Online 1–15.

Andersson, M. 1982. Female choice selects for extreme tail length in a widowbird. Nature.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.

Ardila-Garcia, A. M., G. J. Umphrey, and T. R. Gregory. 2010. An expansion of the genome size dataset for the insect order Hymenoptera, with a first test of parasitism and eusociality as possible constraints. Insect Mol. Biol. 19:337–346.

Attygalle, A. B. 1998. Microchemical Techniques. Pp. 207–294 *in* J. G. Millar and K. F. Haynes, eds. Methods in Chemical Ecology - Volume 1: Chemical Methods. Springer Science & Business Media, New York.

Bagnères, A.-G., and M. C. Lorenzi. 2010. Chemical deception/mimicry using cuticular hydrocarbons. Pp. 282–324 *in* Blomquist GJ, Bagnères A-G (eds) Insect hydrocarbons: biology, biochemistry, and chemical ecology. Cambridge University Press, New York.

Bailey, S. F., N. Rodrigue, and R. Kassen. 2015. The effect of selection environment on the probability of parallel evolution. Mol. Biol. Evol. 32:1436–1448.

Baltimore, D. 1970. RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature 226:1209–1211.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. O. N. Pham, A. D. Prjibelski, A. V Pyshkin, A. V Sirotkin, N. Vyahhi, G. Tesler,

M. A. X. A. Alekseyev, and P. A. Pevzner. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. 19:455–477.

Barghi, N., R. Tobler, V. Nolte, A. M. Jakšić, F. Mallard, K. A. Otte, M. Dolezal, T. Taus, R. Kofler, and C. Schlötterer. 2019. Genetic redundancy fuels polygenic adaptation in *Drosophila*.

Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation. Trends Ecol. Evol. 23:38–44.

Bartlett, J. W., and C. Frost. 2008. Reliability, repeatability and reproducibility: Analysis of measurement errors in continuous variables. Ultrasound Obstet. Gynecol. 31:466–475.

Bastide, H., J. D. Lange, J. B. Lack, A. Yassin, and J. E. Pool. 2016. A variable genetic architecture of melanic evolution in *Drosophila melanogaster*. Genetics 204:1307–1319.

Baur, H., Y. Kranz-Baltensperger, A. Cruaud, J. Y. Rasplus, A. V. Timokhov, and V. E. Gokhman. 2014. Morphometric analysis and taxonomic revision of *Anisopteromalus ruschka* (Hymenoptera: Chalcidoidea: Pteromalidae) - an integrative approach. Syst. Entomol. 39:691–709.

Baur, H., and C. Leuenberger. 2011. Analysis of ratios in multivariate morphometry. Syst. Biol. 60:813–825.

Beatty, J. 2006. Replaying Life's Tape. J. Philos. 103:336–362.

Bell, G. 2017. The Distribution of Abundance in Neutral Communities. Am. Nat. 155:606.

Bergland, A. O., R. Tobler, J. González, P. Schmidt, and D. Petrov. 2016. Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*. Mol. Ecol. 25:1157–1174.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22.

Blaimer, B. 2012. Acrobat ants go global – Origin, evolution and systematics of the genus *Crematogaster* (Hymenoptera: Formicidae). Mol. Phylogenet. Evol. 65:421–436.

Blix, A. S. 2016. Adaptations to polar life in mammals and birds. J. Exp. Biol. 219:1093–1105.

Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. Evolution (N. Y). 57:717–745.

Blomquist, G. J. 2010. Structure and analysis of insect hydrocarbons. In G. J. Blomquist, & A.-G. Bagnères (Eds.), *Insect hydrocarbons: Biology, biochemistry, and chemical ecology*. New York.

Blomquist, G. J., and A.-G. Bagnères. 2010. *Insect Hydrocarbons*. Cambridge University Press, New York.

Bocedi, G., K. E. Atkins, J. Liao, R. C. Henry, J. M. J. Travis, and J. J. Hellmann. 2013. Effects of local adaptation and interspecific competition on species' responses to climate change. Ann. N. Y. Acad. Sci.

Boetzer, M., and W. Pirovano. 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics 15:1–9.

Bolaños, L. M., M. Rosenblueth, A. M. De Lara, A. Migueles-Lozano, C. Gil-Aguillón, V. Mateo-Estrada, F. González-Serrano, C. E. Santibáñez-López, T. García-Santibáñez, and E. Martínez-

Romero. 2019. Cophylogenetic analysis suggests cospeciation between the Scorpion Mycoplasma Clade symbionts and their hosts. PLoS One 14:1–20.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Bolnick, D. I. 2001. Intraspecific competition favours niche width expansion in *Drosophila melanogaster*. Nature 410:463–466.

Bonasio, R., G. Zhang, C. Ye, N. S. Mutti, X. Fang, N. Qin, G. Donahue, P. Yang, Q. Li, C. Li, P. Zhang, Z. Huang, S. L. Berger, D. Reinberg, J. Wang, and J. Liebig. 2010. Genomic Comparison of the Ants *Camponotus floridanus* and *Harpegnathos saltator*. Science (80-.). 329:1068–1071.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. 10:e1003537.

Boyle, J. H., D. Martins, P. M. Musili, and N. E. Pierce. 2018. Population genomics and demographic sampling of the ant-plant *Vachellia drepanolobium* and its symbiotic ants from sites across its range in East Africa. Front. Ecol. Evol. 7:206.

Brand, P., S. R. Ramírez, F. Leese, J. J. G. Quezada-Euan, R. Tollrian, and T. Eltz. 2015. Rapid evolution of chemosensory receptor genes in a pair of sibling species of orchid bees (Apidae: Euglossini). BMC Evol. Biol. 15:176.

Brückner, A., and M. Heethoff. 2017. A chemo-ecologists' practical guide to compositional data analysis. Chemoecology 27:33–46.

Bruneaux, M., M. Nikinmaa, V. N. Laine, K. Lindström, C. R. Primmer, and A. Vasemägi. 2014. Differences in the metabolic response to temperature acclimation in nine-spined stickleback (*Pungitius pungitius*) populations from contrasting thermal environments. J. Exp. Zool. Part A Ecol. Genet. Physiol. 321:550–565.

Buchwald, R., and M. D. Breed. 2005. Nestmate recognition cues in a stingless bee, *Trigona fulviventris*. Anim. Behav. 70:1331–1337.

Buellesbach, J., B. A. Whyte, E. Cash, J. D. Gibson, K. J. Scheckel, R. Sandidge, and N. D. Tsutsui. 2018. Desiccation Resistance and Micro-Climate Adaptation: Cuticular Hydrocarbon Signatures of Different Argentine Ant Supercolonies Across California. J. Chem. Ecol. 44:1101–1114.

Bush, M. B. 1994. Amazonian Speciation: A Necessarily Complex Model. J. Biogeogr. 21:5.

Bush, M. B., and P. E. de Oliveira. 2006. The rise and fall of the Refugial Hypothesis of Amazonian speciation: a paleoecological perspective. Biota Neotrop. 6.

Bushnell, B. 2014. BBMap: A Fast, Accurate, Splice-Aware Aligner. P. *in* 9th Annual Genomics of Energy & Environment Meeting, Walnut Creek, CA, March 17-20, 2014.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Carlson, D. A., U. R. Bernier, and B. D. Sutton. 1998. Elution patterns from capillary GC for methyl-branched alkanes. J. Chem. Ecol. 24:1845–1865.

Cash, E. 2016. Proximate and Ultimate Mechanisms of Nestmate Recognition in Ants. Dissertation. Arizona State University.

Cavener, D., G. Corbett, D. Cox, and R. Whetten. 1986. Isolation of the eclosion gene cluster and the developmental expression of the Gld gene in *Drosophila melanogaster*. EMBO J. 5:2939–2948.

Céréghino, R., C. Leroy, A. Dejean, and B. Corbara. 2010. Ants mediate the structure of phytotelm communities in an ant-garden bromeliad. Ecology 91:1549–1556.

Chaffee, R. R. J., and J. C. Roberts. 1971. Temperature Acclimation in Birds and Mammals. Annu. Rev. Physiol. 33:155–202.

Chamberlain, S. A., J. L. Bronstein, and J. A. Rudgers. 2014. How context dependent are species interactions? Ecol. Lett. 17:881–890.

Charlesworth, B. 2009. Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10:195–205.

Charmantier, A., C. Doutrelant, G. Dubuc-Messier, A. Fargevieille, and M. Szulkin. 2016. Mediterranean blue tits as a case study of local adaptation. Evol. Appl. 9:135–152.

Chaturvedi, S., L. K. Lucas, C. C. Nice, J. A. Fordyce, M. L. Forister, and Z. Gompert. 2018. The predictability of genomic changes underlying a recent host shift in Melissa blue butterflies. Mol. Ecol. 27:2651–2666.

Cheng, C., B. J. White, C. Kamdem, K. Mockaitis, C. Costantini, M. W. Hahn, and N. J. Besansky. 2012. Ecological Genomics of *Anopheles gambiae* Along a Latitudinal Cline: A Population-Resequencing Approach. Genetics 190:1417–1432.

Chomicki, G., and S. S. Renner. 2017. The interactions of ants with their biotic environment. Proc. R. Soc. B Biol. Sci. 284.

Chomicki, G., P. S. Ward, and S. S. Renner. 2015. Macroevolutionary assembly of ant/plant symbioses: *Pseudomyrmex* ants and their ant-housing plants in the neotropics. Proc. R. Soc. B Biol. Sci. 282.

Chung, H., and S. B. Carroll. 2015. Wax, sex and the origin of species: Dual roles of insect cuticular hydrocarbons in adaptation and mating. BioEssays 37:822–830.

Chung, H., D. W. Loehlin, H. D. Dufour, K. Vaccarro, J. G. Millar, and S. B. Carroll. 2014. A single gene affects both ecological divergence and mate choice in *Drosophila*. Science. 343:1148–1151.

Clark, M. S., and M. R. Worland. 2008. How insects survive the cold: Molecular mechanisms - A review. J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 178:917–933.

Clutton-Brock, T. 2009. Sexual selection in females. Anim. Behav. 77:3–11. Elsevier Ltd.

Colinvaux, P. A., and P. E. De Oliveira. 2001. Colinvaux, Deoliveira_2001_Amazon plant diversity and climate through the Cenozoic. Palaeogeogr. Palaeoclimatol. Palaeoecol. 166:51–63.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, H. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science. 307:1928–1933.

Cruaud, A., N. Ronsted, B. Chantarasuwan, L. S. Chou, W. L. Clement, A. Couloux, B. Cousins, G. Genson, R. D. Harrison, P. E. Hanson, M. Hossaert-Mckey, R. Jabbour-Zahab, E. Jousselin, C.

Kerdelhué, F. Kjellberg, *et al.* 2012. An Extreme case of plant-insect codiversification: Figs and fig-pollinating wasps. Syst. Biol. 61:1029–1047.

Csösz, S., H. C. Wagner, M. Bozsó, B. Seifert, W. Arthofer, B. C. Schlick-Steiner, F. M. Steiner, and Z. Pénzes. 2014. *Tetramorium indocile* Santschi, 1927 stat. rev. is the proposed scientific name for *Tetramorium* sp. C sensu Schlick-Steiner *et al*. (2006) based on combined molecular and morphological evidence (Hymenoptera: Formicidae). Zool. Anz. 253:469–481.

D'Ettorre, P., J. Heinze, C. Schulz, W. Francke, and M. Ayasse. 2004. Does she smell like a queen? Chemoreception of a cuticular hydrocarbon signal in the ant *Pachycondyla inversa*. J. Exp. Biol. 207:1085–1091.

Dallerac, R., C. Labeur, J. M. Jallon, D. C. Knipple, W. L. Roelofs, and C. Wicker-Thomas. 2000. A ð9 desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. 97:9449–9454.

Darwell, C. T., and J. M. Cook. 2017a. Cryptic diversity in a fig wasp community - morphologically differentiated species are sympatric but cryptic species are parapatric. Mol. Ecol. 937–950.

Darwell, C. T., and J. M. Cook. 2017b. Cryptic diversity in a fig wasp community — morphologically differentiated species are sympatric but cryptic species are parapatric. Mol. Ecol. 26:937–950.

Darwin, C. 1859. On the origins of species by means of natural selection. London: Murray.

Dasmahapatra, K. K., J. R. Walters, A. D. Briscoe, J. W. Davey, A. Whibley, N. J. Nadeau, A. V. Zimin, D. S. T. Hughes, L. C. Ferguson, S. H. Martin, C. Salazar, J. J. Lewis, S. Adler, S. J. Ahn, D. A. Baker, S. W. Baxter, N. L. Chamberlain, C. Ritika, B. A. Counterman, T. Dalmay, L. E. Gilbert, K. Gordon, D. G. *et al.* 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94–98. Nature Publishing Group.

Davidson, D. W. 1988. Ecological Studies of Neotropical Ant Gardens. Ecology 69:1138–1152.

De Bustos, A., A. Cuadrado, and N. Jouve. 2016. Sequencing of long stretches of repetitive DNA. Sci. Rep., doi: 10.1038/srep36665. Nature Publishing Group.

De Queiroz, K. 2007. Species concepts and species delimitation. Syst. Biol. 56:879–886.

de Vienne, D. M., G. Refrégier, M. López-Villavicencio, A. Tellier, M. E. Hood, and T. Giraud. 2013. Cospeciation vs host-shift speciation: Methods for testing, evidence from natural associations and relation to coevolution. New Phytol. 198:347–385.

Degnan, P. H., A. B. Lazarus, C. D. Brock, and J. J. Wernegreen. 2004. Host-symbiont stability and fast evolutionary rates in an ant-bacterium association: Cospeciation of *Camponotus* species and their endosymbionts, *Candidatus blochmannia*. Syst. Biol. 53:95–110.

Delić, T., P. Trontelj, M. Rendoš, and C. Fišer. 2017. The importance of naming cryptic species and the conservation of endemic subterranean amphipods. Sci. Rep. 7:1–12.

Dembeck, L. M., K. Böröczky, W. Huang, C. Schal, R. R. H. Anholt, and T. F. C. Mackay. 2015. Genetic architecture of natural variation in *Drosophila melanogaster* aggressive behavior. Elife 4.

Dennis, A. B., L. T. Dunning, B. J. Sinclair, and T. R. Buckley. 2015. Parallel molecular routes to cold adaptation in eight genera of New Zealand stick insects. Sci. Rep. 5:1–13.

Deutsch, C. A., J. J. Tewksbury, R. B. Huey, K. S. Sheldon, C. K. Ghalambor, D. C. Haak, and P. R. Martin. 2008. Impacts of climate warming on terrestrial ectotherms across latitude. Proc. Natl. Acad. Sci. U. S. A. 105:6668–6672.

Didan, K. 2015. MOD13A1 MODIS/Terra Vegetation Indices 16-Day L3 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC.

Dieckmann, U., and M. Doebeli. 1999. On the origin of speciesby sympatric speciation. Nature 400:354–357.

Dietemann, V., C. Peeters, J. Liebig, V. Thivet, and B. Holldobler. 2003. Cuticular hydrocarbons mediate discrimination of reproductives and nonreproductives in the ant *Myrmecia gulosa*. Proc. Natl. Acad. Sci. 100:10341–10346.

Dimiceli, C., M. Carroll, R. Sohlberg, D. H. Kim, M. Kelly, and J. R. G. Townshend. 2015. MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC.

Doebeli, M., and U. Dieckmann. 2000. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. Am. Nat. 156:S77–S101.

Dohmen, E., L. P. M. Kremer, E. Bornberg-Bauer, and C. Kemena. 2016. DOGMA: Domain-based transcriptome and proteome quality assessment. Bioinformatics 32:2577–2581.

Doležel, J., J. Bartoš, H. Voglmayr, and J. Greilhuber. 2003. Nuclear DNA content and genome size of trout and human. Cytom. Part A 51 (2):127–128.

Doležel, J., M. Doleželová, and F. J. Novák. 1994. Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). Biol. Plant. 36:351–357.

Dreiss, A. N., S. Antoniazza, R. Burri, L. Fumagalli, C. Sonnay, C. Frey, J. Goudet, and A. Roulin. 2012. Local adaptation and matching habitat choice in female barn owls with respect to melanic coloration. J. Evol. Biol. 25:103–114.

El-Gebali, S., J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. E. Tosatto, and R. D. Finn. 2019. The Pfam protein families database in 2019. Nucleic Acids Res. 47:D427–D432.

Elia, M., A. Khalil, A. G. Bagnères, and M. C. Lorenzi. 2018. Appeasing their hosts: a novel strategy for parasite brood. Anim. Behav. 146:123–134.

Elmer, K. R., and A. Meyer. 2011. Adaptation in the age of ecological genomics: Insights from parallelism and convergence. Trends Ecol. Evol. 26:298–306.

Emery, V. J., and N. D. Tsutsui. 2013. Recognition in a Social Symbiosis: Chemical Phenotypes and Nestmate Recognition Behaviors of Neotropical Parabiotic Ants. PLoS One 8.

Enriquez, T., and H. Colinet. 2019. Cold acclimation triggers lipidomic and metabolic adjustments in the spotted wing drosophila *Drosophila suzukii* (Matsumara). Am. J. Physiol. - Regul. Integr. Comp. Physiol. 316:R751–R763.

Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564–567.

Fahrenholz, H. 1913. Ectoparasiten und Abstammungslehre. Zool. Anz. 41:371–374.

Falcón, T., M. J. Ferreira-Caliman, F. M. Franco Nunes, É. D. Tanaka, F. S. do Nascimento, and M. M. Gentile Bitondi. 2014. Exoskeleton formation in *Apis mellifera*: Cuticular hydrocarbons profiles and expression of desaturase and elongase genes during pupal and adult development. Insect Biochem. Mol. Biol. 50:68–81. Elsevier Ltd.

Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350. Elsevier Ltd.

Feldhaar, H., J. Gadau, and B. Fiala. 2010. Speciation in Obligately Plant-Associated *Crematogaster* Ants: Host Distribution Rather than Adaption Towards Specific Hosts Drives the Process.

Feldmeyer, B., D. Elsner, A. Alleman, and S. Foitzik. 2017. Species-specific genes under selection characterize the co-evolution of slavemaker and host lifestyles. BMC Evol. Biol. 17:1–11. BMC Evolutionary Biology.

Fennessy, J., T. Bidon, F. Reuss, V. Kumar, P. Elkan, M. A. Nilsson, M. Vamberger, U. Fritz, and A. Janke. 2016. Multi-locus Analyses Reveal Four Giraffe Species Instead of One. Curr. Biol. 26:2543–2549. Elsevier Ltd.

Finck, J., E. L. Berdan, F. Mayer, B. Ronacher, S. F. Geiselhardt, H. Chung, S. B. Carroll, T. L. Singer, J. F. Ferveur, R. W. Howard, G. J. Blomquist, B. G. Johansson, T. M. Jones, S. Martin, F. Drijfhout, S. F. Geiselhardt, S. F. Geiselhardt, K. Peschke, Y. Qiu, H. Chung, S. S. Chirala, S. J. Wakil, S. Smith, S.-C. Tsai, K. Hashimoto, A. Wagner, N. Gompel, B. Prud'homme, P. J. Wittkopp, V. A. Kassner, S. B. *et al*.. 2016. Divergence of cuticular hydrocarbons in two sympatric grasshopper species and the evolution of fatty acid synthases and elongases across insects. Sci. Rep. *6* VN-re:*33695*. Nature Publishing Group.

Finet, C., K. Slavik, J. Pu, S. B. Carroll, and H. Chung. 2019. Birth-and-Death Evolution of the Fatty Acyl-CoA Reductase (FAR) Gene Family and Diversification of Cuticular Hydrocarbon Synthesis in *Drosophila*. Genome Biol. Evol. 11:1541–1551.

Fišer, C., C. T. Robinson, and F. Malard. 2018. Cryptic species as a window into the paradigm shift of the species concept. Mol. Ecol. 27:613–635.

Forel, A. 1898. La parabiose chez les fourmis. Bull. la Société Vaudoise des Sci. Nat. 34:380–384.

Fournier, D., M. Tindo, M. Kenne, P. S. Mbenoun Masse, V. van Bossche, E. de Coninck, and S. Aron. 2012. Genetic structure, nestmate recognition and behaviour of two cryptic species of the invasive big-headed ant *Pheidole megacephala*. PLoS One 7.

Frentiu, F. D., and S. F. Chenoweth. 2010. Clines in cuticular hydrocarbons in two *Drosophila* species with independent population histories. Evolution (N. Y). 64:1784–1794.

Funk, D. J., P. Nosil, and W. J. Etges. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. Proc. Natl. Acad. Sci. U. S. A. 103:3209–3213.

Gadau, J., M. Helmkampf, S. Nygaard, J. Roux, D. F. Simola, C. R. Smith, G. Suen, Y. Wurm, and C. D. Smith. 2012. The genomic impact of 100 million years of social evolution in seven ant species. Trends Genet. 28:14–21.

García-Robledo, C., E. K. Kuprewicz, C. L. Staines, T. L. Erwin, and W. J. Kress. 2015. Limited

tolerance by insects to high temperatures across tropical elevational gradients and the implications of global warming for extinction. Proc. Natl. Acad. Sci. 113:680–685.

Gause, G. F. 1932. Experimental Studies on the Struggle for Existence I. Mixed Population of two Species of Yeast. J. Exp. Biol. 9:389–402.

Gautier, M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. Genetics 201:1555–1579.

Gebiola, M., M. M. Monti, R. C. Johnson, J. B. Woolley, M. S. Hunter, M. Giorgini, and P. A. Pedata. 2017. A revision of the *Encarsia pergandiella* species complex (Hymenoptera: Aphelinidae) shows cryptic diversity in parasitoids of whitefly pests. Syst. Entomol. 42:31–59.

Gibbons, T. C., S. M. Rudman, and P. M. Schulte. 2017. Low temperature and low salinity drive putatively adaptive growth differences in populations of threespine stickleback. Sci. Rep. 7:1–9.

Gibbs, A. G., A. K. Chippindale, and M. R. Rose. 1997. Physiological mechanisms of evolved desiccation resistance in *Drosophila melanogaster*. J. Exp. Biol. 200:1821–1832.

Gibbs, A. G., and J. G. Pomonis. 1995. Physical properties of insect cuticular hydrocarbons: The effects of chain length, methyl-branching and unsaturation. Comp. Biochem. Physiol. 112B:243–249.

Giglio, L., and C. Justice. 2015. MOD14A2 MODIS/Terra Thermal Anomalies/Fire 8-Day L3 Global 1km SIN Grid V006. 2015, distributed by NASA EOSDIS Land Processes DAAC.

Goedbloed, D. J., T. Czypionka, J. Altmüller, A. Rodriguez, E. Küpfer, O. Segev, L. Blaustein, A. R. Templeton, A. W. Nolte, and S. Steinfartz. 2017. Parallel habitat acclimatization is realized by the expression of different genes in two closely related salamander species (genus *Salamandra*). Heredity (Edinb). 119:429–437.

Goriely, A. 2016. Decoding germline de novo point mutations. Nat. Genet. 48:823–824. Nature Publishing Group.

Gould, S. J. 1974. The Origin and Function of "Bizarre" Structures: Antler Size and Skull Size in the "Irish Elk," *Megaloceros giganteus*. Soc. Study Evol. 28:191–220.

Gould, S. J. 1989. Wonderful Life: The Burgess Shale and the Nature of History. W.W. Norton & Company, New York.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2013. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29:644–652.

Grant, B. R., and P. R. Grant. 2003. What Darwin's Finches Can Teach Us about the Evolutionary Origin and Regulation of Biodiversity. Bioscience 53:965.

Greene, M. J., and D. M. Gordon. 2003. Cuticular hydrocarbons inform task decision. Nature 423:2190.

Greilhuber, J., J. Doležel, M. A. Lysák, and M. D. Bennett. 2005. The Origin , Evolution and Proposed Stabilization of the Terms 'Genome Size' and 'C-Value' to Describe Nuclear DNA Contents. Ann. Bot. 255–260.

Grundt, H. H., S. Kjolner, L. Borgen, L. H. Rieseberg, and C. Brochmann. 2006. High biological species diversity in the arctic flora. Proc. Natl. Acad. Sci. 103:972–975.

Guimarães, P. R., P. Jordano, and J. N. Thompson. 2011. Evolution and coevolution in mutualistic networks. Ecol. Lett. 14:877–885.

Günther, T., and G. Coop. 2013. Robust identification of local adaptation from allele frequencies. Genetics 195:205–220.

Gustafsson, A. L. S., I. Skrede, H. C. Rowe, G. Gussarova, L. Borgen, L. H. Rieseberg, C. Brochmann, and C. Parisod. 2014. Genetics of cryptic speciation within an arctic Mustard, *Draba nivalis*. PLoS One 9.

Hackl, T., R. Hedrich, J. Schultz, and F. Förster. 2014. Proovread: Large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics 30:3004–3011.

Hadley, N. F. 1981. Cuticular Lipids of Terrestrial Plants and Arthropods: A Comparison of Their Structure, Composition, and Waterproofing Function. Biol. Rev. 56:23–47.

Haenel, Q., M. Roesti, D. Moser, A. D. C. MacColl, and D. Berner. 2019. Predictable genome-wide sorting of standing genetic variation during parallel adaptation to basic versus acidic environments in stickleback fish. Evol. Lett. 3:28–42.

Hahn, C., L. Bachmann, and B. Chevreux. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads — a baiting and iterative mapping approach. Nucleic Acids Res. 41.

Halbritter, A. H., J. M. Alexander, P. J. Edwards, and R. Billeter. 2013. How comparable are species distributions along elevational and latitudinal climate gradients? Glob. Ecol. Biogeogr. 22:1228–1237.

Han, F., S. Lamichhaney, B. Rosemary Grant, P. R. Grant, L. Andersson, and M. T. Webster. 2017. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. Genome Res. 27:1004–1015.

Han, M. V., and C. M. Zmasek. 2009. PhyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics 10:1–6.

Hardin, G. 1960. The Competitive Exclusion Principle. Science (80-.). 131:1292–1297.

Hartke, J., P. P. Sprenger, J. Sahm, H. Winterberg, J. Orivel, H. Baur, T. Beuerle, T. Schmitt, B. Feldmeyer, and F. Menzel. 2019a. Cuticular hydrocarbons as potential mediators of cryptic species divergence in a mutualistic ant association. Ecol. Evol. 9:1–17.

Hartke, J., T. Schell, E. Jongepier, H. Schmidt, P. P. Sprenger, and B. Feldmeyer. 2019b. Hybrid genome assembly of a neotropical mutualistic ant. Genome Biol. Evol. 11:2306–2311.

He, B., and P. N. Adler. 2002. The genetic control of arista lateral morphogenesis in *Drosophila*. Dev. Genes Evol. 212:218–229.

Heethoff, M., M. Laumann, G. Weigmann, and G. Raspotnig. 2011. Integrative taxonomy: Combining morphological, molecular and chemical data for species delineation in the parthenogenetic *Trhypochthonius tectorum* complex (Acari, Oribatida, Trhypochthoniidae). Front. Zool. 8:2. BioMed Central Ltd.

Helmkampf, M., E. Cash, and J. Gadau. 2015. Evolution of the Insect Desaturase Gene Family with an Emphasis on Social Hymenoptera. Mol. Biol. Evol. 32:456–471.

Hennig, C. 2007. Cluster-wise assessment of cluster stability. Comput. Stat. Data Anal. 52:258–271.

Henry, P., and M. A. Russello. 2013. Adaptive divergence along environmental gradients in a climate-change-sensitive mammal. Ecol. Evol. 3:3906–3917.

Herre, E. A., N. Knowlton, U. G. Mueller, and S. A. Rehner. 1999. The evolution of mutualisms: exploring the paths between conflict and cooperation. Trends Ecol. Evol. 14:49–53.

Higashi, M., G. Takimoto, and N. Yamamura. 1999. Sympatric speciation by sexual selection. Nature 402:523–526.

Higgins, P. A. T. 2007. Biodiversity loss under existing land use and climate change: An illustration using northern South America. Glob. Ecol. Biogeogr. 16:197–204.

Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd, D. B. Lowry, M. L. Poss, L. K. Reed, A. Storfer, and M. C. Whitlock. 2016. Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. Am. Nat. 188:379–397.

Hoeksema, J. D., and E. M. Bruna. 2000. Pursuing the big questions about interspecific mutualism: A review of theoretical approaches. Oecologia 125:321–330.

Hoffmann, A. A. 1990. Acclimation for desiccation resistance in *Drosophila melanogaster* and the association between acclimation responses and genetic variation. J. Insect Physiol. 36:885–891.

Hoffmann, A. A., M. Turelli, and G. M. Simmons. 1986. Unidirectional Incompatibility Between Populations of *Drosophila simulans*. Evolution (N. Y). 40:692–701.

Holt, C., and M. Yandell. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12:491.

Hosokawa, T., Y. Kikuchi, N. Nikoh, M. Shimada, and T. Fukatsu. 2006. Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. PLoS Biol. 4:1841–1851.

Hubbell, S. P. 2005. Neutral theory in community ecology and the hypothesis of functional equivalence. Funct. Ecol. 19:166–172.

Hubbell, S. P. 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press.

Hudson, E. J., and T. D. Price. 2014. Pervasive reinforcement and the role of sexual selection in biological speciation. J. Hered. 105:821–833.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Huerta-Sánchez, E., X. Jin, Asan, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, P. Ni, B. Wang, X. Ou, Huasang, J. Luosang, Z. X. P. Cuo, K. Li, G. Gao, Y. Yin, W. Wang, X. Zhang, X. Xu, H. Yang, Y. Li, J. Wang, J. Wang, and R. Nielsen. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512:194–197.

Hume, J. P., and D. Martill. 2019. Repeated evolution of flightlessness in *Dryolimnas* rails (Aves: Rallidae) after extinction and recolonization on Aldabra. Zool. J. Linn. Soc. 000:666–672.

Ingram, V. M. 1957. Gene mutations in human haemoglobin: The chemical differences between normal and sickle cell haemoglobin. Nature 180:326–328.

Itino, T., S. J. Davies, H. Tada, Y. Hieda, M. Inoguchi, T. Itioka, S. Yamane, and T. Inoue. 2001. Cospeciation of ants and plants. Ecol. Res. 16:787–793.

Ivory-Church, J., F. D. Frentiu, and S. F. Chenoweth. 2015. Polymorphisms in a desaturase 2 ortholog associate with cuticular hydrocarbon and male mating success variation in a natural population of *Drosophila serrata*. J. Evol. Biol. 28:1600–1609.

Janz, N., K. Nyblom, and S. Nylin. 2001. Evolutionary Dynamics of Host-Plant Specialization: a Case Study of the Tribe Nymphalini. Evolution (N. Y). 55:783.

Janzen, D. H. 1966. Coevolution of Mutualism Between Ants and Acacias in Central America. Evolution (N. Y). 20:249–275.

Janzen, D. H. 1985. On Ecological Fitting. Oikos 45:308.

Janzen, D. H. 1979. When is it Coevolution ? Evolution (N. Y). 34:611–612.

Janzen, D. H., J. M. Burns, Q. Cong, W. Hallwachs, T. Dapkey, R. Manjunath, M. Hajibabaei, P. D. N. Hebert, and N. V. Grishin. 2017. Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology. Proc. Natl. Acad. Sci. U. S. A. 114:8313–8318.

Jeffreys, H. 1961. Theory of Probability. 3rd ed. Clarendon Press, Oxford.

Jiang, N., L. Wang, J. Chen, L. Wang, L. Leach, and Z. Luo. 2014. Conserved and divergent patterns of DNA methylation in higher vertebrates. Genome Biol. Evol. 6:2998–3014.

Jiggins, C. D., R. E. Naisbit, R. L. Coe, and J. Mallet. 2001. Reproductive isolation caused by colour pattern mimicry. Nature 411:302–305.

Jongepier, E., and S. Foitzik. 2016. Ant recognition cue diversity is higher in the presence of slavemaker ants. Behav. Ecol. 27:304–311.

Jousselin, E., S. Van Noort, V. Berry, J. Y. Rasplus, N. Rønsted, J. C. Erasmus, and J. M. Greeff. 2008. One fig to bind them all: Host conservatism in a fig wasp community unraveled by cospeciation analyses among pollinating and nonpollinating fig wasps. Evolution (N. Y). 62:1777–1797.

Kaeuffer, R., C. L. Peichel, D. I. Bolnick, and A. P. Hendry. 2011. Parallel and nonparallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. Evolution (N. Y). 66:402–418.

Kamilar, J. M., and N. Cooper. 2013. Phylogenetic signal in primate behaviour, ecology and life history. Philos. Trans. R. Soc. B 368:20120341.

Kang, C., T. N. Sherratt, Y. E. Kim, Y. Shin, J. Moon, U. Song, J. Y. Kang, K. Kim, and Y. Jang. 2017. Differential predation drives the geographical divergence in multiple traits in aposematic frogs. Behav. Ecol. 28:1122–1130.

Karger, D. N., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, H. P. Linder, and M. Kessler. 2017. Climatologies at high resolution for the earth's land surface areas. Sci. Data 4:1–20.

Kassambara, A., and F. Mundt. 2016. Factoextra: extract and visualize the results of multivariate data

analyses. R Packag. version 1:2016.

Kaur, R., M. Stoldt, E. Jongepier, B. Feldmeyer, F. Menzel, E. Bornberg-Bauer, and S. Foitzik. 2019. Ant behaviour and brain gene expression of defending hosts depend on the ecological success of the intruding social parasite. Philos. Trans. R. Soc. B Biol. Sci. 374.

Kawakita, A., A. Takimura, T. Terachi, T. Sota, and M. Kato. 2004. Cospeciation Analysis of an Obligate Pollination Mutualism: Have *Glochidion* Trees (Euphorbiaceae) and Pollinating *Epicephala* Moths (Gracillariidae) Diversified in Parallel? Evolution (N. Y). 58:2201.

Kefi, M., V. Balabanidou, V. Douris, G. Lycett, R. Feyereisen, and J. Vontas. 2019. Two functionally distinct CYP4G genes of *Anopheles gambiae* contribute to cuticular hydrocarbon biosynthesis. Insect Biochem. Mol. Biol. 110:52–59.

Keilwagen, J., M. Wenk, J. L. Erickson, M. H. Schattat, J. Grau, and F. Hartung. 2016. Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. 44.

Kim, D., B. Langmead, and S. L. Salzberg. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12.

Kirkpatrick, M., and V. Ravigne. 2012. Speciation by Natural and Sexual Selection: Models and Experiments. Author(s): Mark Kirkpatrick and Virginie Ravigné Reviewed work(s): Source: The. 159.

Kleeberg, I., F. Menzel, and S. Foitzik. 2017. The influence of slavemaking lifestyle, caste and sex on chemical profiles in *Temnothorax* ants: Insights into the evolution of cuticular hydrocarbons. Proc. R. Soc. B Biol. Sci. 284.

Kleidon, A., and S. Lorenz. 2001. Deep roots sustain Amzonian rainforest in climate model simulations of the last ice age. Geophys. Res. Lett. 28:2425–2428.

Klepsatel, P., M. Gáliková, C. D. Huber, and T. Flatt. 2014. Similarities and differences in altitudinal versus latitudinal variation for morphological traits in *Drosophila melanogaster*. Evolution (N. Y). 68:1385–1398.

Klingenberg, C. P. 2016. Size, shape, and form: concepts of allometry in geometric morphometrics. Dev. Genes Evol. 226:113–137. Development Genes and Evolution.

Kofler, R., R. V. Pandey, and C. Schlötterer. 2011. PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27:3435–3436.

Kolaczkowski, B., A. D. Kern, A. K. Holloway, and D. J. Begun. 2011. Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. Genetics 187:245–260.

Konorov, E. A., M. A. Nikitin, K. V Mikhailov, S. N. Lysenkov, M. Belenky, P. L. Chang, S. V. Nuzhdin, and V. A. Scobeyeva. 2017. Genomic exaptation enables *Lasius niger* adaptation to urban environments. BMC Evol. Biol. 17:39. BMC Evolutionary Biology.

Korf, I. 2004. Gene finding in novel genomes. BMC Bioinformatics 9:1–9.

Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol. Biol. Evol. 35:1547–1549.

Lahav, S., V. Soroker, A. Hefetz, and R. Aviv. 1999. Direct Behavioral Evidence for Hydrocarbons as Ant Recognition Discriminators. Naturwissenschaften 86:246–249.

Lai, Y. T., C. K. L. Yeung, K. E. Omland, E. L. Pang, Y. Hao, B. Y. Liao, H. F. Cao, B. W. Zhang, C. F. Yeh, C. M. Hung, H. Y. Hung, M. Y. Yang, W. Liang, Y. C. Hsu, C. Te Yao, L. Dong, K. Lin, and S. H. Li. 2019. Standing genetic variation as the predominant source for adaptation of a songbird. Proc. Natl. Acad. Sci. U. S. A. 116:2152–2157.

Lande, R. 1981. Models of speciation by sexual selection on polygenic traits (mating preferences/sexual dimorphism/genetic correlation/runaway process). 78:3721–3725.

Leal, L. C., C. C. Jakovac, P. E. D. Bobrowiec, J. L. C. Camargo, and P. E. C. Peixoto. 2017. The role of parabiotic ants and environment on epiphyte composition and protection in ant gardens. Sociobiology 64:276–283.

Leavitt, D. H., J. Starrett, M. F. Westphal, and M. Hedin. 2015. Multilocus sequence data reveal dozens of putative cryptic species in a radiation of endemic Californian mygalomorph spiders (Araneae, Mygalomorphae, Nemesiidae). Mol. Phylogenet. Evol. 91:56–67.

Lee, E., G. A. Helt, J. T. Reese, M. C. Munoz-torres, C. P. Childers, R. M. Buels, L. Stein, I. H. Holmes, C. G. Elsik, and S. E. Lewis. 2013. Web Apollo: a web-based genomic annotation editing platform. Genome Biol. 14.

Lehmann, F. O. 1999. Ambient temperature affects free-flight performance in the fruit fly *Drosophila melanogaster*. J. Comp. Physiol. - B Biochem. Syst. Environ. Physiol. 169:165–171.

Leigh, J. W., and D. Bryant. 2015. POPART: Full-feature software for haplotype network construction. Methods Ecol. Evol. 6:1110–1116.

Lenoir, A., P. D'Ettorre, and C. Errard. 2001. Chemical Ecology and Social Parasitism in Ants. Annu. Rev. Entomol. 46:573–599.

Lenoir, A., C. Malosse, and R. Yamaoka. 1997. Chemical mimicry between parasitic ants of the genus *Formicoxenus* and their host Myrmica (Hymenoptera, Formicidae). Biochem. Syst. Ecol. 25:379–389.

Leonhardt, S. D., F. Menzel, V. Nehring, and T. Schmitt. 2016. Ecology and Evolution of Communication in Social Insects. Cell 164:1277–1287. Elsevier Inc.

Lewis, J. S., M. L. Farnsworth, C. L. Burdett, D. M. Theobald, M. Gray, and R. S. Miller. 2017. Biotic and abiotic factors predicting the global distribution and population density of an invasive large mammal. Sci. Rep. 7:1–12. Nature Publishing Group.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Li, H., H. Higashi, and K. Tamura. 2006. Estimation of boiling and melting points of light, heavy and complex hydrocarbons by means of a modified group vector space method. Fluid Phase Equilib. 239:213–222.

Li, Y. F., J. C. Costello, A. K. Holloway, and M. W. Hahn. 2008. "Reverse ecology" and the power of

population genomics. Evolution (N. Y). 62:2984–2994.

Liang, S., R. C. Zhou, S. S. Dong, and S. H. Shi. 2008. Adaptation to salinity in mangroves: Implication on the evolution of salt-tolerance. Chinese Sci. Bull. 53:1708–1715.

Liaw, A., and M. Wiener. 2002. Classification and Regression by randomForest. R News 2/3:18–22.

Liebig, J., C. Peeters, N. J. Oldham, C. Markstädter, and B. Hölldobler. 2000. Are variations in cuticular hydrocarbons of queens and workers a reliable signal of fertility in the ant *Harpegnathos saltator*? Proc. Natl. Acad. Sci. U. S. A. 97:4124–4131.

Lihoreau, M., and C. Rivault. 2009. Kin recognition via cuticular hydrocarbons shapes cockroach social life. Behav. Ecol. 20:46–53.

Liu, S., E. D. Lorenzen, M. Fumagalli, B. Li, K. Harris, Z. Xiong, L. Zhou, T. S. Korneliussen, M. Somel, C. Babbitt, G. Wray, J. Li, W. He, Z. Wang, W. Fu, X. Xiang, C. C. Morgan, A. Doherty, M. J. O'Connell, J. O. McInerney, *et al*. 2014. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. Cell 157:785–794.

Lomsadze, A., V. Ter-Hovhannisyan, Y. O. Chernoff, and M. Borodovsky. 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 33:6494–6506.

Longino, J. T. 2003. The *Crematogaster* (Hymenoptera, Formicidae, Myrmicinae) of Costa Rica. Zootaxa 151:1–150.

Luu, K., E. Bazin, and M. G. B. Blum. 2017. pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol. Ecol. Resour. 17:67–77.

Machado, C. A., E. A. Herre, S. McCafferty, and E. Bermingham. 1996. Molecular phylogenies of fig pollinating and non-pollinating wasps and the implications for the origin and evolution of the fig-fig wasp mutualism. J. Biogeogr. 23:531–542.

MacMillan, H. A., J. M. Knee, A. B. Dennis, H. Udaka, K. E. Marshall, T. J. S. Merritt, and B. J. Sinclair. 2016. Cold acclimation wholly reorganizes the *Drosophila melanogaster* transcriptome and metabolome. Sci. Rep. 6:1–14. Nature Publishing Group.

MacPherson, A., and S. L. Nuismer. 2017. The probability of parallel genetic evolution from standing genetic variation. J. Evol. Biol. 30:326–337.

Mallard, F., V. Nolte, R. Tobler, M. Kapun, and C. Schlötterer. 2018. A simple genetic basis of adaptation to a novel thermal environment results in complex metabolic rewiring in *Drosophila*. Genome Biol. 19:1–15. Genome Biology.

Mallet, J., A. Meyer, P. Nosil, and J. L. Feder. 2009. Space, sympatry and speciation. J. Evol. Biol. 22:2332–2341.

Maloiy, G. M. O., J. M. Z. Kamau, A. Shkolnik, M. Meir, and R. Arieli. 1982. Thermoregulation and metabolism in a small desert carnivore: the Fennec fix (*Fennecus zerda*) (Mammalia)*. J. Zool. 198:279–291.

Manenti, R., M. Denoël, and G. F. Ficetola. 2013. Foraging plasticity favours adaptation to new habitats in fire salamanders. Anim. Behav. 86:375–382. Elsevier Ltd.

Mani, G. S., and B. C. Clarke. 1990. Mutational order: A major stochastic process in evolution. Proc. R. Soc. B Biol. Sci. 240:29–37.

Marron, M. T., T. A. Markow, K. J. Kain, and A. G. Gibbs. 2003. Effects of starvation and desiccation on energy metabolism in desert and mesic *Drosophila*. J. Insect Physiol. 49:261–270.

Martin, S. J., and F. P. Drijfhout. 2009. A review of ant cuticular hydrocarbons. J. Chem. Ecol. 35:1151–1161.

Martin, S. J., H. Helanterä, and F. P. Drijfhout. 2008. Evolution of species-specific cuticular hydrocarbon patterns in *Formica* ants. Biol. J. Linn. Soc. 95:131–140.

Mayle, F. E., M. J. Burn, M. Power, and D. H. Urrego. 2009. Past Climate Variability in South America and Surrounding Regions. 14:89–112.

McKenzie, S. K., I. Fetter-Pruneda, V. Ruta, and D. J. C. Kronauer. 2016. Transcriptomics and neuroanatomy of the clonal raider ant implicate an expanded clade of odorant receptors in chemical communication. Proc. Natl. Acad. Sci. U. S. A. 113:14091–14096.

McKenzie, S. K., and D. J. C. Kronauer. 2018. The genomic architecture and molecular evolution of ant odorant receptors. Genome Res. 28:1757–1765.

Melo-Sampaio, P. R., P. Passos, A. Fouquet, A. L. D. C. Prudente, and O. Torres-Carvajal. 2019. Systematic review of *Atractus schach* (Serpentes: Dipsadidae) species complex from the Guiana Shield with description of three new species. Syst. Biodivers. 17:207–229. Taylor & Francis.

Menzel, F., B. B. Blaimer, and T. Schmitt. 2017a. How do cuticular hydrocarbons evolve? Physiological constraints and climatic and biotic selection pressures act on a complex functional trait. Proc. R. Soc. B 284:20161727.

Menzel, F., and N. Blüthgen. 2010. Parabiotic associations between tropical ants: Equal partnership or parasitic exploitation? J. Anim. Ecol. 79:71–81.

Menzel, F., N. Blüthgen, and T. Schmitt. 2008a. Tropical parabiotic ants: Highly unusual cuticular substances and low interspecific discrimination. Front. Zool. 5:16.

Menzel, F., N. Blüthgen, T. Tolasch, J. Conrad, U. Beifuß, T. Beuerle, and T. Schmitt. 2013. Crematoenones - a novel substance class exhibited by ants functions as appeasement signal. Front. Zool. 10:1–12.

Menzel, F., K. E. Linsenmair, and N. Blüthgen. 2008b. Selective interspecific tolerance in tropical *Crematogaster-Camponotus* associations. Anim. Behav. 75:837–846.

Menzel, F., S. Morsbach, J. H. Martens, P. Räder, S. Hadjaje, M. Poizat, and B. Abou. 2019. Communication vs. waterproofing: the physics of insect cuticular hydrocarbons. J. Exp. Biol. 222.

Menzel, F., J. Orivel, M. Kaltenpoth, and T. Schmitt. 2014. What makes you a potential partner? Insights from convergently evolved ant-ant symbioses. Chemoecology 24:105–119.

Menzel, F., and T. Schmitt. 2012. Tolerance requires the right smell: First evidence for interspecific selection on chemical recognition cues. Evolution (N. Y). 66:896–904.

Menzel, F., T. Schmitt, and B. B. Blaimer. 2017b. The evolution of a complex trait: cuticular hydrocarbons in ants evolve independent from phylogenetic constraints. J. Evol. Biol. 30:1372–1385.

Menzel, F., M. Zumbusch, and B. Feldmeyer. 2018. How ants acclimate: Impact of climatic

conditions on the cuticular hydrocarbon profile. Funct. Ecol. 32:657–666.

Mercer, K., Á. Martínez-Vásquez, and H. R. Perales. 2008. Asymmetrical local adaptation of maize landraces along an altitudinal gradient. Evol. Appl. 1:489–500.

Mikheyev, A. S., and T. A. Linksvayer. 2015. Genes associated with ant social behavior show distinct transcriptional and evolutionary patterns. Elife 2015:1–17.

Minard, G., V. Tran Van, F. H. Tran, C. Melaun, S. Klimpel, L. K. Koch, K. Ly Huynh Kim, T. Huynh Thi Thuy, H. Tran Ngoc, P. Potier, P. Mavingui, and C. Valiente Moro. 2017. Identification of sympatric cryptic species of *Aedes albopictus* subgroup in Vietnam: new perspectives in phylosymbiosis of insect vector. Parasites and Vectors 10:1–14. Parasites & Vectors.

Modesti, M., and R. Kanaar. 2001. Homologous recombination: From model organisms to human disease. Genome Biol. 2:1–5.

Montero-Pau, J., A. Gomez, and J. Muñoz. 2008. Application of an inexpensive and high-throughput genomic DNA extraction method for the molecular ecology of zooplanktonic diapausing eggs. Limnol. Oceanogr.: Methods 6, 2008, 218–222 - 0218.pdf. Limnol. Oceanogr. Methods 6:218–222.

Mori, A., R. Visicchio, M. F. Sledge, D. A. Grasso, F. Le Moli, S. Turillazzi, S. Spencer, and G. R. Jones. 2000. Behavioural assays testing the appeasement allomone of *Polyergus rufescens* queens during host-colony usurpation. Ethol. Ecol. Evol. 12:315–322.

Morris, S. C. 2010. Evolution: Like any other science it is predictable. Philos. Trans. R. Soc. B Biol. Sci. 365:133–145.

Neupert, S., A. DeMilto, F. Drijfhout, S. Speller, and R. M. M. Adams. 2018. Host colony integration: *Megalomyrmex* guest ant parasites maintain peace with their host using weaponry. Anim. Behav. 139:71–79. Elsevier Ltd.

Noonan, B. P., and P. Gaucher. 2005. Phylogeography and demography of Guianan harlequin toads (Atelopus): Diversification within a refuge. Mol. Ecol. 14:3017–3031.

Norman, V., H. Darras, C. Tranter, S. Aron, and W. O. H. Hughes. 2016. Cryptic lineages hybridize for worker production in the harvester ant *Messor barbarus*. Biol. Lett. 12.

Norris, L. C., B. J. Main, Y. Lee, T. C. Collier, A. Fofana, A. J. Cornel, and G. C. Lanzaro. 2015. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. Proc. Natl. Acad. Sci. U. S. A. 112:815–820.

Nosil, P. 2012. Ecological speciation. Oxford University Press, Oxford, UK.

Novák, P., P. Neumann, J. Pech, J. Steinhaisl, and J. Macas. 2013. RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29:792–793.

Nuismer, S. L., A. Macpherson, and E. B. Rosenblum. 2012. Crossing the threshold: Gene flow, dominance and the critical level of standing genetic variation required for adaptation to novel environments. J. Evol. Biol. 25:2665–2671.

Nygaard, S., H. Hu, C. Li, M. Schiøtt, Z. Chen, Z. Yang, Q. Xie, C. Ma, Y. Deng, R. B. Dikow, C. Rabeling, D. R. Nash, W. T. Wcislo, S. G. Brady, T. R. Schultz, G. Zhang, and J. J. Boomsma.

2016. Reciprocal genomic evolution in the ant-fungus agricultural symbiosis. Nat. Commun. 7:1–9.

Nygaard, S., G. Zhang, M. Schiøtt, C. Li, Y. Wurm, H. Hu, J. Zhou, L. Ji, F. Qiu, M. Rasmussen, H. Pan, F. Hauser, A. Krogh, C. J. P. Grimmelikhuijzen, J. Wang, and J. J. Boomsma. 2011. The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. Genome Res. 21:1–10.

Oke, K. B., G. Rolshausen, C. LeBlond, and A. P. Hendry. 2017. How parallel is parallel evolution? A comparative analysis in fishes. Am. Nat. 190.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2016. vegan: Community Ecology Package.

Orivel, J., C. Errard, and A. Dejean. 1997. Ant gardens: Interspecific recognition in parabiotic ant species. Behav. Ecol. Sociobiol. 40:87–93.

Orivel, J., and C. Leroy. 2010. The diversity and ecology of ant gardens (Hymenoptera: Formicidae; Spermatophyta: Angiospermae). Myrmecological News 14:73–85.

Orr, H. A. 2005. The probability of parallel adaptation. Evolution (N. Y). 59:216–220.

Otte, T., M. Hilker, and S. Geiselhardt. 2018. Phenotypic Plasticity of Cuticular Hydrocarbon Profiles in Insects. J. Chem. Ecol. 44:235–247. Journal of Chemical Ecology.

Otto, F. 1990. DAPI Staining of Fixed Cells for High-Resolution Flow Cytometry of Nuclear DNA. In: Darzynkiewickz Z, Crissman HA, eds. Methods Cell Biol. 33:105–110.

Oxley, P. R., L. Ji, I. Fetter-Pruneda, S. K. McKenzie, C. Li, H. Hu, G. Zhang, and D. J. C. Kronauer. 2014. The genome of the Clonal raider ant *Cerapachys biroi*. Curr. Biol. 24:451–458.

Pamminger, T., S. Foitzik, K. C. Kaufmann, N. Schützler, and F. Menzel. 2014. Worker personality and its association with spatially structured division of labor. PLoS One 9:1–8.

Paradis, E. 2010. pegas: An R package for population genetics with an integrated-modular approach. Bioinformatics 26:419–420.

Parkash, R., D. D. Aggarwal, D. Singh, C. Lambhod, and P. Ranga. 2013. Divergence of water balance mechanisms in two sibling species (*Drosophila simulans* and *D. melanogaster*): Effects of growth temperatures. J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 183:359–378.

Parkash, R., D. Singh, and C. Lambhod. 2014. Sex-specific differences in desiccation resistance and the use of energy metabolites as osmolytes in *Drosophila melanogaster* flies acclimated to dehydration stress. J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 184:193–204.

Parmentier, T., K. Yéo, W. Dekoninck, and T. Wenseleers. 2017. An apparent mutualism between Afrotropical ant species sharing the same nest. Behav. Ecol. Sociobiol. 71. Behavioral Ecology and Sociobiology.

Parra, G., K. Bradnam, and I. Korf. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061–1067.

Pask, G. M., J. D. Slone, J. G. Millar, P. Das, J. A. Moreira, X. Zhou, J. Bello, S. L. Berger, R. Bonasio, C. Desplan, D. Reinberg, J. Liebig, L. J. Zwiebel, and A. Ray. 2017. Specialized odorant

receptors in social insects that detect cuticular hydrocarbon cues and candidate pheromones. Nat. Commun. 8:1–10. Springer US.

Patalano, S., A. Vlasova, C. Wyatt, P. Ewels, F. Camara, P. G. Ferreira, C. L. Asher, T. P. Jurkowski, A. Segonds-Pichon, M. Bachman, I. González-Navarrete, A. E. Minoche, F. Krueger, E. Lowy, M. Marcet-Houben, J. L. Rodriguez-Ales, F. S. Nascimento, S. Balasubramanian, T. Gabaldon, J. E. Tarver, *et al.* 2015. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. Proc. Natl. Acad. Sci. 112:13970–13975.

Pérez-Lachaud, G., J. C. Bartolo-Reyes, C. M. Quiroa-Montalván, L. Cruz-López, A. Lenoir, and J. P. Lachaud. 2015. How to escape from the host nest: Imperfect chemical mimicry in eucharitid parasitoids and exploitation of the ants' hygienic behavior. J. Insect Physiol. 75:63–72.

Petfield, D., S. F. Chenoweth, H. D. Rundle, and M. W. Blows. 2005. Genetic variance in female condition predicts indirect genetic variance in male sexual display traits. Proc. Natl. Acad. Sci. U. S. A. 102:6045–6050.

Petit, R. J., and L. Excoffier. 2009. Gene flow and species delimitation. Trends Ecol. Evol. 24:386–393.

Pfenninger, M., S. Patel, L. Arias-Rodriguez, B. Feldmeyer, R. Riesch, and M. Plath. 2015. Unique evolutionary trajectories in repeated adaptation to hydrogen sulphide-toxic habitats of a neotropical fish (*Poecilia mexicana*). Mol. Ecol. 24:5446–5459.

Pfenninger, M., and K. Schwenk. 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. BMC Evol. Biol. 7:1–6.

Podos, J., and S. Nowicki. 2004. Beaks, Adaptation, and Vocal Evolution in Darwin's Finches. Bioscience 54:501.

Potter, S., J. G. Bragg, M. P. K. Blom, J. E. Deakin, M. Kirkpatrick, M. D. B. Eldridge, and C. Moritz. 2017. Chromosomal speciation in the genomics era: Disentangling phylogenetic evolution of rock-wallabies. Front. Genet. 8:1–18.

Price, M. N., P. S. Dehal, and A. P. Arkin. 2009. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Mol. Bol. Evol. 26:1641–1650.

Pringle, E. G., S. R. Ramírez, T. C. Bonebrake, D. M. Gordon, and R. Dirzo. 2012. Diversification and phylogeographic structure in widespread *Azteca* plant-ants from the northern Neotropics. Mol. Ecol. 21:3576–3592.

Punta, M., P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. L. Sonnhammer, S. R. Eddy, A. Bateman, and R. D. Finn. 2012. The Pfam protein families database. Nucleic Acids Res. 40:290–301.

Pyhäjärvi, T., M. B. Hufford, S. Mezmouk, and J. Ross-Ibarra. 2013. Complex Patterns of Local Adaptation in Teosinte. Genome Biol. Evol. 5:1594–1609.

Qiu, Y., C. Tittiger, C. Wicker-Thomas, G. Le Goff, S. Young, E. Wajnberg, T. Fricaux, N. Taquet, G. J. Blomquist, and R. Feyereisen. 2012. An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. Proc. Natl. Acad. Sci. U. S. A. 109:14858–14863.

Quek, S. P., S. J. Davies, T. Itino, and N. E. Pierce. 2004. Codiversification in an ant-plant mutualism: Stem texture and the evolution of host use in *Crematogaster* (Formicidae: Myrmicinae)

inhabitants of Macaranga (Euphorbiaceae). Evolution (N. Y). 58:554–570.

Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, and R. Lopez. 2005. InterProScan: Protein domains identifier. Nucleic Acids Res. 33:116–120.

Quinlan, A. R., and I. M. Hall. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Qvarnström, A., M. Ålund, S. E. Mcfarlane, and P. M. Sirkiä. 2016. Climate adaptation and speciation: Particular focus on reproductive barriers in *Ficedula* flycatchers. Evol. Appl. 9:119–134.

R Core, T. 2018. R Foundation for Statistical Computing; Vienna, Austria: 2014. R A Lang. Environ. Stat. Comput. 2013.

Rabosky, D. L., and I. J. Lovette. 2008. Density-dependent diversification in North American wood warblers. Proc. R. Soc. B Biol. Sci. 275:2363–2371.

Rajpurohit, S., R. Hanus, V. Vrkoslav, E. L. Behrman, A. O. Bergland, D. Petrov, J. Cvačka, and P. S. Schmidt. 2017. Adaptive dynamics of cuticular hydrocarbons in *Drosophila*. J. Evol. Biol. 30:66–80.

Rajpurohit, S., C. C. Oliveira, W. J. Etges, and A. G. Gibbs. 2013. Functional genomic and phenotypic responses to desiccation in natural populations of a desert drosophilid. Mol. Ecol. 22:2698–2715.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. 67:901–904.

Rand, A. S. 1985. Tradeoffs in the evolution of frog calls. Proc. Anim. Sci. 94:623–637.

Randall, J. A. 1993. Bahavioural adaptations of desert rodents (Heteromyidae).

Rane, R. V., L. Rako, M. Kapun, S. F. Lee, and A. A. Hoffmann. 2015. Genomic evidence for role of inversion 3RP of *Drosophila melanogaster* in facilitating climate change adaptation. Mol. Ecol. 24:2423–2432.

Ranson, H., B. Jensen, J. M. Vulule, X. Wang, J. Hemingway, and F. H. Collins. 2000. Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. Insect Mol. Biol. 9:491–497.

Reeve, H. K., and P. W. Sherman. 1993. Adaptation and the Goals of Evolutionary Research. Q. Rev. Biol. 68:1–32.

Reidenbach, K. R., C. Cheng, F. Liu, C. Liu, N. J. Besansky, and Z. Syed. 2014. Cuticular differences associated with aridity acclimation in African malaria vectors carrying alternative arrangements of inversion 2La. Parasites and Vectors 7:1–13.

Revell, L. J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.

Rieseberg, L. H., and J. H. Willis. 2007. Plant speciation. Science. 317:910–914.

Rinehart, J. P., A. Li, G. D. Yocum, R. M. Robich, S. A. L. Hayward, and D. L. Denlinger. 2007. Up-regulation of heat shock proteins is essential for cold survival during insect diapause. Proc. Natl. Acad. Sci. U. S. A. 104:11130–11137.

Rivas, M. J., M. Saura, A. Pérez-Figueroa, M. Panova, T. Johansson, C. André, A. Caballero, E. Rolán-Alvarez, K. Johannesson, and H. Quesada. 2018. Population genomics of parallel evolution in gene expression and gene sequence during ecological adaptation. Sci. Rep. 8:1–12.

Riveron, J., T. Boto, and E. Alcorta. 2009. The effect of environmental temperature on olfactory perception in *Drosophila melanogaster*. J. Insect Physiol. 55:943–951.

Rúa, M. A., A. Antoninka, P. M. Antunes, V. B. Chaudhary, C. Gehring, L. J. Lamit, B. J. Piculell, J. D. Bever, C. Zabinski, J. F. Meadow, M. J. Lajeunesse, B. G. Milligan, J. Karst, and J. D. Hoeksema. 2016. Home-field advantage? Evidence of local adaptation among plants, soil, and arbuscular mycorrhizal fungi through meta-analysis. BMC Evol. Biol. 16:1–15.

Rubin, B. E. R., and C. S. Moreau. 2016. Comparative genomics reveals convergent rates of evolution in ant-plant mutualisms. Nat. Commun. 7. Nature Publishing Group.

Rubin, B. E. R., C. S. Moreau, S. Shigenobu, H. Watanabe, M. Hattori, Y. Sakaki, H. Ishikawa, E. F. Kirkness, L. Van Valen, S. Paterson, C. Pal, M. D. Maciá, A. Oliver, I. Schachar, A. Buckling, C. T. Bergstrom, M. Lachmann, M. Doebeli, N. Knowlton, D. H. Janzen *et al.* 2016. Comparative genomics reveals convergent rates of evolution in ant–plant mutualisms. Nat. Commun. 7:12679. Nature Publishing Group.

Rundle, H. D., and P. Nosil. 2005. Ecological speciation. Ecol. Lett. 8:336–352.

Running, S., Q. Mu, and M. Zhao. 2017. MOD16A2 MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V006. 2017, distributed by NASA EOSDIS Land Processes DAAC.

Savarit, F., and R. Fénéron. 2014. Imperfect chemical mimicry explains the imperfect social integration of the inquiline ant *Ectatomma parasiticum* (Hymenoptera: Formicidae: Ectatomminae). Myrmecological News 20:7–14.

Scheffers, B. R., L. De Meester, T. C. L. Bridge, A. A. Hoffmann, J. M. Pandolfi, R. T. Corlett, S. H. M. Butchart, P. Pearce-Kelly, K. M. Kovacs, D. Dudgeon, M. Pacifici, C. Rondinini, W. B. Foden, T. G. Martin, C. Mora, D. Bickford, and J. E. M. Watson. 2016. The broad footprint of climate change from genes to biomes to people. Science. 354.

Schell, T., B. Feldmeyer, H. Schmidt, B. Greshake, O. Tills, M. Truebano, S. D. Rundle, J. Paule, I. Ebersberger, and M. Pfenninger. 2017. An Annotated Draft Genome for *Radix auricularia* (Gastropoda, Mollusca). Genome Biol. Evol. 9:585–592.

Schlenke, T. A., and D. J. Begun. 2004. Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. Proc. Natl. Acad. Sci. 101:1626–1631.

Schluter, D. 2001. Ecology and the origin of species. Trends Ecol. Evol. 16:372–380.

Schluter, D. 2009. Evidence for ecological speciation and its alternative. Science. 323:737–741.

Schneider, D. 1964. Insects antennae. Annu. Rev. Entomol 9:103–122.

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

Schuler, H., K. Köppler, S. Daxböck-Horvath, B. Rasool, S. Krumböck, D. Schwarz, T. S. Hoffmeister, B. C. Schlick-Steiner, F. M. Steiner, A. Telschow, C. Stauffer, W. Arthofer, and M.

Riegler. 2016. The hitchhiker's guide to Europe: The infection dynamics of an ongoing *Wolbachia* invasion and mitochondrial selective sweep in *Rhagoletis cerasi*. Mol. Ecol. 25:1595–1609.

Schultz, T. R., S. A. Solomon, U. G. Mueller, P. Villesen, J. J. Boomsma, R. M. M. Adams, and B. Norden. 2002. Cryptic speciation in the fungus-growing ants *Cyphomyrmex longiscapus* Weber and *Cyphomyrmex muelleri* Schultz and Solomon, new species (Formicidae, Attini). Insectes Soc. 49:331–343.

Schwander, T., D. Arbuthnott, R. Gries, G. Gries, P. Nosil, and B. J. Crespi. 2013. Hydrocarbon divergence and reproductive isolation in *Timema* stick insects. BMC Evol. Biol. 13:151.

Scriven, J. J., P. R. Whitehorn, D. Goulson, and M. C. Tinsley. 2016. Niche partitioning in a sympatric cryptic species complex. Ecol. Evol. 6:1328–1339.

Seebacher, F. 2009. Responses to temperature variation: Integration of thermoregulation and metabolism in vertebrates. J. Exp. Biol. 212:2885–2891.

Seifert, B. 2008. Removal of allometric variance improves species separation in multi-character discriminant functions when species are strongly allometric and exposes diagnostic characters. Myrmecological News 11:91–105.

Servedio, M. R., G. S. Van Doorn, M. Kopp, A. M. Frame, and P. Nosil. 2011. Magic traits in speciation: "magic" but not rare? Trends Ecol. Evol. 26:389–397.

Sexton, J. P., P. J. McIntyre, A. L. Angert, and K. J. Rice. 2009. Evolution and Ecology of Species Range Limits. Annu. Rev. Ecol. Evol. Syst. 40:415–436.

Shafer, A. B. A., and J. B. W. Wolf. 2013. Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology. Ecol. Lett. 16:940–950.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V Kriventseva, and E. M. Zdobnov. 2015. Genome analysis BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.

Simola, D. F., L. Wissler, G. Donahue, R. M. Waterhouse, M. Helmkampf, J. Roux, S. Nygaard, K. M. Glastad, D. E. Hagen, L. Viljakainen, J. T. Reese, B. G. Hunt, D. Graur, E. Elhaik, E. V. Kriventseva, J. Wen, B. J. Parker, E. Cash, E. Privman, C. P. Childers *et al.* 2013. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. Genome Res. 23:1235–1247.

Slabbekoorn, H., and M. Peet. 2003. Birds sing at higher pitch in urban noise. Nature 424:267–268.

Slater, G. S. C., and E. Birney. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 11:1–11.

Slone, J. D., G. M. Pask, S. T. Ferguson, J. G. Millar, S. L. Berger, D. Reinberg, J. Liebig, A. Ray, and L. J. Zwiebel. 2017. Functional characterization of odorant receptors in the ponerine ant, *Harpegnathos saltator*. Proc. Natl. Acad. Sci. 201704647.

Smadja, C., and R. K. Butlin. 2009. On the scent of speciation: The chemosensory system and its role in premating isolation. Heredity (Edinb). 102:77–97.

Smit, A. F. A., R. Hubley, and P. Green. 2008. RepeatModeler Open-1.0. 2008–2015. Inst. Syst. Biol. Seattle, WA, USA.

Smith, C. D., A. Zimin, C. Holt, E. Abouheif, R. Benton, E. Cash, V. Croset, C. R. Currie, E. Elhaik, C. G. Elsik, M.-J. Favé, V. Fernandes, J. Gadau, J. D. Gibson, D. Graur, K. J. Grubbs, D. E. Hagen, M. Helmkampf, J.-A. Holley, H. Hu, *et al.* 2011a. Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). Proc. Natl. Acad. Sci. U. S. A. 108:5673–5678.

Smith, C. R., C. D. Smith, H. M. Robertson, M. Helmkampf, A. Zimin, M. Yandell, C. Holt, H. Hu, E. Abouheif, R. Benton, E. Cash, V. Croset, C. R. Currie, E. Elhaik, C. G. Elsik, M.-J. Favé, V. Fernandes, J. D. Gibson, D. Graur, W. Gronenberg, *et al.* Draft genome of the red harvester ant *Pogonomyrmex barbatus*. Proc. Natl. Acad. Sci. U. S. A. 108:5667–72.

Smith, K. L., L. J. Harmon, L. P. Shoo, and J. Melville. 2011c. Evidence of constrained phenotypic evolution in a cryptic species complex of agamid lizards. Evolution (N. Y). 65:976–992.

Sobel, J. M., G. F. Chen, L. R. Watt, and D. W. Schemske. 2009. The biology of speciation. Evolution (N. Y). 64:295–315.

Somero, G. N. 2010. The physiology of climate change: How potentials for acclimatization and genetic adaptation will determine "winners" and "losers." J. Exp. Biol. 213:912–920.

Sprenger, P. P., L. H. Burkert, B. Abou, W. Federle, and F. Menzel. 2018. Coping with the climate: Cuticular hydrocarbon acclimation of ants under constant and fluctuating conditions. J. Exp. Biol. 221.

Sprenger, P. P., J. Hartke, B. Feldmeyer, J. Orivel, T. Schmitt, and F. Menzel. 2019. Influence of Mutualistic Lifestyle, Mutualistic Partner, and Climate on Cuticular Hydrocarbon Profiles in Parabiotic Ants. J. Chem. Ecol. 45:741–754.

Sprenger, P. P., and F. Menzel. 2020. Cuticular hydrocarbons in ants (Hymenoptera: Formicidae) and other insects: how and why they differ among individuals, colonies, and species. Myrmecological News 30:1–26.

Stanke, M., O. Keller, I. Gunduz, A. Hayes, S. Waack, and B. Morgenstern. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34:W435–W439.

Steiner, F. M., S. Csősz, B. Markó, A. Gamisch, L. Rinnhofer, C. Folterbauer, S. Hammerle, C. Stauffer, W. Arthofer, and B. C. Schlick-Steiner. 2018. Turning one into five: Integrative taxonomy uncovers complex evolution of cryptic species in the harvester ant *Messor* "*structor*." Mol. Phylogenet. Evol. 127:387–404.

Stinchcombe, J. R., and H. E. Hoekstra. 2008. Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. Heredity (Edinb). 100:158–170.

Stinziano, J. R., R. J. Sové, H. D. Rundle, and B. J. Sinclair. 2015. Rapid desiccation hardening changes the cuticular hydrocarbon profile of *Drosophila melanogaster*. Comp. Biochem. Physiol. -Part A Mol. Integr. Physiol. 180:38–42. Elsevier Inc.

Stork, N. E. 2018. How Many Species of Insects and Other Terrestrial Arthropods Are There on Earth? Annu. Rev. Entomol. 63:31–45.

Störtkuhl, K. F., and A. Fiala. 2011. The smell of blue light: A new approach toward understanding an olfactory neuronal network. Front. Neurosci. 5:1–11.

Ströher, P. R., C. Li, and M. R. Pie. 2013. Exon-primed intron-crossing (EPIC) markers as a tool for ant phylogeography. Rev. Bras. Entomol. 57:427–430.

Struck, T. H., J. L. Feder, M. Bendiksby, S. Birkeland, J. Cerca, V. I. Gusarov, S. Kistenich, K. Larsson, L. H. Liow, M. D. Nowak, B. Stedje, L. Bachmann, and D. Dimitrov. 2018. Finding Evolutionary Processes Hidden in Cryptic Species. Trends Ecol. Evol. 33:153–163. Elsevier Ltd.

Suen, G., C. Teiling, L. Li, C. Holt, E. Abouheif, E. Bornberg-Bauer, P. Bouffard, E. J. Caldera, E. Cash, A. Cavanaugh, O. Denas, E. Elhaik, M.-J. Favé, J. Gadau, J. D. Gibson, D. Graur, K. J. Grubbs, D. E. Hagen, T. T. Harkins, M. Helmkampf, *et al.* 2011. The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle. PLoS Genet. 7.

Suh, A., L. Smeds, and H. Ellegren. 2015. The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. PLoS Biol. 13:1–18.

Supek, F., M. Bošnjak, N. Škunca, and T. Šmuc. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS One 6.

Svarovskaia, E. S., S. R. Cheslock, W. H. Zhang, W. S. Hu, and V. K. Pathak. 2003. Retroviral mutation rates and reverse transcriptase fidelity. Front. Biosci. 8:117–134.

Svensson, E. I. 2012. Non-ecological speciation, niche conservatism and thermal adaptation: How are they connected? Org. Divers. Evol. 12:229–240.

Swain, R. B. 1980. Trophic Competition among Parabiotic Ants. Insectes Soc. 27:377–390.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10:512–26.

Taylor, E. B., and J. D. McPhail. 2000. Historical contingency and ecological determinism interact to prime speciation in sticklebacks, Gasterosteus. Proc. R. Soc. B Biol. Sci. 267:2375–2384.

Temin, H. M., and S. Mizutani. 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature 226:1211–1213.

Tewksbury, J. J., R. B. Huey, and C. A. Deutsch. 2008. Ecology: Putting the heat on tropical animals. Science. 320:1296–1297.

Therkildsen, N. O., A. P. Wilder, D. O. Conover, S. B. Munch, H. Baumann, and S. R. Palumbi. 2019. Contrasting genomic shifts underlie parallel phenotypic evolution in Response To Fishing. Science. 490:487–490.

Thibert-Plante, X., and S. Gavrilets. 2013. Evolution of mate choice and the so called magic traits in ecological speciation. Ecol. Lett. 16:1004–1013.

Thomas, M. L., and L. W. Simmons. 2009. Male-derived cuticular hydrocarbons signal sperm competition intensity and affect ejaculate expenditure in crickets. Proc. R. Soc. B Biol. Sci. 276:383–388.

Thomas, M. L., and L. W. Simmons. 2008. Sexual dimorphism in cuticular hydrocarbons of the Australian field cricket *Teleogryllus oceanicus* (Orthoptera: Gryllidae). J. Insect Physiol. 54:1081–1089.

Thomas, M. L., and L. W. Simmons. 2011. Short-term phenotypic plasticity in long-chain cuticular hydrocarbons. Proc. R. Soc. B Biol. Sci. 278:3123–3128.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

Thompson, J. N., C. Schwind, P. R. Guimarães, and M. Friberg. 2013. Diversification through multitrait evolution in a coevolving interaction. Proc. Natl. Acad. Sci. U. S. A. 110:11487–11492.

Thompson, K. A., M. M. Osmond, and D. Schluter. 2019. Parallel genetic evolution and speciation from standing variation. Evol. Lett. 3:129–141.

Torres Dowdall, J., C. A. Handelsman, E. W. Ruell, S. K. Auer, D. N. Reznick, and C. K. Ghalambor. 2012. Fine-scale local adaptation in life histories along a continuous environmental gradient in Trinidadian guppies. Funct. Ecol. 26:616–627.

Tsutsui, N. D., A. V Suarez, J. C. Spagna, and J. S. Johnston. 2008. The evolution of genome size in ants. BMC Evol. Biol. 8:64.

Türke, M., B. Fiala, K. E. Linsenmair, and H. Feldhaar. 2010. Estimation of dispersal distances of the obligately plant-associated ant *Crematogaster decamera*. Ecol. Entomol. 35:662–671.

Turner, T. L., E. C. Bourne, E. J. Von Wettberg, T. T. Hu, and S. V. Nuzhdin. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. Nat. Genet. 42:260–263.

Urban, J. M., J. Bliss, C. E. Lawrence, and S. A. Gerbi. 2015. Sequencing ultra-long DNA molecules with the Oxford Nanopore MinION. BioRxiv Prepr., doi: 10.1002/dvdy.

Urban, M. C., J. J. Tewksbury, and K. S. Sheldon. 2012. On a collision course: Competition and dispersal differences create no-analogue communities and cause extinctions during climate change. Proc. R. Soc. B Biol. Sci. 279:2072–2080.

van't Hof, A. E., P. Campagne, D. J. Rigden, C. J. Yung, J. Lingley, M. A. Quail, N. Hall, A. C. Darby, and I. J. Saccheri. 2016. The industrial melanism mutation in British peppered moths is a transposable element. Nature 534:102–105. Nature Publishing Group.

Van Valen, L. 1973. A New Evolutionary Law. Evol. Theory 1:1–30.

van Wilgenburg, E., M. R. E. Symonds, and M. A. Elgar. 2011. Evolution of cuticular hydrocarbon diversity in ants. J. Evol. Biol. 24:1188–1198.

van Zweden, J. S., and P. d'Ettorre. 2010. Nestmate recognition in social insects and the role of hydrocarbons. Insect Hydrocarb. Biol. Biochem. Chem. Ecol. 222–243.

Vantaux, A., A. Dejean, A. Dor, and J. Orivel. 2007. Parasitism versus mutualism in the ant-garden parabiosis between *Camponotus femoratus* and *Crematogaster levior*. Insectes Soc. 54:95–99.

Vicente, R. E., W. Dáttilo, and T. J. Izzo. 2014. Differential recruitment of *Camponotus femoratus* (Fabricius) ants in response to ant garden herbivory. Neotrop. Entomol. 43:519–525.

Violle, C., D. R. Nemergut, Z. Pu, and L. Jiang. 2011. Phylogenetic limiting similarity and competitive exclusion. Ecol. Lett. 14:782–787.

Vodă, R., L. Dapporto, V. Dincă, and R. Vila. 2015a. Cryptic matters: Overlooked species generate most butterfly beta-diversity. Ecography (Cop.). 38:405–409.

Vodă, R., L. Dapporto, V. Dincă, and R. Vila. 2015b. Why do cryptic species tend not to co-occur? A case study on two cryptic pairs of butterflies. PLoS One 10:e0117802.

Wagner, D., M. Tissot, and D. Gordon. 2001. Task-related environment alters the cuticular hydrocarbon composition of harvester ants. J. Chem. Ecol. 27:1805–1819.

Waldvogel, A. M., B. Feldmeyer, M. Exposito-Alonso, T. Bataillon, A. Bergland, T. Flatt, F. Guillaume, R. Kofler, T. Mock, C. Rellstab, G. Rolshausen, O. Savolainen, K. Schmid, I. Schmitt, and M. Pfenninger. 2019. Evolutionary genomics can improve prediction of species' responses to climate change. Evol. Lett., doi: 10.1002/evl3.154.

Waldvogel, A. M., A. Wieser, T. Schell, S. Patel, H. Schmidt, T. Hankeln, B. Feldmeyer, and M. Pfenninger. 2018. The genomic footprint of climate adaptation in *Chironomus riparius*. Mol. Ecol. 27:1439–1456.

Walkinshaw, E., Y. Gai, C. Farkas, D. Richter, E. Nicholas, K. Keleman, and R. L. Davis. 2015. Identification of genes that promote or inhibit olfactory memory formation in *Drosophila*. Genetics 199:1173–1182.

Walther, G.-R., E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J.-M. Fromentin, O. Hoegh-Guldberg, and F. Bairlein. 2002. Ecological responses to recent climate change. Nature 416:389–395.

Wang, X., R. L. Edwards, A. S. Auler, H. Cheng, X. Kong, Y. Wang, F. W. Cruz, J. A. Dorale, and H. W. Chiang. 2017. Hydroclimate changes across the Amazon lowlands over the past 45,000 years. Nature 541:204–207. Nature Publishing Group.

Weber, R. E. 2007. High-altitude adaptations in vertebrate hemoglobins. Respir. Physiol. Neurobiol. 158:132–142.

West-Eberhard, M. J. 1984. Sexual Selection, Social Competition, and Speciation. Q. Rev. Biol. 58:156–183.

Wickham, H. 2016. *ggplot2: Elegant graphics for data analysis* , 2nd ed. Springer-Verlag, New York, NY.

Wilson, J. S., S. L. Clark, K. A. Williams, and J. P. Pitts. 2012. Historical biogeography of the arid-adapted velvet ant *Sphaeropthalma arota* (Hymenoptera: Mutillidae) reveals cryptic species. J. Biogeogr. 39:336–352.

Wingett, S. W., S. Andrews, and R. S. Hamilton. 2018. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Research 1–12.

Winkelmann, K., M. J. Genner, T. Takahashi, and L. Rüber. 2014. Competition-driven speciation in cichlid fish. Nat. Commun. 5: 3412.

Wittkopp, P. J., B. L. Williams, J. E. Selegue, and S. B. Carroll. 2003. *Drosophila* pigmentation evolution: Divergent genotypes underlying convergent phenotypes. Proc. Natl. Acad. Sci. U. S. A. 100:1808–1813.

Wolak, M. E., D. J. Fairbairn, and Y. R. Paulsen. 2012. Guidelines for estimating repeatability. Methods Ecol. Evol. 3:129–137.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. Proc. Natl. Acad. Sci. U. S. A. 106:13875–13879.

Wu, H., X. Guang, M. B. Al-Fageeh, J. Cao, S. Pan, H. Zhou, L. Zhang, M. H. Abutarboush, Y. Xing, Z. Xie, A. S. Alshanqeeti, Y. Zhang, Q. Yao, B. M. Al-Shomrani, D. Zhang, J. Li, M. M. Manee, Z. Yang, L. Yang, Y. Liu, *et al*. 2014. Camelid genomes reveal evolution and adaptation to desert environments. Nat. Commun. 5.

Wurm, Y., J. Wang, O. Riba-Grognuz, M. Corona, S. Nygaard, B. G. Hunt, K. K. Ingram, L. Falquet, M. Nipitwattanaphon, D. Gotzek, M. B. Dijkstra, J. Oettler, F. Comtesse, C. J. Shih, W. J. Wu, C. C. Yang, J. Thomas, E. Beaudoing, S. Pradervand, V. Flegel, *et al*. 2011. The genome of the fire ant *Solenopsis invicta*. Proc. Natl. Acad. Sci. 108:5679–5684.

Youngsteadt, E., J. A. Baca, J. Osborne, and C. Schal. 2009. Species-Specific Seed Dispersal in an Obligate Ant-Plant Mutualism. PLoS One 4:e4335.

Youngsteadt, E., P. G. Bustios, and C. Schal. 2010. Divergent chemical cues elicit seed collecting by ants in an obligate multi-species mutualism in lowland amazonia. PLoS One 5:e15822.

Youngsteadt, E., S. Nojima, C. Häberlein, S. Schulz, and C. Schal. 2008. Seed odor mediates an obligate ant-plant mutualism in Amazonian rainforests. Proc. Natl. Acad. Sci. 105:4571–4575.

Yu, D. W. 1994. The Structural Role of Epiphytes in Ant Gardens. Biotropica 26:222.

Zhou, X., A. Rokas, S. L. Berger, J. Liebig, A. Ray, and L. J. Zwiebel. 2015. Chemoreceptor evolution in Hymenoptera and its implications for the evolution of eusociality. Genome Biol. Evol. 7:2407–2416.

ACKNOWLEDGEMENTS

Information removed for privacy purposes

Curriculum vitae

Information removed for privacy purposes