Genetic analysis of wing pattern and pheromone composition in two sister species of *Heliconius* butterflies

Bruna Cama

MSc by Research

University of York

Biology

November 2016

Abstract.

The problem of sympatric speciation and speciation with gene flow has been of great interest to evolutionary biologists for years. While both concepts were highly debated in the past, it is now known that both events are possible. Ecological factors can come into play to encourage species divergence in sympatry, with different selective pressures giving rise to different adaptive phenotypes within the same territory, thereby forming separate races within a population that can eventually evolve into fully separate species. When this happens, gene flow may be partially retained between the incipient species as they diverge. But gene flow and divergent selection do not affect all parts of the genome equally: novel adaptations that promote speciation tend to accumulate in specific regions known as islands of divergence, that can act as reproductive barriers between the diverging taxa. Understanding the genetics of traits that are leading speciation is thus of great importance in the study of evolution, and *Heliconius* butterflies offer a great model system for researching this phenomenon. In this study, hybrid individuals are used to characterize the genetic structure of two traits that are likely to have this role, the colour pattern and the pheromone blend, in two sister species of *Heliconius* butterflies, *H. pardalinus butleri* and *H. elevatus*. Several putative colour pattern loci were found, potentially arranged into 2-3 linkage groups, consistent with the arrangement observed in other *Heliconius* species. The biosynthetic machinery underlying pheromone production in *Heliconius* was also partially characterized for the first time by investigating correlations between compounds within the F2 hybrids, which also offered insight into the genetic organization of pheromone genes.

Table of Contents

Abstract. ... 2

Table of Contents ... 3

List of Figures ... 5

List of Tables ... 7

Acknowledgements ... 8

Author's declaration ... 9

General Introduction ... 10

1. Speciation and speciation genes ... 10

1.1 Sympatry and speciation with gene flow ... 12

1.2 Magic traits ... 15

1.3 Hybrid speciation ... 16

1.4 *Heliconius* butterflies as model organisms ... 18

Phenotypic analysis of the *H. pardalinus* and *H. elevatus* colour pattern. ... 23

1. Introduction ... 23

1.1 The colour pattern as a reproductive barrier ... 23

1.2 What makes a colour pattern? The Nymphalid groundplan and its variations. ... 24

1.3 The *Heliconius* pattern: mimicry and assortative mating ... 25

1.4 The *Heliconius* mimicry toolbox. ... 27

1.5 The silvaniform colour pattern and its genetic architecture ... 29

2. Materials and methods ... 31

3. Results ... 34

3.1 Colour pattern analysis. ... 34

3.2 Inferred linkage between traits. ... 43

3.3 Quantitative variation. ... 44

4. Discussion ... 45

4.1 Multilocus architecture of the colour pattern ... 45

4.2 Quantitative variation in the black spots. ... 47

4.3 Different developmental interpretations of the loci can affect phenotype scoring. ... 49

4.4 *H. elevatus* and *H. pardalinus*'s colour patterns share their genetic architecture with other silvaniforms. 50

4.5 The white invasion dilemma: a re-activated *H. elevatus* allele? 51

5. Conclusion. 52

Characterizing the pheromone blend of *H. pardalinus* and *H. elevatus.* 54

1. Introduction 54

1.1 Insect pheromones: important means of communication 54

1.2 Lepidopteran pheromone biosynthesis: a general look 55

1.3 Pheromones in butterflies 59

1.4 *Heliconius* pheromones: a relatively unexplored topic 61

2. Materials and methods 62

3. Results 66

3.1 Preliminary analysis of the pure species 66

3.2 F2, F1 and BC analysis 72

3.3 In-depth analysis of F2 hybrids 73

4. Discussion 84

5. Conclusions. 95

Concluding remarks 1

References. 3

List of Figures

Figure 1. Phylogeny of the silvaniform and melpomene clades of Heliconius (adapted from Heliconius Genome Consortium, 2012). 20
Figure 2. The overlapping ranges of H. pardalinus and H. elevatus (Rosser, 2012). Triangles and squares indicate collection sites. 22
Figure 3A. The general Nymphalid groundplan (picture from http://sites.biology.duke.edu/nijhout/patterns2.html- Nijhout, n.d.). B. The derived Heliconius groundplan showing the variable black elements (picture from Nijhout et al., 1990). 24
Figure 4. The H. elevatus and H. pardalinus butleri colour patterns. Photos by Lucie Queste. 26
Figure 5. The main Heliconius colour pattern genes that have been researched in depth thus far. An additional gene of this kind is presumably found on chromosome 1, and may contain the K locus from H. numata. Picture adapted from Jiggins et al. (2017). 28
Figure 6. Phenotypes of all scored traits. The pure species phenotypes are shown as well as the heterozygous phenotypes if any. 33
Figure 7A. Pure species phenotypes and putative genotypes, with elements of the phenotype highlighted. B. Comparison of phenotypes seen in the pure species, F1, F2 (though not all F2 phenotypes are shown) and backcrosses. Putative genotypes are also reported. 35
Figure 8. Different phenotypes of Yb and white invasion on different Rb backgrounds. 39
Figure 9. Various degrees of black expansion in F2 individuals. Notice how it can affect the discal area and the apical area independently. It is also shown that extensive amounts of black expansion on the forewing do not necessarily affect the hindwing as well. 44
Figure 10. A variety of phenotypes found in the backcrosses to H. pardalinus. Notice variations in the shape of the hindwing rays. 47
Figure 11. Comparison between white invasion in an F2 hybrid (left) and H. elevatus bari (right). Picture of H. elevatus bari from http://www.heliconius.net/. 52
Figure 12. Fatty acid metabolism (Mann et al., unpublished). C(n) indicates the number of carbons in the main chain, Z(n) indicates the position of a double bond, if present (the number of double bonds is given after the number of carbons: C(n):0 for no double bonds, C(n):1 for one double bond). –COOH denotes fatty acids. Notice the desaturation event occurs early on, as C18:0-COOH (stearic acid) is transformed into Z9-C18:1-COOH (oleic acid). Other relevant compound classes represented here include alcohols (-OH), acetates (-Ac) and aldehydes (-CO). C(n) on its own represents alkanes, while Z(m)-C(n):1 represents alkenes. 57
Figure 13. Principal component analysis of the pure species. Red = H. elevatus, cyan = H. pardalinus. 67
Figure 14. Correlations between compounds in the pure species. Colours represent Pearson's correlation coefficient. Only significant correlations ($p<0.05$) are shown. Different families of compounds are indicated by the small shapes. 71
Figure 15. PCA of the logged concentrations of the pheromone blend in F2 individuals with different colours showing the individuals from the feeding experiments. Red = fed on P. edulis, black = fed on P. serratodigitata, cyan = F2 that were not part of the feeding experiment. 74

Figure 16. PCA of logged concentrations of the pheromone blend in the 10 individuals from the feeding experiment. Red = fed on P. serratodigitata, cyan = P. edulis. 75
Figure 17. PCA on the logged pheromone blends of the full dataset. Black = F2 (Peru), orange = F2 (York), cyan = H. pardalinus, red = H. elevatus, purple = backcrosses to H. pardalinus, green = F1. 76
Figure 18. F2 pheromone blend components. Concentration histograms for the 11-methylpentacosane normalized dataset. Each graph represents a compound. None show bimodal or trimodal distributions expected of control by a single Mendelian locus. On each plot, the x-axis represents the compound's concentration, the y-axis is a simple count of individuals. Each bar reports the number of individuals whose concentration falls within a specific range. 78
Figure 19. Correlations between compounds in the F2 standardized by C27-29. Colours represent Pearson's correlation coefficient. Only significant values ($p<0.05$) are shown. 79
Figure 20. Correlations between compounds in the F2 standardized by C25Me. Colours represent Pearson's correlation coefficient. Only significant values ($p<0.05$) are shown. 80
Figure 21. Chemical structures of the compounds belonging to the main clusters A-G..... 82
Figure 22. Histograms for the averaged concentrations of compounds belonging to each cluster in the C25Me dataset. Cluster names are indicated at the top of each graph. Cluster A= C21-23 alkenes; B= C20-23 alkanes; C= large even numbered alkanes; D= vanillin derivatives; E= phytol derivatives; F= complex esters; G= minor alkenes; J= all alkanes. For a complete list of compounds in each cluster, see Table 5..... 83
Figure 23. Histograms for the averaged concentrations of compounds belonging to each cluster in the C27-29 dataset. Cluster names are indicated at the top of each graph. Cluster A= C21-23 alkenes; B= C20-23 alkanes; C= large even numbered alkanes; D= vanillin derivatives; E= phytol derivatives; F= complex esters; G= minor alkenes; H= alkenes and phytol derivatives; I= large odd-numbered alkanes. For a complete list of compounds in each cluster, see Table 5. 84
Figure 24. The FA metabolic pathway showing the most prominent H. pardalinus and H. elevatus specific compounds. Even-numbered alkanes and alkenes, as well as their derivatives, minor compounds, and compounds of unclear structure are not included in this figure. 86
Figure 25. Correlation matrix for logged compound concentrations in the backcross. Colours represent Pearson's correlation coefficient. Only significant correlations ($p<0.05$) are shown. 89
Figure 26. Cluster A. Note that in ?-docosene, neither the double bond's position nor its orientation (E or Z) are known..... 90
Figure 27. The two most robust alkane clusters, B (C20-23 in length) and C (C24-28 in length). 91
Figure 28. Cluster D. Notice the similarity in structure between homovanillyl-alcohol and syringaldehyde..... 92
Figure 29. Cluster E. The similarity in structure between phytol and HHFA is evident. The acetate groups potentially derived from phytol degradation are circled. 93
Figure 30. Cluster F. The main chain's length is the same in both esters, but the functional group's length changes. The black circles indicate the propionate and butyrate group, the blue ones the acetate portion of the functional group..... 94
Figure 31. Cluster G. The positions of the double bonds remain unknown, the ones represented here are just some of the possibilities. 95

List of Tables

Table 1. Chi-squared tests for Mendelian frequencies of different putative loci, assuming complete or incomplete dominance. In the backcrosses, the expected ratios refer to heterozygous:pardalinus phenotype, as pure elevatus phenotypes are expected to be completely absent. 42
Table 2. Chi-square tests for linkage based on phenotype frequencies. All combinations with the apical band (A) are calculated only with the York frequencies. In=white invasion. If $p<0.05$, traits may be linked. 44
Table 3. Median, 25% (Q1) and 75% (Q3) quartiles for the concentrations (nmol) of compounds from H. elevatus and H. pardalinus. Also shown are the results of the Kruskal-Wallis test between the two species. 69
Table 4. List of all compounds included in the whole dataset analysis. Median, 25% quartile (Q1) and 75% quartile (Q3) of the compounds' concentrations (nmol) are reported for comparisons between the different broods. 73
Table 5. The main clusters of compounds detected by R(corrplot). 81

Acknowledgements

I would like first and foremost to acknowledge Dr. Kanchon Dasmahapatra, who gave me the opportunity to carry out my MSc by Research. He was of great help and an amazing supervisor. I've learned a lot during MSc by Research and my final year project and it is thanks to him.

Next I would like to thank Dr. Stefan Schulz for having me in his lab at Braunschweig TU. His guidance was of paramount importance for this project. Likewise, I would like to thank Florian Mann for being supportive and for teaching me all the secrets of GCMS. He was tremendous help and meeting him really impacted me. Without him this project would not have been possible, or as enjoyable.

I would also like to acknowledge the other researchers in the Dasmahapatra lab, mainly Lucie Queste, Jake Morris and Neil Rosser, for being not only good research companions, but also good people in general, and for having contributed to making my research experience in York very pleasant.

Author's declaration

I, Bruna Cama, declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. I furthermore declare that all the work presented here was carried out by me unless otherwise stated, that it is original and that it has never been published before. Florian Mann identified most of the compounds reported here, and carried out some general preliminary analysis. Jake Morris first collected the pheromones of *H. elevatus* and *H. pardalinus* and carried out the analysis of the pure species' pheromone blend, which I reproduced here. I began investigating the colour pattern in 2014-2015 as part of my undergraduate final year project: this included phenotypical analysis of F1 and F2 hybrids generated in York only. This created some preliminary knowledge to base my current work on. However I repeated the analysis completely for my MSc by Research, which brought substantial changes to my original explanations for the phenotypes.

Chapter 1

General Introduction

1. Speciation and speciation genes

The study of how organisms respond to environmental changes is of crucial interest in Evolutionary Biology. Every natural community is constantly faced with challenges in the form of natural selection, and each organism must rise to the occasion in order to survive and pass down its genes. Natural communities consist in more than one kind of organism, so the diversity of a community depends in large part on the number of species it comprises, a parameter known as species richness. This makes species the fundamental units of biodiversity. In light of this, speciation (alongside extinction and adaptation) plays an important role in modifying its patterns. Speciation is the process by which new species arise from pre-existing populations, and it has become one of the most discussed topics in evolution. The formation of new species through a process of divergence is a problematic topic like many in Evolutionary Biology, due to the impossibility of observing it in real time. Yet, the origin of new traits and their success (or lack thereof) ultimately depends on genes: in order for evolution to take place, the traits must be heritable and an individual must be able to pass them down to its offspring. Thus, the study of speciation has benefitted greatly from the development of genetic techniques that allow us to study its effects on the genome more so than on a macroscopic scale (Wu & Ting, 2004). To better understand changes in biodiversity patterns, it is crucial to investigate the genetic mechanisms that lead to the origin of new species.

Noor & Feder (2006) highlighted the usefulness of genomics as a tool for research on natural populations, which offer an insight into the process of speciation and reproductive isolation that is not influenced by artificial factors. In addition to this, with genomic approaches one can explore the evolutionary history of new taxa that do not have a history as laboratory model organisms (Noor & Feder, 2006). Aside from this, there are many ways genetics can help answering important questions about speciation (Noor & Feder, 2006; Nosil & Schluter, 2011). These include how many genes are generally involved in speciation and to which extent do they affect it, or what regions of the genome are most likely to play a part in the process, or how does genetic

incompatibility in hybrids influence the formation of novel species. According to Nosil and Schluter (2011), these questions are relevant to the interests of geneticists, ecologists and evolutionary biologists alike, as their solutions can advance our understanding of speciation and the origins of species diversity, and they offer several examples of how they can be of use. For example, the classes of genes involved most closely with speciation and their effect on the phenotype can paint a picture of what environmental pressures and genomic conflicts (which arise when different genes in the same genome are subject to different rules of inheritance) are leading to adaptation. In addition to that, they can be informative as to what kind of selection is acting on those genes- whether it's divergent, stabilizing or convergent (Nosil & Schluter, 2011). In the same paper, Nosil and Schluter refer to any gene of interest to the evolution of a species which has had a substantial effect on the development of reproductive isolation as a "speciation gene".

There are many possible definitions of "speciation gene", but what they all have in common is an emphasis on the reduction of gene flow at that locus between the two developing species, with reproductive isolation (i.e. the complete lack of genetic exchanges between the two species' gene pools) as the final outcome. Orr and Presgraves (2000) for example defined them as genes that have deleterious effects when they are transferred to a different species, thus reducing hybrid fitness. Mihola et al. (2009) use a similar definition as they describe what they call a mammalian speciation gene, a methyltransferase involved in chromatin remodeling and responsible for hybrid sterility in mice. Both of these definitions imply that any form of gene flow between the incipient species would be immediately impeded in the presence of such genes, since the hybrids have no means of mating back to the parental populations, and thus no part of the genome could be transferred from one population to the other. Wu and Ting (2004) choose a more generalistic approach: while speciation genes ultimately promote reproductive isolation, their effect does not always need to be immediately as drastic as hybrid sterility or inviability, and gene flow can occasionally keep occurring between the populations even as they diverge into different species.

In fact, it is tempting to consider speciation as a process that affects the whole genome at once. Mayr, a major figure in evolutionary biology, strongly argued in favour of an interpretation of genomes as complexes of genes whose effects are tightly interconnected and co-dependent. He implied that a genome would act as a large unit in the face of natural selection and genetic drift rather than a set of independent elements, and as a consequence, changes experienced by one locus would affect all the others as well (Mayr, 1963). However, increasing evidence points to a

genic view of speciation, which considers each gene as a unit with its own effect on species divergence and its own fitness value. In this context, speciation genes can have a range of effects that range from weak ecological, physiological and/or behavioural discrepancies between the diverging populations to complete hybrid inviability. Additionally, while speciation genes exert their influence by inhibiting gene flow, the rest of the genome that is not involved in speciation may continue to be exchanged between the emerging species (Wu and Ting, 2004). This results in the formation of so-called "islands of divergence", genomic regions that are highly differentiated in comparison to the rest of the genome, which in the context of the "island" simile represents the "sea" of gene flow.

The debate between supporters of the whole-genome speciation concept and the genic one is closely tied to the problem of speciation with gene flow. The formers tend to argue that in order for species to form, strong barriers (of usually geographical nature) must exist between the diverging populations, while the latters are supportive of a scenario where new species can arise in the face of gene flow and a lack of reproductive isolation.

1.1 Sympatry and speciation with gene flow

Speciation is most commonly exemplified as a process of vicariance, by which two subpopulations of the same species diverge due to geographic barriers to reproduction, until they become unable to produce viable offspring, even in the event that they were to reunite- at which point they are considered fully separate species (Coyne & Orr, 2004). This situation is actually known as allopatric speciation and it is identified by a complete lack of gene flow between the two diverging subpopulations (Coyne & Orr, 2004). This is a rather extreme occurrence, as it is also possible for two lineages to diverge without any strict barriers to reproduction. The extreme opposite to allopatric speciation, sympatric speciation, is what happens when two lineages diverge from the same population while consistently occupying the same range, with no apparent substructure (Coyne & Orr, 2004; Kawecki, 2004). Fitzpatrick et al. (2008) add to these prerequisites the necessity for panmixia (random mating, so that every individual has equal chances to pass down its genes) in the original population. While it may be tempting to classify the different modes of speciation into discrete categories, allopatric and sympatric speciation may simply represent the two ends of a continuum. With that being clarified, most instances of speciation likely occur with some degree of gene flow between the diverging taxa.

The problem of sympatry and speciation-with-gene-flow has been of great interest to evolutionary biologists for years. This is because, due to the way natural selection works, which implies selective mating between the most adaptive individuals, it can be difficult to imagine how traits can diverge with no hurdles to reproduction and no environmental differences to drive the evolution of different adaptations. Effectively, gene flow tends to homogenize subpopulations, as recombination between two diverging taxa tends to break up the adaptive genetic combinations that are being developed by the two intermixing populations. Disruptive selection on the other hand tends to separate them, so speciation in sympatry may seem like a paradox (Coyne & Orr, 2004).

However, sympatric speciation between two intermixing populations is possible if the two are subjected to different ecological optima, and thus, to divergent selection (Bolnick & Fitzpatrick, 2007). This phenomenon is known as ecological speciation, and it is well distinct from models of speciation where ecological factors do not play a central role, such as those that revolve around chance events (like population bottlenecks or genetic drift) or sexual selection (Rundle & Nosil, 2005). It needs to be said that while ecological speciation is being presented here in the context of sympatry, it is not inherently a feature of sympatric speciation, and can in fact have a large role regardless of whether the two diverging populations exist in sympatry, allopatry or in any intermediate state (Rundle & Nosil, 2005). When ecological speciation occurs in sympatry, divergent selection shapes the populations' phenotypes so they form a barrier to homogenization in the same way as a geographical feature of the territory in the context of allopatric speciation. However not all traits are equally subjected to selection, and not all regions of the genome have equal chances of being exchanged during mating, which means that the genomes of closely related, recently diverged species often show specific islands of divergence rather than widespread differentiation (Harrison & Larson, 2014). In such a situation, mating between the species and the subsequent exchange of alleles can still occur, but the areas of divergence will remain stably untouched, in which case the islands themselves represent the species boundary.

It is worth mentioning that in order for islands of divergence to form, it is necessary that genetic variation already be present in the population. However it is important to distinguish between cases where such variation arose in sympatry and cases where it arose in allopatry, with the two populations of interest becoming sympatric only following secondary contact (Harrison & Larson, 2014). For example, in the case of *Ficedula* flycatchers, a popular example of non homogeneous

genome divergence, the formation of islands between the sympatric species *F. albicollis* and *F. hypoleuca* was probably favoured by the presence of polymorphisms resulting from periods of allopatry due to glaciations. Later, the two populations came into secondary contact and re-established gene flow but these polymorphic loci were not subject to homogenization in the same measure as the rest of the genome, and thus, even with the species having been reunited, they continued diverging in sympatry (Ellegren et al., 2012).

It is still unclear what may make a genomic region more prone to becoming an island of divergence in the face of gene flow, but some broad patterns have been identified. The most obvious trend is that regions that are naturally subject to less recombination due to their location are also the ones more likely to become highly divergent. These include centromeres and sex chromosomes (Harrison & Larson, 2014), as well as inversions and rearrangements (Navarro & Barton, 2003).

In recently diverged species with very similar genomes, a chromosome rearrangement becomes an area of differentiation. This is because even in the event of hybridization, the rearranged region will not be involved in recombination between heterozygous homologous chromosomes due to its structure being discrepant between the two homologs (but the exchange will occur naturally if the two chromosomes are homozygous and thus sport the same inversion). As a consequence, alleles that are contributing to species divergence may accumulate there and keep diverging in the face of gene flow across the rest of the genome as they would remain unaffected by it (Nachman & Payseur, 2012).

Selection dictates what alleles accumulate in the low-recombination zones. While in high-recombination zones exchanges between the two sub-populations may maintain normal levels of nucleotide variation, the islands of divergence remain untouched in the two sub-populations, and thus tend to lose diversity due to stabilizing selection. This loss of diversity causes the effects of selection to be magnified in the low-recombination zones, as deleterious alleles are more quickly eliminated and beneficial ones are easily preserved (Nachman & Payseur, 2012). This explains why advantageous alleles that contribute to sub-population divergence under differing selective pressures will tend to accumulate there.

This trend is only visible in relatively recently diverged species: species that diverged with gene flow in the remote past tend to have more uniformly differentiated genomes not unlike those of

species that diverged in allopatry, due to a phenomenon called “genetic hitchhiking” (Feder, Gejji et al., 2012; Feder et al., 2012; Via, 2012). Initially, only the loci that are under direct divergent selection become differentiated, but this effect tends to extend to the nearby loci due to physical linkage: this is known as “divergence hitchhiking” and it leads to enlargement of the islands of divergence. As more and more loci begin diverging due to this, the effect gradually extends to the genome as a whole, at which point it is known as “genome hitchhiking”. This final step in the process of speciation effectively means that loci cannot introgress across any part of the genomes. At this point the effect of genomic barriers to gene flow has acquired the same magnitude as the effect of geographical barriers in allopatrically diverged species (Feder et al., 2012). This implies that the absence of islands of divergence does not necessarily indicate that species diverged with a complete lack of gene flow.

1.2 Magic traits

If traits under divergent selection found in islands of divergence also happen to have an effect on the phenotype that pleiotropically influences mate choice and favours reproductive isolation, they are known as “magic traits”. In their presence the separation between species in the face of gene flow is facilitated. Effectively, a single gene is controlling both ecological fitness and non-random mating (Bolnick & Fitzpatrick, 2007; Servedio, 2011).

A classic example of sympatric ecological speciation where ecological divergence affects mate choice as well is observed in *Rhagoletis pomonella,* the apple maggot fly. This species is broadly distributed in North America all the way to Mexico and it commonly infests hawthorn, yet certain populations in the Northern United States successfully colonized apples and formed a separate race, which is to this day not considered a species on its own, but it could be taken as an instance of an incipient species (Bush, 1969). Mating in *Rhagoletis* is strongly associated with the host plant: courtship behaviour is synchronized with fruiting periods as males typically establish a territory on a fruit of the same plant they emerged from, where they wait for a female to mate with. The females in turn lay eggs within that same fruit. This implies that subpopulations of the original *R. pomonella* race that chose to use apple rather than hawthorn as their host easily became reproductively isolated from the hawthorn-feeding species, without finding themselves in a situation of allopatry (Bush, 1969). This change in ecology has a genetic basis: the apple-feeding behaviour is associated with a chromosomal inversion that originated in the Mexican populations of *R. pomonella* and later got introduced in the Northern populations via introgression (Feder,

2003). This event essentially reconciles allopatric and sympatric speciation as the inversion formed in allopatry, yet when it was introduced in a new population it favoured ecological specialization thus allowing the apple and hawthorn races to diverge without geographical barriers. Hybridization as a source of novel alleles can play an important role in speciation (Harrison & Larson, 2014). Under strong divergent selection, one may expect the two races to eventually become fully distinct species.

1.3 Hybrid speciation

Speciation by hybridization is a type of speciation that necessarily occurs in sympatry, since in order to hybridize, two species have to coexist within the same habitat. It is a phenomenon wherein a new species originates from the intercrossing of two other species, from which it becomes distinct (Mavares, 2006). In order for this to happen, F1 individuals must be viable and at least one sex must be fertile and able to backcross to a parental species, introducing new alleles in it. If these alleles are adaptive, a new lineage can form and give rise to a new species (Jiggins et al., 2008). However backcrossing is not necessary for the formation of a new species if the F1 individuals can interbreed and form a stable lineage (so, if both sexes are fertile) (Jiggins et al., 2008). Moreover, the hybrid phenotype itself must cause reproductive isolation of the hybrid lineage from its parental taxa (Mavares, 2006). There are two kinds of hybrid speciation: polyploid and homoploid. The former occurs when the number of chromosomes of the hybrid doubles, as a result of selfing (Mallet, 2007). The latter on the other hand involves no change in chromosome number. Some of the best-known examples of hybrid speciation are observed in plants. At least 25% of modern plant species are known to hybridize naturally, (Mallet, 2005), and in these organisms polyploid speciation is particularly common. For example, the sunflower *Helianthus anomalus* is a hybrid species derived from *H. annuus* and *H. petiolaris* (Rieseberg, 1995). The plasticity of a plant's genome and physiology means that two different species are able to hybridize without the offspring suffering major consequences from genome incompatibilities and polyploidy. In animals, hybrid speciation is thought to be much rarer and only possible in case of homoploidy (where the chromosome number is the same between the two parental species) (Mallet, 2007).

It is a common misconception that in the event that hybridization between two animal species does occur, the F1 crosses are usually non adaptive and infertile, and that in the possibility that backcrosses may occur, the resulting F2 would be even less fit for survival. This essentially paints

hybridization as a rare, unfavourable phenomenon overall. However, this is not always the case: within distinct groups of animals, crossing occurs at varying, but often relatively high rates, and as seen in *Rhagoletis* flies, it is not always disadvantageous. Another common example of adaptive introgression is seen between two species of *Anopheles* mosquitos, *A. coluzzii* and *A. gambiae,* both of which are malaria vectors (Norris et al., 2015). *A. gambiae* carries an island of divergence containing a single nucleotide polymorphism (SNP) that confers resistance to several insecticides. Upon introduction of new insecticide-laced bed nets in Mali, Malian populations of both species were subjected to new selective pressures, that favoured hybridization as the F1 hybrids and the backcrosses to *A. coluzzii* that carried the SNP became advantaged against the new control measures (Norris et al., 2015). Thus, the SNP successfully introgressed from *A. gambiae* to *A. coluzzii* while previous to the introduction of the insecticide nets the hybrids were disadvantaged in comparison to the two pure species.

Another classic example of adaptive introgression comes from warfarin resistance in mice (Song et al., 2011; Hedrick, 2013): warfarin is a rodenticide that was deployed in the 1950s against house mice (*Mus musculus domesticus*), which developed resistance to its anticoagulant effect in around a decade from its introduction. Polymorphisms in the gene *vkorc1,* involved in vitamin K metabolism, can confer resistance to warfarin, and Song et al. (2011) demonstrated that some of the resistant variants of this gene in house mice were actually derived from *M. spretus,* the Algerian mouse. In *M. spretus,* the polymorphism probably represented an adaptation to dietary requirements that was selected independently of warfarin usage, since the pesticide was not used extensively on this species, and this is plausible given that *M. spretus'* diet is poor in vitamin K and this would require particularly efficient metabolism of this blood coagulation factor. (Hedrick, 2013; Song et al., 2011). Following positive selection at the resistant *vkorc1* allele *vkorc1*^{spr} *Vkorc1*^{spr} crossed species barriers following hybridization between *M. spretus* and *M. musculus domesticus,* where it was readily subjected to novel positive selective pressures dictated by warfarin usage (Song et al., 2011).

With the notion that hybridization in the animal world may often have neutral or even positive effects rather than debilitating ones, the fact that it is fairly common seems less surprising. The amount of species that partake in hybridization varies with the taxon: certain taxa include a very high percentage of hybridizing species, while in others it is less common. In *Heliconiinae* butterflies, hybridization is quite common as it is thought to involve 26% of all species, with

Heliconius being a very active genus in these regards (Mallet, 2005). By hybridizing, *Heliconius* butterflies have managed to exchange colour pattern genes in the past, with distantly related species acquiring the same patterns, as is the case for *H. elevatus* and *H. melpomene* (which will be covered in the next paragraphs) (Heliconius Genome Consortium, 2012). This is a good example of adaptive introgression since these butterflies depend on their aposematic colouration to avoid predation, and the more common a warning pattern is, the stronger the signal sent to potential predators.

1.4 *Heliconius* butterflies as model organisms

Heliconius butterflies are one the most important animal models in the study of evolution, and in particular, speciation (Merrill, 2015). This genus, which comprises over 40 species resulting from a relatively recent evolutionary radiation, populates Central and South America (with one species being found in the Southern United States as well), reaching maximum levels of diversity in the Amazon basin (Rosser et al., 2012). They are best known for their aposematic colouration: their striking, highly contrasting colours offer them protection against predators by signalling that the butterfly is unpalatable. Originally *Heliconius* butterflies were studied as models for mimicry, in particular Müllerian mimicry (Bates, 1862), which had an important effect on *Heliconius* evolution. Müllerian mimicry is a phenomenon by which different species of poisonous or unpalatable animals evolve convergently to adopt the same aposematic colouration. Bates first described the *Heliconius* mimicry system, where different species within a certain area of the neotropics would often adopt the same model warning pattern, but different territories would have different models (Bates, 1862). Additionally each territory may have more than one model, each model being known as a "mimicry ring", so within the same territory closely related species may evolve strikingly divergent colour patterns (Bates, 1862). Thus, there are cases where distantly related species can sport similar colour patterns, and cases where sister species or different races of the same species evolve divergently to become part of different mimicry rings. For example *H. cydno*, which is closely related to *H. melpomene,* mimics members of the distantly related *H. sapho* clade, whose species are often black and white and rather different in biology from species in the *melpomene/cydno* clade (Kronforst & Papa, 2015). Mimicry was the starting point for the study of *Heliconius* biology, and today it is known that the *Heliconius* history is rich in phenomena that are not easily observed in other models of the animal kingdom.

Heliconius butterflies in fact provide some of the most well-known examples of speciation with gene flow and hybrid speciation, both of which have left still-observable traces on their genome (Merrill et al., 2015, Supple et al., 2015, Mavarez et al., 2006). As the divergence between *Heliconius* species is relatively recent (Kozak et al., 2015) the genomes of many sister and non-sister species remain compatible to this day and the F1 hybrids can potentially not only survive, but also backcross, making introgression an important aspect of this genus' evolution (*Heliconius* Genome Consortium, 2012, Martin 2013). These events, which have occurred between closely related species as well as between distant relatives within the genus (Zhang et al., 2016, Pardo-Diaz et al., 2012), have led to the formation of new species in the past. These may have benefitted from the acquisition of novel colour pattern alleles from one of the parental species (this would be the case if the novel alleles allowed the new species to become part of a mimicry ring). This overall makes *Heliconius* an excellent animal model system for the study of hybrid speciation, horizontal gene transfer and the species continuum since these concepts are appropriately exemplified in this genus' history (Merrill et al., 2015). Their relevance to the study of speciation genomics have led to the sequencing of the *Heliconius melpomene* genome, which is now used as a reference genome in *Heliconius* genetics: it first presented evidence for introgression between *H. melpomene'*s close relative, *Heliconius timareta*, and the more distantly related silvaniform species, *Heliconius elevatus*. When compared to the *Bombyx mori* genome it revealed expansions in several gene families including chemosensory/olfactory genes and opsins (*Heliconius* Genome Consortium, 2012). The latters are expected because due to being diurnal, butterflies are more visually-oriented than moths, and colour pattern recognition is of great importance in the search for a mate. The formers may appear more surprising: given the large effect of visual signals in mate choice, one may expect chemical signals to play a much smaller role. Actually, chemosensation is crucial for species recognition, since chemical cues like pheromones can be used by co-mimetic species to discriminate against one another and in favour of conspecifics (Meròt et al., 2015).

H. melpomene and its close relatives (such as *H. cydno*) and co-mimics (such as *H. erato*) are the best studied members of the *Heliconius* genus. Research on them has led to an understanding of the genetic architecture of mimicry and colour pattern evolution, on the role of assortative mating in ecological separation between species, and on sympatric speciation (Merrill et al., 2015). However the *Heliconius* genus includes several poorly studied species. The silvaniform clade has received relatively little attention, with most studies focusing on *H. numata,* but it is interesting in its own right. Silvaniform *Heliconius* generally have very different colour patterns from those observed in the *melpomene/cydno* clade, as they usually mimic the genus *Melinaea,* and one species, *H. numata,* is very plastic in the variety of colour patterns it can adopt within the same territory. Phylogenetic relationships between species in the silvaniform and *melpomene* clades can be seen in Fig. 1.

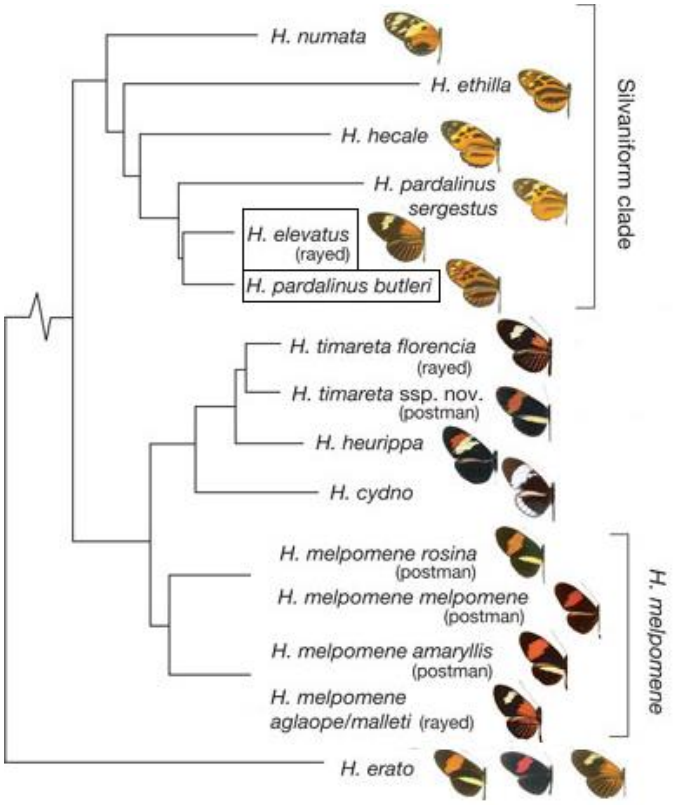


Figure 1. Phylogeny of the silvaniform and *melpomene* clades of *Heliconius* (adapted from *Heliconius* Genome Consortium, 2012).

This study focuses on *H. pardalinus butleri* and its sister species, *H. elevatus*, two sympatric silvaniform species that live in Perù and across the Amazon Basin (*Fig. 2*). *H. elevatus* is believed to have formed by hybridization with *H. melpomene* (*Heliconius* Genome Consortium, 2012), making it a good model for adaptive introgression. Both are a good study system for sympatric speciation in the silvaniform clade, as there are no geographical barriers between them but several ecological barriers are believed to have driven their divergence, including but not limited to their colour pattern, their sex pheromone blend, their flight dynamics, wing shape and preferred food plant (Queste, 2015). Introgression of colour pattern genes from *H. melpomene* into *H. elevatus* allowed the latter to join the rayed mimicry ring to which several other species belong. This may have occurred via backcrossing of the F1 hybrids to the parental

silvaniform species, and by the subsequent reproductive isolation of individuals carrying the introduced allele. This situation is analogous to what may have happened in *H. heurippa,* another hybrid species. The result of adaptive introgression at the colour pattern loci is that the outcomes of phylogenetic analysis differ depending on the region of the genome that is taken into account. Regions that are unlinked to colour loci produce a signal that reflects that of the whole genome, with different species clustering with their close relatives, whereas in phylogenetic trees based on colour loci the species that share a pattern cluster together regardless of their general taxonomic relationship (*Heliconius* Genome Consortium, 2012). This is observed, in fact, in *H. elevatus*: phylogenetic analyses based on its main colour locus place it closest to rayed races of *H. melpomene* even though they are very distant relatives (*Heliconius* Genome Consortium, 2012). Because this locus is linked to loci controlling assortative mating in *H. melpomene* (Merrill, 2015), *H. elevatus* must have inherited a behaviour that cause it to mate preferentially with individuals of the same colours (Queste, 2015).

Thus, colour pattern acted as a magic trait to promote the lineage's divergence. When *H. elevatus* is compared to its sister species *H. pardalinus,* 95% of their genomes show no divergence and is consistent with a situation of complete admixture (F_{ST}=0), but the remaining 5% is strongly divergent between the two (Dasmahapatra, in prep). The aim of this study is to use F2 hybrids to characterize the genetic architecture responsible for the colour pattern of the two species. The two main priorities are to assess whether this architecture is based on multiple loci or a single polymorphic locus (as seen in *H. numata*), and whether the arrangement of the loci mimics that observed in other *Heliconius* species, regardless of whether they are multiple or a single one. In addition to this, this research projects aims at characterizing the genetic structure of another trait that is thought to be important in *Heliconius* speciation: the composition of male sex pheromones. Unlike the colour pattern, male sex pheromones have not been studied in-depth in these species before, and remain poorly characterized in the *Heliconius* genus as a whole as most work on these butterflies' chemical ecology has been carried out on a different class of pheromones, the antiaphrodisiacs. This study attempts for the first time to explore the genetic organization of loci responsible for male sex pheromone production, with a detailed analysis of hybrid phenotypes.

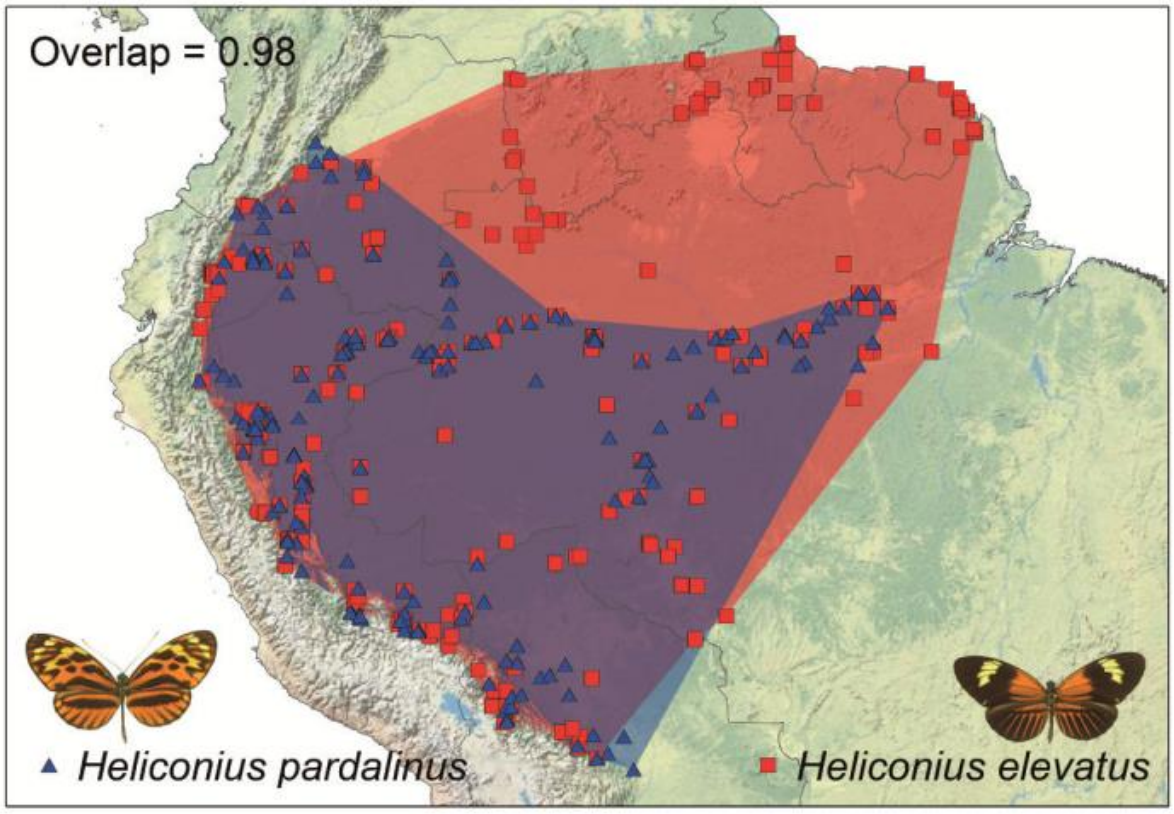


Figure 2. The overlapping ranges of *H. pardalinus* and *H. elevatus* (Rosser, 2012). Triangles and squares indicate collection sites.

Chapter 2

Phenotypic analysis of the *H. pardalinus* and *H. elevatus* colour pattern.

1. Introduction

1.1 The colour pattern as a reproductive barrier

The wing colour patterns are perhaps Lepidopterans' most striking trait. Their functions range from crypsis (avoiding detection by blending in the background) to aposematism (signalling toxicity with bright alarming colours). In addition, as butterflies are very visually-oriented organisms, the colour pattern is thought to be one of the main methods they use to recognize their conspecifics, even when the patterns that distinguish two different species are minute, as seen in some species of Lycaenid butterflies (Fordyce et al., 2002). Lycaenid wing patterns are usually made up of several small spots whose arrangement, size and colours varies between species, occasionally in very subtle ways. In the two recently diverged parapatric North American species *Lycaeides idas* and Californian *Lycaeides melissa* this difference is given simply by a slight variation in the size of certain spots (Fordyce et al., 2002). In spite of the small difference, *L. idas* males are able to distinguish their conspecific females by the colour pattern alone, and discriminate positively towards them when presented with the choice to mate with them or with *L. melissa* populations. While pheromones play an important part in this discrimination (*L. idas* live females whose hindwing pattern was scraped off were still preferentially approached by males of their own species), the discrimination remained true when males were presented with odourless wing models (Fordyce et al., 2002). However, the allopatric population of Nevadan *L. melissa* shows no preference towards their own species, implying a stronger pressure on maintaining the discrimination in parapatric species where interspecific encounters are most likely. In fact, while *L. idas* and Californian *L. melissa* are parapatric, the latter is adapted to a more mountainous habitat than the former, potentially making the exchange of genes disadvantageous as the two species have different requirements, and thus explaining why there may be pressure to reinforce reproductive barriers. But Nevadean *L. melissa* do not usually encounter either *L. idas* or their Californian conspecifics, so they have no pressure to develop barrier reinforcement (Fordyce et al.,

2002). In *Heliconius,* the colour pattern has a central role in mate choice (Jiggins et al., 2001), as individuals belonging to one mimicry ring will mate preferentially with partners sporting the same pattern.

1.2 What makes a colour pattern? The Nymphalid groundplan and its variations.

All Nymphalid wing patterns including the ones seen in *Heliconius* follow a basic groundplan (Nijhout et al., 1990) (*Fig. 3A*): a nonspecific ancestral Nymphalid butterfly has a wing that can be divided into three regions- the base, the centre and the border. Each of these has its own basic fixed elements that form symmetrical smooth bands along the wing (plus spots in the case of the border region) and that can be evolutionarily modified to change size or shape, to disappear altogether, or to become fused together (Nijhout, 2002). These elements are peculiar compared to the colour patterns seen in mammals because they form individually: each evolutionarily derived

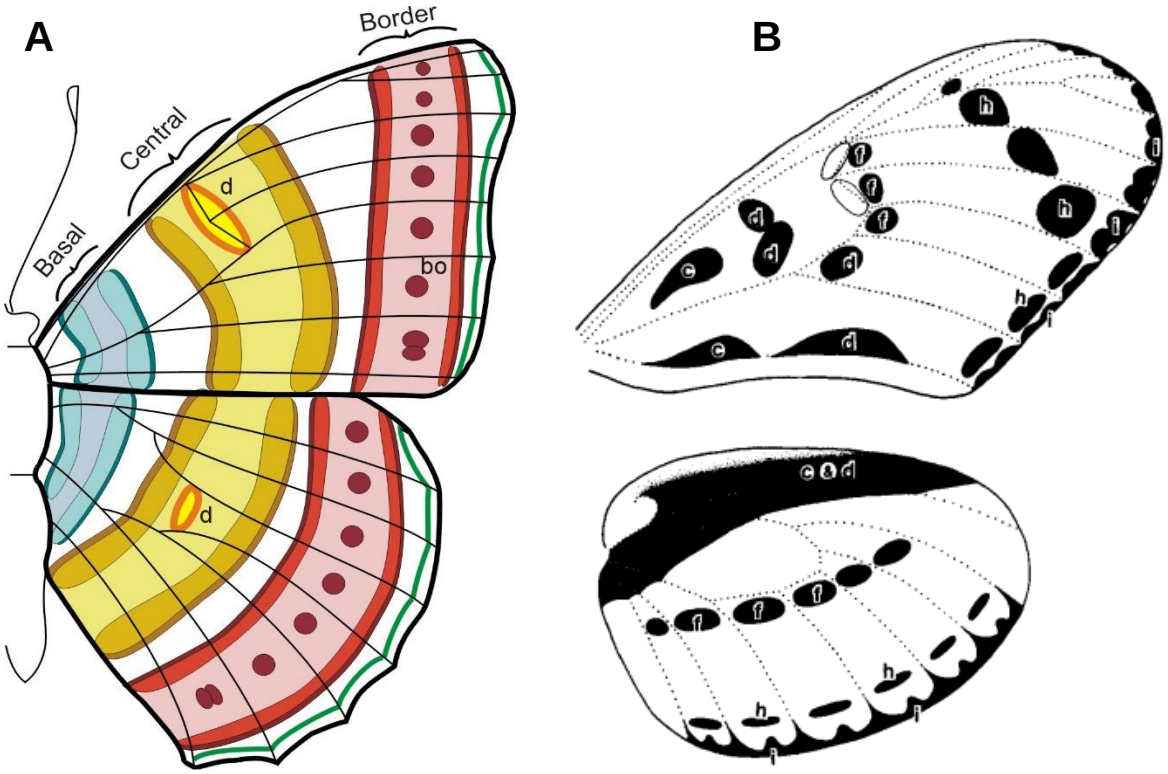


Figure 3A. The general Nymphalid groundplan (picture from http://sites.biology.duke.edu/nijhout/patterns2.html Nijhout, n.d.). B. The derived *Heliconius* groundplan showing the variable black elements (picture from Nijhout et al., 1990).

element on a butterfly's wing can be traced back to the original basic groundplan, whereas the same cannot be said about each stripe of a zebra or each spot of a leopard (Nijhout, 2002). This means that as evolution proceeds, butterflies are "free" to tweak their own pattern as their

ecology and lifestyle dictate, thus giving rise to the striking cryptic or aposematic patterns and the eyespots seen in many species of Nymphalids.

The *Heliconius* groundplan is derived from the Nymphalid one (*Fig. 3B*). It is most easily observed in silvaniform species since their patterns are mostly comprised of small melanic (black) dots on wide orange or white backgrounds, whereas the more *melpomene*-like patterns have simpler, more minimalistic coloured shapes with large black areas (Nijhout, n. d.). The black dots of the basic *Heliconius* groundplan can expand and fuse with each other to form the large black areas seen in the *melpomene/cydno* clade, and the areas in-between the black elements can evolve different colours or become white (Nijhout et al., 1990).

1.3 The *Heliconius* pattern: mimicry and assortative mating

The *Heliconius* colour pattern, in addition to being a cue for mate recognition, has enough ecological importance due to its aposematic function to effectively act as a magic trait in at least some species. Many species of *Heliconius* are sympatric, yet they manage to remain reproductively isolated. As stated before this phenomenon in animals can be caused by genomic incompatibilities that make the F1 hybrids non viable or sterile, resulting in post-zygotic isolation and the subsequent lack of gene flow. This is not always the case for *Heliconius* hybrids, which are often not only healthy, but also able to produce offspring by crossing with one another or by backcrossing to parental species (Mallet et al., 2007). In spite of this, post-zygotic ecological barriers do exist: in the wild hybrids are strongly selected against due to their pattern not falling within either parental mimicry ring and thus not being recognized by predators as a warning signal (Merrill et al., 2012). Another factor acting against hybridization is disruptive sexual selection: in *H. cydno* and *H. melpomene,* while F1 hybrids were shown to readily hybridize with each other in captivity, parental species often discriminated against them (Naisbit et al., 2001). So the *Heliconius* colour pattern prevents hybrids from backcrossing to their parental species, while also being an important aspect of pre-zygotic reproductive isolation between species. Much like a parental species will discriminate against hybrids, it will also discriminate against species which sport a different colour pattern, even when the two species would be partially interfertile (Merrill et al., 2015, Kronforst & Papa, 2015). In fact, mate preference has been found to be genetically linked to colour pattern loci (Kronforst et al., 2006; Merrill et al., 2011). *H. cydno galanthus* and *H. pachinus* are distinguishable by the colour of the bands on their wings, which are white in the former (white being a structural colour) and yellow in the latter. Colour pattern and mate preference both map

to the *wingless* gene region (Kronforst et al., 2006). Because *wingless* is responsible for the production of red/orange ommochromes (a type of pigment) in other Nymphalid species, it makes sense that it would retain that function in *Heliconius pachinus,* given that the yellow pigment is also an ommochrome (Kronforst et al., 2006). Similarly, in *H. melpomene* and *H. cydno,* male preference for red patterns (seen in *H. melpomene*) rather than white ones (seen in *H. cydno*) maps to the same genetic region as the loci that control red pattern elements (mainly the B locus for the red forewing band) (Merrill et al., 2011). Thus, the *Heliconius* colour pattern is under divergent ecological selection and at the same time it dictates mating preference, potentially making it a good example of a magic trait, at least in some species. (Merrill et al., 2015).

H. elevatus and *H. pardalinus,* the models of this study, are sister species but they belong to different mimicry rings despite the massive overlap of their ranges. While *H. elevatus* has a typical rays and dennis pattern that mimics those found in Amazonian *H. melpomene* races, *H. pardalinus* has complex spotted patterns which mimic the colours of butterflies from the genus *Melinaea,* that belongs to a different Nymphalid sub-clade, the Ithomiini (*Fig. 4*).

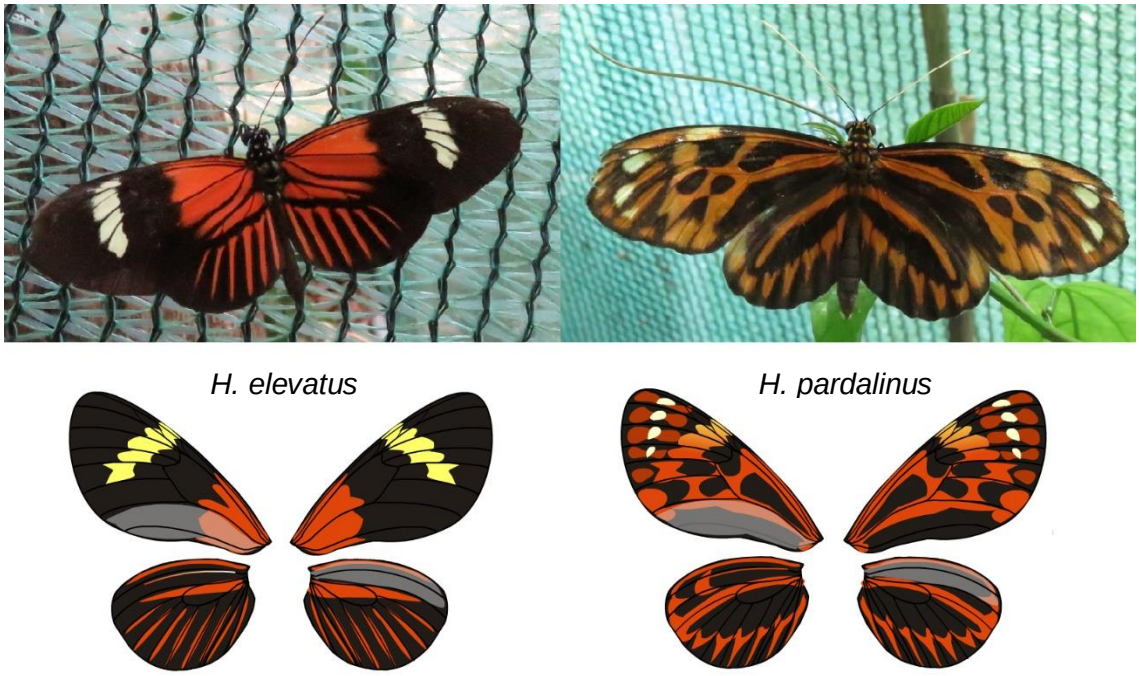


Figure 4. The *H. elevatus* and *H. pardalinus butleri* colour patterns. Photos by Lucie Queste.

H. elevatus and *H. pardalinus* can produce viable hybrids, but rarely do so in nature. This means that there must be reproductive barriers between the two species, and it is likely that colour pattern constitutes a "magic trait". Indeed, although *Heliconius* hybrids are often fertile (Naisbit et al., 2001), the colour pattern is subject to strong disruptive selection, the main opposing force to hybridization. In a mimicry ring, it is advantageous to convergently evolve the same colour pattern (Naisbit et al., 2001). Thus it makes sense that colouration may influence mate choice. This is seen in *H. melpomene* and *H. cydno*, with assortative mating between individuals bearing the same pattern, and this effect is stronger in sympatric populations where there is more pressure to avoid mixing (Jiggins et al., 2001). It is via this association between mate choice and colour pattern that "pure" colour patterns are maintained in sympatric, recently diverged species of *Heliconius,* and colour pattern genes, with their evolution driven by strong disruptive selective pressures, can form clear islands of divergence.

1.4 The *Heliconius* mimicry toolbox.

Most studies of the *Heliconius* colour pattern have focused on the *melpomene*/*cydno* clade and on the *H. melpomene* co-mimics such as *H. erato* (Sheppard et al., 1985; Nadeau et al., 2014). *H. melpomene* is often a mimic of *H. erato,* from which it is believed to have diverged 15-20 million years ago (making the two rather distant relatives). Across their range in the Neotropics they have evolved to become co-mimics, with the model colour pattern varying with the territory (Bates, 1862). Both of them exist in a wide variety of races, whose colours vary from the typical postman pattern consisting of a few simple band shapes on a black background to a more complex rayed pattern, but races of both species found in the same territory will almost always be co-mimic (Sheppard et al., 1985, Kronforst & Papa, 2015). The alleles involved in this process are homologous between the two (Nadeau et al., 2014).

Broadly speaking, the *Heliconius* mimicry toolbox seems to be comprised of a few large genes of developmental significance, whose function is sometimes linked to the production of the pigments themselves, as is the case for *wingless,* mentioned in 1.3. Within these genes, several loci are found which correspond to *cis*-regulatory elements of the main gene. The action of these loci produces differently shaped spots (Kronforst & Papa, 2015; Jiggins et al., 2017). This implies quite a flexible control mechanism for the colour pattern, where different regulatory elements can become activated or inactivated or acquire new allelic variants and easily create large-scale modifications in the looks of the wing.

The *H. melpomene* colour pattern is controlled by a few loci that have been very well-characterized in a series of studies. Originally, Turner & Crane (1962) had described several putative loci for *H. melpomene* (12 loci) and *H. erato* (16), each controlling one or few elements of the pattern, but with the advent of molecular techniques in genetics, most of these traits were conflated into fewer loci. In rayed *melpomene,* the main locus that controls the presence of the typical basal orange elements on both fore and hindwing, commonly referred to as "dennis", is the D locus, which also controls the presence of hindwing rays. D is closely linked to another locus, B, which controls the presence of red pigmentation in the forewing band of postman races (Sheppard et al., 1985). In *H. erato* the main red and orange elements are controlled by the D locus, which is homologous to *H. melpomene*'s B/D loci (Baxter et al., 2008). The rays were at first described as a separate locus, R (Turner & Crane, 1962), however with further studies it was confirmed that the rays phenotype is given by an allele of the D locus, D^R (Sheppard et al., 1985).

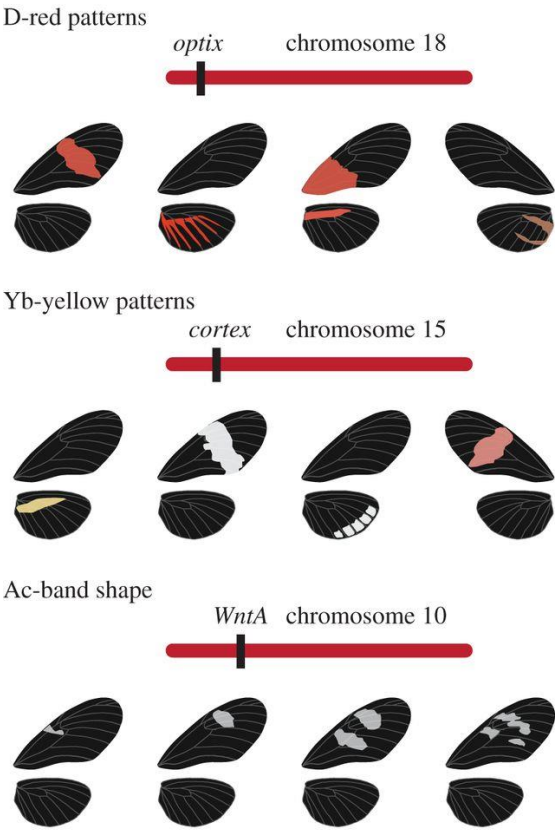


Figure 5. The main *Heliconius* colour pattern genes that have been researched in depth thus far. An additional gene of this kind is presumably found on chromosome 1, and may contain the K locus from *H. numata.* Picture adapted from Jiggins et al. (2017).

It is now known that all the red/orange elements of the *melpomene/erato* patterns may be attributable to the action of a single gene, *optix,* that controls production of the red pigment (*Fig. 5*). All the variations in shape of the red/orange elements are given by the *cis*-regulatory elements associated with *optix,* each of whom acts as a locus with its own set of alleles to determine the distribution and shape of different red shapes (Reed et al., 2011, Supple et al., 2013; Wallbank et al., 2006; Lewis et al., 2016,). Thus, the *cis*-regulatory region of *optix* appears to be organized as a set of tightly linked consecutive genetic "cassettes", each determining the positional information for one red element of the pattern- B and D being an example of such regulatory modules (Supple et al., 2013).

A similar system is found in the *WntA* gene, which encodes an extracellular signalling molecule active during the butterfly's development. *WntA* is responsible for variations in the shape of melanic elements on the forewing collectively known as forewing shutter, found in several subspecies of *H. melpomene, H. erato* and *H. cydno* (*Fig. 5*). Much like *optix, WntA* has a very conserved sequence across the whole genus, meaning that its variable phenotypes too are given by *cis*-regulatory elements (Kronforst & Papa, 2015).

Similarly to the orange elements and their B/D loci and the melanic elements with *WntA*, the yellow and white elements of *H. melpomene* wings are given by another small group of closely linked loci known as Yb-N-Sb, each of whom controls the presence, shape and intensity of different white or yellow elements (Sheppard et al., 1985, Mallet, 1989). Very recent studies by Nadeau et al. (2016) have revealed that in several species of *Heliconius* Yb and other loci responsible variations in the yellow/white elements map to a gene called *cortex* (*Fig. 5*), which is also associated with colour pattern variations in the moth *Biston betularia.* This gene also produces the pink colour seen in some *H. melpomene* races by controlling the presence of white scales on the red band region (Nadeau et al., 2016). It appears to have a conserved function across Lepidoptera and it seems to have undergone rapid evolution in this taxon (Nadeau et al., 2016; Jiggins et al., 2017). In addition to red/orange and white/yellow loci, *H. melpomene,* as well as *H. cydno* and the more distantly related *H. erato,* also carry several small effect QTLs associated with the shape and size of melanic spots (Martin et al., 2012, Papa et al., 2013).

1.5 The silvaniform colour pattern and its genetic architecture

Compared to the *melpomene* clade, the silvaniform colour patterns have received little attention. Silvaniforms do not usually fall within the postman or rayed mimicry rings, and instead mimic butterflies of the related genus *Melinaea,* developing patterns characterized by complex arrangements of melanic spots and stripes on orange, yellow and white backgrounds (Joron et al., 2006). *Melinaea* butterflies are themselves toxic and very variable in colour, but unlike *Heliconius,* they do not tend to form strict mimicry rings: rather, across their geographical distribution, these Ithomiine adopt a complex mosaic of local variants, with several types of differently patterned butterflies inhabiting the same range (Joron et al., 2006). This implies that *H. numata*, whose different morphs mimic several different species of *Melinea,* is under a different kind of selective pressure than *H. melpomene/erato*. While evolutionary convergence towards a single pattern is encouraged in the latters, *H. numata* instead had to evolve a method for genetic control of its

pattern that would allow it to flexibly shift from one layout to the other depending on the types of *Melinaea* patterns present in that territory (Joron et al., 2006). Thus, sympatry with *Melinaea* led to a phenomenon known as polymorphic mimicry: the coexistence of *Melinaea* species with different colour patterns promoted the evolution of polymorphisms in *H. numata,* whose multiple races defined by the different patterns are now arranged in a mosaic that reflects the distribution of the *Melinaea* species it mimics (Joron, 1999).

*H. numata'*s complex colour pattern is given by a single polymorphic supergene locus P formed from chromosomal rearrangements (Joron et al., 2011). It is homologous to *H. melpomene'*s N-Yb-Sb complex (Joron et al., 2006), with each polymorphism corresponding to a different type of *Melinaea* pattern, and within the same area one can expect to find several different *H. numata* phenotypes (Joron et al., 1999). Additionally, races of *H. numata* are able to interbreed. There is a clear dominance hierarchy between certain alleles of the P locus, so crosses between individuals with different phenotypes do not produce an intermediate pattern, implying that the cost to exchanging alleles that one would expect in strict mimicry rings is not present in *H. numata'*s case (Jones et al., 2012). Le Poul et al. (2014) studied this dominance hierarchy in depth and found that sympatric pairs of alleles generally have very strict dominance relationships, so the heterozygous hybrid phenotypes fall into one of the two parental mimicry rings and suffer no loss in fitness. Sympatric alleles can have incomplete dominance however: for example, crosses between the sympatric alleles *tarapotensis/acruella* or *elegans/aurora* produce hybrids with intermediate phenotypes (Le Poul et al., 2014). However these hybrids' pattern also comes at no cost, as while it is distinct from the parental phenotypes, it still falls within a separate mimicry ring (Le Poul et al., 2014). Meanwhile, in parapatric species, dominance relationships are less strict and how dominant each morph is depends on the dominance relationships of the single independent pattern elements, with black elements generally taking priority over orange and yellow elements and orange elements over yellow ones (Le Poul et al., 2014). Of all the *H. numata* morphs, the one carring the *bicoloratus* allele can recombine with all the others, and this allowed mapping of the P locus, which turned out to correspond to the *cortex* gene much like white/yellow pattern elements in other *Heliconius* species (Nadeau et al., 2016).

However what is true in *H. numata*'s case cannot be assumed to hold true for the rest of the silvaniform clade. *H. hecale* and *H. ismenius* are two silvaniform species whose colour pattern was also analyzed, and it was shown that they have a multilocus architecture comparable to that

observed in *H. melpomene* (Huber et al., 2015). This means that we cannot predict with certainty the genetic organization of the *H. elevatus* and *H. pardalinus* colour patterns. These species' colour pattern genetics have never been studied before, so the question is whether they have a single-locus architecture like *H. numata*, or a multilocus architecture more like *H. hecale* and *H. ismenius*. The aim of this study is to characterize the genetic structure underlying generation of the colour pattern in *pardalinus* and *elevatus,* using hybrid wing phenotypes and their frequencies as indicators of the number of loci that are affecting these traits and of the dominance relationships between their alleles. In addition to this, another important aim is to search for evidence of homologies between them and other clades, as well as evidence of introgression from *H. melpomene* into *H. elevatus.* In the process, it will be clarified whether their genetic structure is more similar to *H. numata* or to *hecale* and *ismenius.*

2. Materials and methods

422 F2 hybrids between *Heliconius elevatus* and *Heliconius pardalinus* were captively bred by the Dasmahapatra laboratory prior to the beginning of this research project, in 2014, and used for colour pattern analysis. 148 of these hybrids were bred in York. These 148 individuals came from 5 separate broods. 274 more F2 individuals were bred in Perù in 2015-2016 from 8 different mothers. In addition, 116 F1 hybrids (35 from York and 81 from Perù) were used for determining the heterozygous phenotype. In addition to these, a single brood of 34 backcrosses to *H. pardalinus* (B3) was also scored.

Both pairs of wings from these 572 individuals were scanned twice at a resolution of 300dpi in true colours against a white background, once for the ventral side and once for the dorsal side. Scanning front and back was necessary due to some elements of the pattern only being visible on one side. where needed, The images' brightness and contrast were enhanced in Adobe Lightroom in order to make the single pattern elements clearer to the naked eye.

Identification and scoring of the phenotype was based on the *Heliconius* groundplan proposed by Nijhout & Wray (1990), which described the major areas of phenotypic variations and the main shapes making up the pattern of *H. melpomene. H. melpomene'*s phenotypes do not present the same pattern elements as the F2 hybrids here described, but the layout of the wing remains very similar and thus parallels can be drawn.

In total, 9 elements were scored plus 3 smaller details that appeared more ambiguous due to showing high levels of variation in the parental species as well *(Fig. 6)*. The 9 major elements were: the shape of the dennis marking, the shape of the horizontal bar on the hindwing, the shape of the hindwing rays, the colour of the white band on the forewing (which is faded in *H. pardalinus* and full in *H. elevatus*), the presence of the orange band on the forewing, of two dumbbell-shaped spots in the discal region of the forewing (counted as two separate traits, but later analysed as a single trait), of orange rays and of a secondary white band in the apical region of the forewing (as two separate traits). The three minor elements were the rim spots, the *elevatus* tornal dot and the *pardalinus* hindwing bridge (*Fig. 6*). While they were initially scored they were eliminated from subsequent analysis due to the aforementioned high levels of variation.

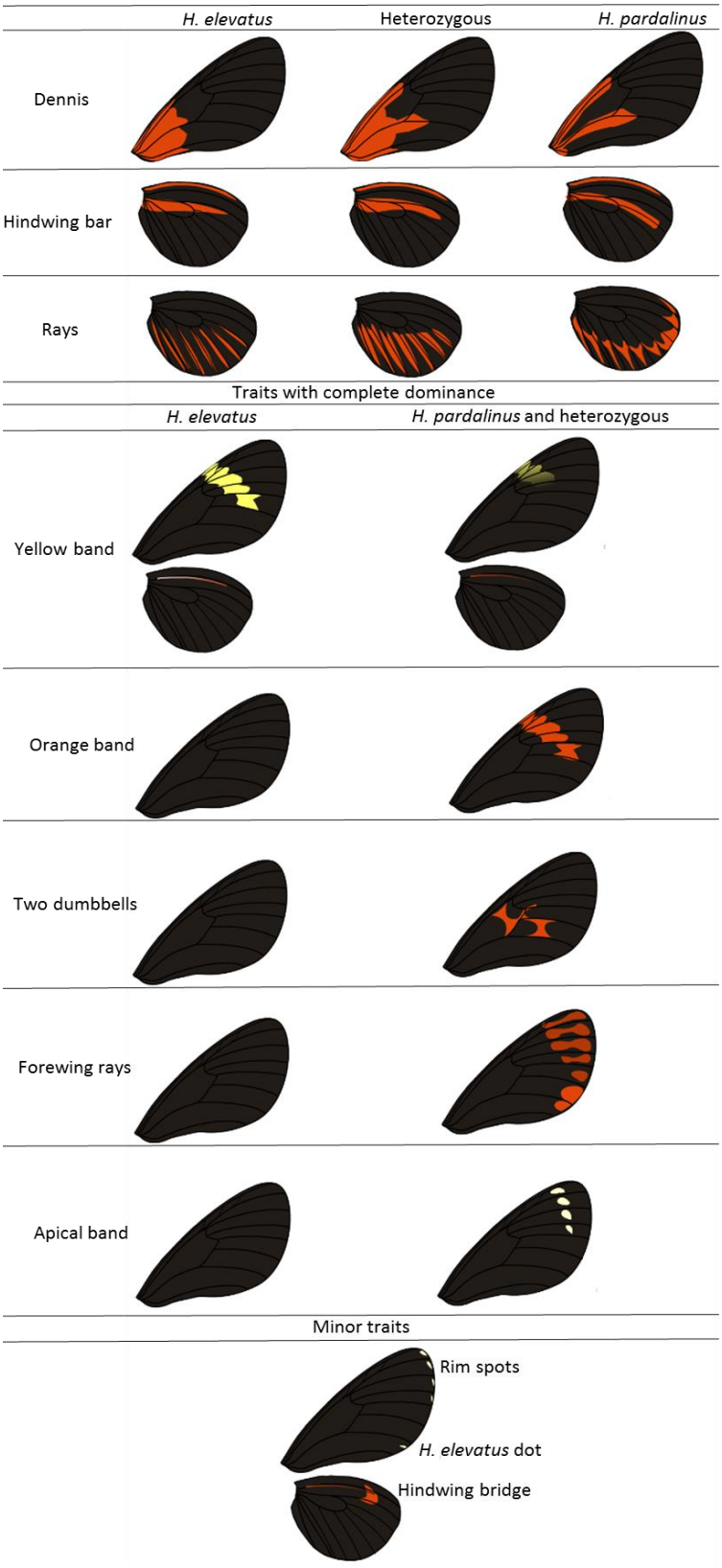


Figure 6. Phenotypes of all scored traits. The pure species phenotypes are shown as well as the heterozygous phenotypes if any.

The Chi-squared test for Goodness of Fit was used to determine whether the frequencies of each phenotype for each trait followed Mendelian expectations, and thus whether the single traits were potentially controlled by a single gene. In the case of complete dominance, the expectation is the frequencies of the two phenotypes of a trait should be consistent with a 3:1 ratio, while in the case of incomplete dominance the ratio should be 1:2:1 as the trait will appear to have three phenotypes rather than two. In addition, evidence for linkage between traits was also tested with pairwise Chi-squared tests, assuming a 9:3:3:1 ratio of parental : recombinant : recombinant : parental phenotypes.

3. Results

3.1 Colour pattern analysis.

Several putative loci were identified for the colour pattern. Most of them display the expected frequencies for Mendelian loci with two alleles and incomplete dominance, but even in the loci that fit this model some degree of quantitative variation in the shape and size of the spots was detected, suggesting the presence of additional loci of small quantitative effect. Examples of F1, backcross and F2 phenotypes are in Fig. . Note that occasionally some individuals had to be left out of the analysis due to wing damage that prevented accurate scoring of the phenotypes.

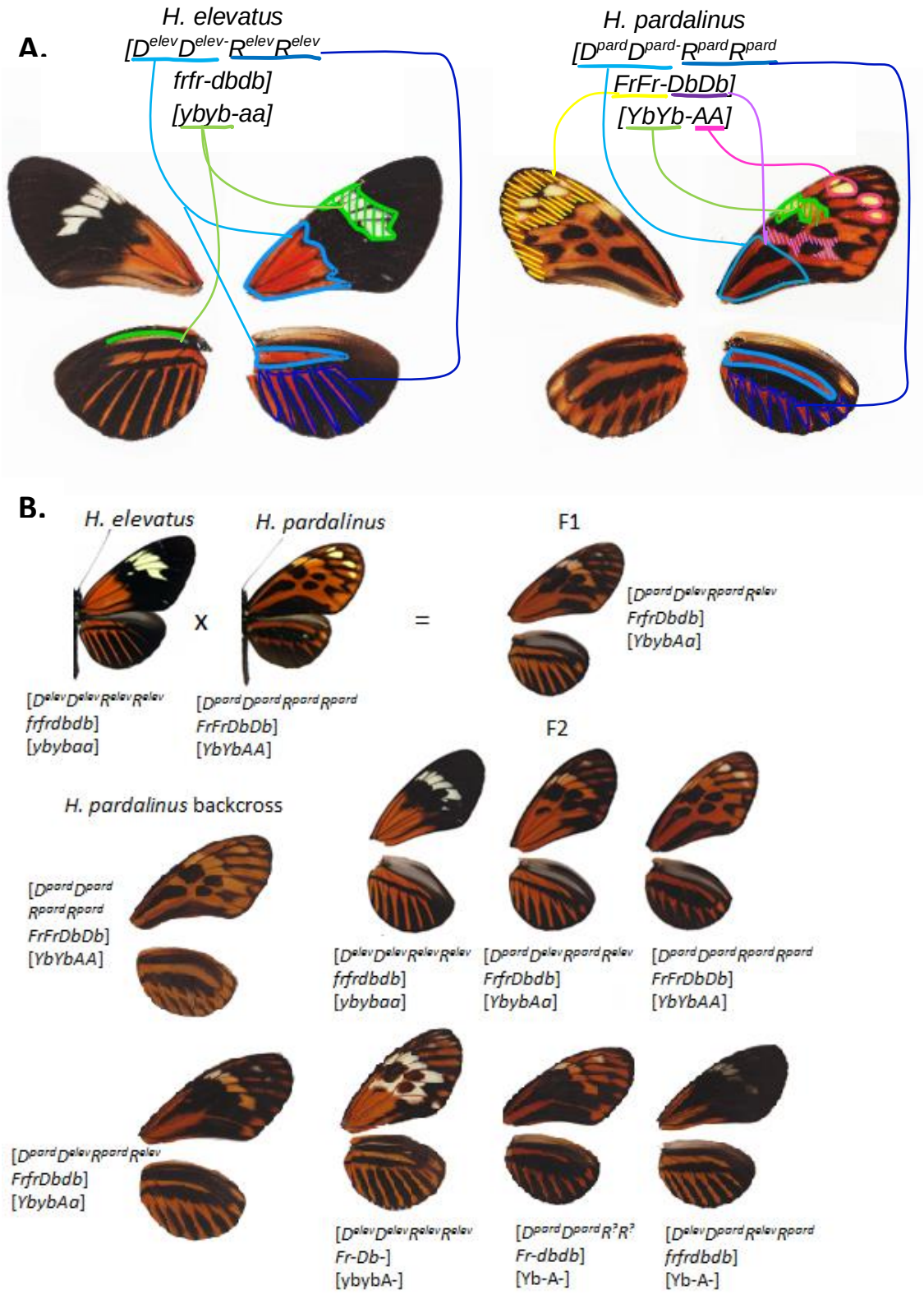


Figure 7A. Pure species phenotypes and putative genotypes, with elements of the phenotype highlighted. *B.* Comparison of phenotypes seen in the pure species, F1, F2 (though not all F2 phenotypes are shown) and backcrosses. Putative genotypes are also reported.

Dennis, hindwing bar and rays- The dennis (hereby tentatively referred to as D), the horizontal hindwing bar (B) and the rays (R) are the most prominent features of the *H. elevatus* and *H. pardalinus* pattern. The dennis and the bar are monomorphic in the parental species, while the rays are monomorphic in *elevatus* but can take on a variety of phenotypes in *pardalinus.* In *H. elevatus,* the combined action of these three loci produces the typical pattern seen in rayed *H. melpomene* races that inhabit the Amazon basin, with a uniform orange area in the basal region of the forewing, a wedge-shaped hindwing bar and long, narrow, tack-shaped rays (*Fig. 6-7*). Thus, the *H. elevatus* genotype for these three traits would be $D^{elev}D^{elev}/B^{elev}B^{elev}/R^{elev}R^{elev}$. Conversely in *H. pardalinus* the genotype is $D^{pard}D^{pard}/B^{pard}B^{pard}/R^{pard}R^{pard}$ and the dennis appears as a pair of narrow orange bands running along the contours of the discal cell, separated by black areas. The hindwing bar is long and curved and spans most of the hindwing's width, and the rays are tend to be short, wide and Y-shaped (*Fig. 7*), albeit there is a degree of quantitative variation in their shape, and there are *H. pardalinus butleri* populations with noticeably different rays.

All these alleles have incomplete dominance, so heterozygous individuals ($D^{elev}D^{pard}/B^{elev}B^{pard}/R^{elev}/R^{pard}$), including the entirety of the F1 hybrids, have phenotypes that look like neither parental species. In these individuals the spots that make up the dennis take on an intermediate shape between the two homozygous phenotypes and the hindwing bar is curved, but not as prominently as in $B^{pard}B^{pard}$. The rays are commonly long but wider than those seen in $R^{elev}R^{elev}$, and while the F1 have a stable rays phenotype, some quantitative variation is observed in regards to how Y-like their shape is in the backcrosses and in the F2: $R^{elev}R^{pard}$ individuals can have really pronounced branching in their rays, thus adopting a more *pardalinus*-like shape, while others only have a small indentation (*Fig. 11*). In the broods produced in Peru, a second *pardalinus* R allele R^{dil} was introduced from a population of *H. pardalinus dilatus,* which are very similar to *H. pardalinus* save for the rays' shape, and the corresponding homozygous phenotype has wide rays that are joint together in a single jagged orange spot.

The expected proportions of F2 phenotypes for a trait with incomplete dominance are 1:2:1, with the heterozygous phenotype appearing twice as frequently as either homozygous parental phenotype. The dennis and hindwing bar appear with frequencies consistent with the expectation of incomplete dominance. The frequencies for the dennis are 91:208:117 for $D^{elev}D^{elev}$:$D^{elev}D^{pard}$:$D^{pard}D^{pard}$ (Chi-square, X^2=2.17, df=2, p=0.34), those for the hindwing bar are 111:218:88 for $B^{elev}B^{elev}$:$B^{elev}B^{pard}$:$B^{pard}B^{pard}$ (Chi-square, X^2=3.4, df=2, p=0.18). However for the rays,

the frequencies of $R^{elev}R^{elev}$:$R^{elev}R^{pard}$:$R^{pard}R^{pard}$ are 37:77:34 for the York broods, which fit Mendelian expectations (Chi-square, X^2=0.36, df=2, p=0.83), and 57:155:57 for the Peru broods, wherein they do not fit expectations (Chi-square, X^2=6.25, df=2, p<0.05).

In the backcross brood, born from an F1 mother and a *pardalinus* father, one would expect a 1:1 ratio of *pardalinus* homozygous:heterozygous alleles, and this expectation is met by the frequencies of the three phenotypes. The frequencies are 14:19 for $D^{elev}D^{pard}$:$D^{pard}D^{pard}$ (Chi-square, X^2=0.48, df=1, p=0.49), 15:18 for $B^{elev}B^{pard}$:$B^{pard}B^{pard}$ (Chi-square, X^2=0.12, df=1, p=0.73) and 15:18 for $R^{elev}R^{pard}$:$R^{pard}R^{pard}$ (Chi-square, X^2=0.12, df=1, p=0.73).

Linkage between D, B and R. With the assumption that the genotype can reliably be predicted from the phenotype thanks to incomplete dominance, a putative arrangement for the D, B and R loci can be inferred using the frequency of recombinant individuals. Knowing that a recombinant phenotype indicates that a crossover has occurred between the chromosomes that carry the loci, and knowing that the chances of a crossover between two given loci is directly proportional to their physical distance on the chromosome, one can use the percentage of recombinant phenotypes to infer the distance between loci and evaluate their linkage. The unit of measure of distance between loci is the centimorgan (cM). Two loci are located at 1 cM from one another if in a single generation the proportion of crossover events (and thus, recombinant phenotypes) between them is 0.01 (1%). Linkage and genetic distance between D, B and R could only be inferred for the York broods due to the presence of extraneous alleles in the Peruvian broods. The percentage of recombinants is 25.7% (25.7 cM) between dennis and bar, 31.8% (31.8 cM) between dennis and rays and 27.7% (27.7 cM) between bar and rays.

Female *Heliconius* butterflies do not form recombinant gametes, so in the backcross brood no recombinant phenotypes should be observed: all individuals are expected to carry fully parental genotypes, either $D^{pard}D^{pard}/B^{pard}B^{pard}/D^{pard}D^{pard}$ or $D^{elev}D^{pard}/B^{elev}B^{pard}/R^{elev}/R^{pard}$. In fact, no recombination is observed between the D and the R locus, but 8 seemingly recombinant individuals were observed between the dennis/rays and the bar (26.5% of the brood). It is possible that these were not actually recombinants, and that quantitative variation affected the phenotype to a point where it was indistinguishable from that of a recombinant.

Orange forewing band. This locus, Ob, controls the presence of an orange band on the forewing, which is always present in *H. pardalinus* (ObOb) and always absent in *H. elevatus* (obob). All F1

hybrids and backcrosses have an orange band, with no observed instances of any intermediate phenotype. This suggests complete Mendelian dominance of the *H. pardalinus* allele. The observed frequencies in the F2, 321:92 (Ob-:obob), do not differ significantly from the expected 3:1 ratio (Chi-square, X^2=1.49, df=1, p=0.22).

Two dumbbells. The *pardalinus* colour pattern is characterized by several spots in the central region and along the border of the wing, distinguishing it from the *elevatus* forewing pattern which is mostly black and features no spots other than the dennis and the white band. These spots are not easy to score due to displaying the highest amount of quantitative variation in their F2 phenotypes of any other pattern element. Although the F1 phenotypes for both traits are more variable in shape than one would expect from two simple 1-locus systems, they usually still look distinct from the parental phenotypes.

The dumbbells are absent in *elevatus* and present in *pardalinus* (wherein they are both fully visible from both the ventral and the dorsal side of the forewing), and in most F1s they are partially present, with only 1 York F1 individual out of 35 (1 out of 80 for the Peru F1) showing no dumbbells at all. If one considers the two dumbbells as two separate loci and assumes partial presence as the heterozygous phenotypes, they would be expected to each have 1:2:1 ratios (*elev*:incomplete:*pard*) in the F2s, but these expectations are not met in either the York or the Peru broods. Instead, assuming them as a single trait with co-dominance or incomplete dominance produces frequencies of 46:71:31 for the York F2, not significantly different from the expected ratio (Chi-square, X^2=3.28, df=2, p=0.19). In the Peru F2 even with this method the frequencies (66:157:46) remain significantly different from the expectations (Chi-square, X^2=10.5, df=2, p<0.01). For the purpose of these calculations, any phenotype with at least one incomplete dumbbell was counted as heterozygous. In the backcrosses, there are no individuals without dumbbells at all and the frequencies of 20:14 are not significantly different from the expected 1:1 ratio (Chi-square, X^2=0.74, df=1, p=0.39). It is worth noting that there is an alternative way to score the two dumbbells, assuming they are a Mendelian trait with complete dominance of the *pardalinus* allele and that the incomplete phenotypes are an effect of QTLs. In this case the frequencies become 300:110, which are consistent with the expected 3:1 ratio (Chi-square, X^2=0.42, df=1, p=0.64).

Forewing rays. The other variable feature of the *pardalinus* wing, these rays are observable along the edge of *H. pardalinus'*s forewing, and are always absent in *H. elevatus*. They appear incomplete and narrower than they are in *H. pardalinus* in all the F1 individuals from both York and Peru where they could be observed, however due to the apical portion of the wing being the most exposed to wear and tear, they could not always be scored with confidence. Attempts to score the three phenotypes in the F2 resulted in frequencies of 26:82:33 (absent:incomplete:present) in the York broods, significantly different from the expected frequencies (Chi-square, X^2=6.94, df=2, p<0.05) and 66:141:57 in the Peru broods, which fit the expectations (Chi-square, X^2=1.84, df=2, p=0.4). In the *pardalinus* backcrosses the forewing rays are always present to some extent and attempts to score the heterozygous and the pure *pardalinus* phenotype produced frequencies of 12:11, which fit expectations (Chi-square, X^2=0.01, df=1, p=0.99) albeit many wings could not be scored due to damage around the edges. Like the dumbbells, the forewing rays can be scored as a simple Mendelian locus with complete dominance of the *pardalinus* allele, assuming the heterozygous phenotype is given by separate QTLs, and the cumulative frequencies thus obtained, 295:90, satisfy the 3:1 expectations (Chi-square, X^2=0.45, df=2, p=0.5).

White band. This locus (Yb) controls the presence of white scales on the forewing band and on the subcostal vein of the hindwing. The parental phenotypes are monomorphic: in *H. elevatus,* the ybyb genotype produces

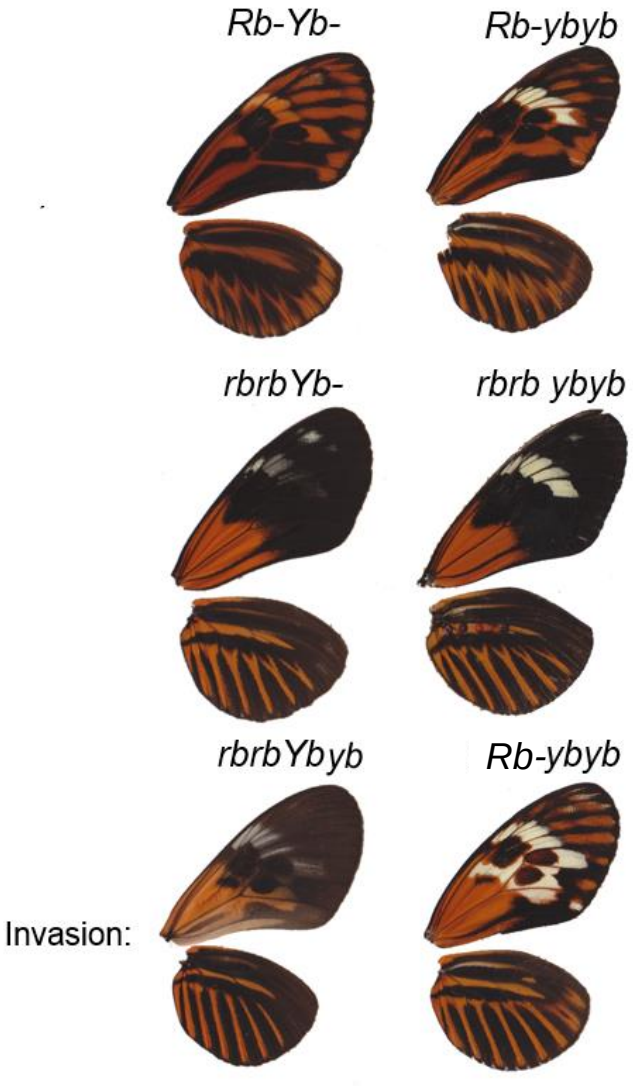


Figure 8. Different phenotypes of Yb and white invasion on different Rb backgrounds.

an intense, uniformly white band and white scales on the subcostal hindwing vein, whereas in *H. pardalinus* (YbYb) the white band is faded and no white scales are seen on the hindwing (*Fig. 8*). The F1 individuals all display the *pardalinus* phenotype with a faded band and no white scales on the vein, suggesting complete dominance of the *H. pardalinus* allele and making Yb a Mendelian locus. In the F2 hybrids, phenotypes resembling the heterozygous *H. melpomene* Ybyb phenotype are occasionally observed: the heterozygous phenotype is comprised of both white scales and shaded scales with a different reflectance from the rest of the wing, better visible on the ventral surface. but this phenotype can only be observed on a black background, a condition that is often not met in the F2 and is never met in the F1. Due to this, the heterozygous individuals had to be scored alongside the homozygous *pardalinus* individuals, assuming complete dominance. Thus, the expected proportion of Yb-:ybyb phenotypes in the F2 broods is 3:1. The observed frequencies of 301:114 are consistent with these expectations (Chi-square, X^2=1.23, df=1, p=0.27).

White invasion. Occasionally, white scales are observed outside of the region normally occupied by the white band, in the central area of the wing (*Fig. 8*). This phenotype was nicknamed “white invasion” in reference to the fact that white scales are never observed in this area in either parental species. Its own phenotype appears to depend on that of the white band: white invasion forms is especially evident when it appears alongside the ybyb phenotype as it forms intense white spots, yet it is not restricted to ybyb individuals. Occasionally it appears in individuals that sport a faded white band, wherein it takes on a different phenotype: a mix of white scales and lightly coloured scales with different reflectance from the rest of the black background. In the F2 broods the phenotype frequencies of this trait are 321:93 (absent:present) consistent with the expected 3:1 ratio (Chi-square, X^2=1.49, df=1, p=0.22). The extent to which white invasion affects the central region of the wing varies, but it always follows the contours of spots that are ordinarily present there, such as the dumbbells.

Apical band. The apical region of the *H. pardalinus* forewing sports a second white band that bisects the forewing rays and is comprised of three spots. The F1 phenotype is variable: in individuals where the apical portion of the forewing was relatively undamaged at least one spot of the apical band could always be observed, in both the York and the Peru broods. In the F2 however the band does not behave like a mendelian locus: while the York F2s show at least one spot of the band with a frequency of 88:38 (present:absent) which is not significantly different from Mendelian predictions (Chi-square, X^2=1.52, df=1, p=0.21) in the Peru F2 it appears less often

than expected, with a frequency of 148:91 of present:absent (Chi-square, X^2=21.11, df=1, p<0.0001). Scoring incomplete apical bands in the F2 as heterozygous gives frequencies of 38:58:30 (absent:incomplete:present) for the York broods and 91:11:138 for the Peru ones. While the formers are not significantly different from the frequencies expected from a heterozygous trait (Chi-square, X^2=1.81, df=2, p=0.4) the latters are still significantly different (Chi-square, X^2=35.85, df=2, p<0.0001).

Table 1. Chi-squared tests for Mendelian frequencies of different putative loci, assuming complete or incomplete dominance. In the backcrosses, the expected ratios refer to heterozygous:*pardalinus* phenotype, as pure *elevatus* phenotypes are expected to be completely absent.

York

	Phenotype frequencies				Chi-square test		
Trait	***elev***	**het.**	***pard***	**exp. ratio**	**χ2**	**p-value**	**df**
dennis	32	76	40	1:2:1	0.97	0.62	2
hindwing bar	37	82	29	1:2:1	2.59	0.27	2
rays	37	77	34	1:2:1	0.36	0.83	2
forewing rays	26	82	33	1:2:1	26.24	<0.0001	2
dumbbells	46		102	1:3	3.28	0.19	1
orange band	26		120	1:3	3.65	0.06	1
white band	39		109	1:3	0.08	0.77	1
invasion	27		121	1:3	3.25	0.07	1
apical band	38		88	1:3	1.81	0.4	1

Peru

	Phenotype frequencies				Chi-square test		
Trait	***elev***	**het.**	***pard***	**exp. ratio**	**χ2**	**p-value**	**df**
dennis	59	132	77	1:2:1	2.48	0.3	2
hindwing bar	74	136	59	1:2:1	1.71	0.42	2
rays	57	155	57	1:2:1	6.25	<0.05	2
forewing rays	66	157	46	1:2:1	1.84	0.4	2
dumbbells	66		198	1:3	10.5	<0.05	1
orange band	66		201	1:3	0.01	0.99	1
white band	75		192	1:3	1.2	0.27	1
invasion	66		200	1:3	0	1	1
apical band	91		148	1:3	35.85	<0.0001	1

Joint datasets

	Phenotype frequencies				Chi-square test		
Trait	***elev***	**het.**	***pard***	**exp. ratio**	**χ2**	**p-value**	**df**
dennis	91	208	117	1:2:1	2.17	0.34	2
hindwing bar	111	218	88	1:2:1	3.4	0.18	2
rays	94	232	91	1:2:1	5.34	0.07	2
forewing rays	92	239	79	1:2:1	10.7	<0.001	2
dumbbells	112		300	1:3	0.93	0.33	1
orange band	92		321	1:3	1.49	0.22	1
white band	114		301	1:3	1.23	0.27	1
invasion	93		321	1:3	1.49	0.22	1
apical band	129		236	1:3	20.28	<0.0001	1

Backcross to *H. pardalinus*

	Phenotype frequencies				Chi-square test		
Trait	***elev***	**het.**	***pard***	**exp. ratio**	**χ2**	**p-value**	**df**
dennis		14	19	1:1	0.48	0.49	2
hindwing bar		15	18	1:1	0.12	0.73	2
rays		14	19	1:1	0.12	0.73	2
forewing rays		12	11	1:1	0	0.99	2
dumbbells		20	14	1:1	0.74	0.39	1
orange band			33	1	0	1	1
white band			33	1	0	1	1
invasion			34	1	0.48	0.48	1
apical band		5	18	1:1	6.26	<0.05	1

3.2 Inferred linkage between traits.

While a linkage map could only be attempted for the dennis, bar and rays due to them having the clearest heterozygous phenotype, it is possible to infer linkage between traits or lack thereof by checking for 9:3:3:1 ratios between different phenotypes: such ratios indicate that two loci are unlinked. This was attempted between the traits with potentially Mendelian frequencies: forewing rays, dumbbells, orange band, white band, white invasion and apical band (albeit the apical band only behaves like a simple locus in the York F2). This analysis focused on the phenotypes, so each trait was considered as the result of a single locus with complete dominance, with partial phenotypes and complete *pardalinus* phenotypes scored together. The results are shown in Table 2. This analysis revealed that the orange band and the forewing rays are potentially given by the same locus (as there are no recombinant phenotypes of the two traits), and it suggested the possibility of linkage between the dumbbells and the forewing rays/orange band, and between the white band and the apical band. Discrepancies were observed between the frequencies of recombinants with the apical band between the York and Peru broods, due to the band's non Mendelian frequencies. However even in the Peru broods the frequencies of parental phenotypes between the white band and the apical band were higher than the frequencies of recombinants. In the Peru broods the apical band segregates with the forewing rays/orange band and with the dumbbells with patterns significantly different from 9:3:3:1 but this is probably due not to actual linkage but to the odd frequencies of the apical band phenotype in these F2s. White invasion appears linked to both white and orange pattern elements, but this is likely to be because white invasion only happens within the boundaries of regular orange elements. This is further explained in the Discussion.

Table 2. Chi-square tests for linkage based on phenotype frequencies. All combinations with the apical band (A) are calculated only with the York frequencies. In=white invasion. If p<0.05, traits may be linked.

Loci	Phenotype combinations	Exp. Ratio	Parental	Recomb.	Recomb.	Parental	χ^2	p-value	df
Fr/Dbs	FrDb : Frdb : frDb : frdb	9:3:3:1	257	57	40	51	51.54	**<0.0001**	3
Fr/Ob	FrOb : Frob : frOb : frob	9:3:3:1	313	0	0	92	359.42	**<0.0001**	3
Fr/Yb	FrYb : Fryb : frYb : fryb	9:3:3:1	235	64	70	35	9.25	**<0.05**	3
Fr/A	FrA : Fra : frA : fra	9:3:3:1	73	14	28	9	5	0.17	3
Dbs/Ob	DbOb : Dbob : dbOb: dbob	9:3:3:1	263	58	41	52	52.52	**<0.0001**	3
Dbs/Yb	DbYb : Dbyb : dbYb : dbyb	9:3:3:1	219	81	85	29	2.05	0.56	3
Dbs/A	DbA : Dba : dbA : dba	9:3:3:1	65	22	22	15	7.24	0.06	3
Ob/Yb	ObYb : Obyb : obYb : obyb	9:3:3:1	242	70	66	35	6.08	0.11	3
Ob/A	ObA : Oba : obA : oba	9:3:3:1	73	14	28	9	5	0.17	3
Yb/A	YbA : Yba : ybA : yba	9:3:3:1	75	13	16	21	29.58	**<0.0001**	3
In/Ob	InOb : Inob : inOb : inob	9:3:3:1	262	51	50	54	52.77	**<0.0001**	3
In/Dbs	InDb : Indb : inDb : indb	9:3:3:1	222	81	99	35	8.03	**<0.05**	3
In/Yb	InYb : Inyb : inYb : inyb	9:3:3:1	191	45	97	52	54.22	**<0.0001**	3
In/Ab	InAb : Inab : inAb : inab	9:3:3:1	98	23	11	27	43.54	**<0.0001**	3

3.3 Quantitative variation.

Variations in the shape of various spots suggests that some QTLs are affecting the formation of the pattern, and that their action mostly influences the shape and size of the black spots. Most F1 and F2 individuals show some degree of quantitative variation in the shape of the melanic spots. The most strongly affected regions are the central and apical portion of the forewing (where the dumbbells and the forewing rays are usually found, as well as part of the white band's final cell), the lower portion of the dennis, and the region surrounding the hindwing bar (*Fig. 9*).

Figure 9. Various degrees of black expansion in F2 individuals. Notice how it can affect the discal area and the apical area independently. It is also shown that extensive amounts of black expansion on the forewing do not necessarily affect the hindwing as well.

4. Discussion

4.1 Multilocus architecture of the colour pattern

Phenotype analysis showed that *H. elevatus* and *H. pardalinus* have evolved a multilocus system for colour pattern generation. Heterozygous phenotypes and parental phenotypes detected in the F2 broods are slightly more variable in shape than homozygous phenotypes seen in the pure species, a phenomenon also observed by Mallet in *H. erato* and *H. melpomene* (Mallet, 1989). The amount of elements of the pattern whose appearance varies among F2 individuals may not correspond to the actual number of genes controlling pattern formation, but while this number currently remains unclear, the complete absence of linkage between the white elements and the orange elements of the pattern suggests the presence of more than one locus. In light of this it is possible to make an attempt at defining the two pure species' genotype: for *pardalinus,* [$D^{pard}D^{pard}R^{pard}R^{pard}$*FrFrDbDb*][*YbYbAA*], for *elevatus* [$D^{elev}D^{elev}R^{elev}R^{elev}$*frfrdbdb*][*ybybaa*], with the orange elements and the white elements apparently falling into two separate linkage groups. Moreover, the presence of QTLs that control the size and presence of different melanic spots was inferred.

Linkage between *Yb* and *A*. The *Yb* locus and the *A* locus segregate with a ratio that is significantly different from the expected 9:3:3:1 ratio for unlinked loci. The parental phenotypes consist in faded band/present apical band (*Yb-/A-*, typical of *pardalinus*) and full band/absent apical band (*ybyb/aa,* typical of *elevatus*) and they both appear more often than the recombinant phenotypes (*Yb-/aa* and *ybyb/A-*) in the F2, with frequencies of 75:13:16:21 against the expected ~70:23:23:8. This strongly suggests linkage between these traits. The fact that they are both associated with the presence of white scales suggests homology with *H. melpomene'*s *N-Yb-Sb* linkage group that controls the appearance of the majority of white elements on its colour pattern and it is located on chromosome 15 (Papa et al., 2008; Jiggins et al., 2005). With the frequency of recombinants being 29/125 (23.2%), it is tempting to say that these two loci may lie at 23.2cM from one another. However the impossibility to score the entire F2 brood and to distinguish heterozygous individuals means that the estimated number of recombinants is necessarily imprecise. It is clear they are not the same locus, which is consistent with *melpomene'*s linkage group, whose loci are distinct, even though their relative distances are much shorter than those one could infer from these crosses (Naisbit et al., 2003).

It may appear counter-intuitive that the Yb locus and the orange band locus, which control the colour of exactly the same area of the wing, are actually completely unlinked. However they clearly produce four different phenotypes: orange band with faded white band (*Fr-Yb-, pardalinus* phenotype), orange band with full white band (*Fr-ybyb,* recombinant), no orange band with faded band (*frfrYb-,* recombinant) and no orange band with full white band (*frfrybyb, elevatus* phenotype). This is not surprising: a similar spread of phenotypes is observed in crosses between different non-rayed races of *H. erato,* where they are given by the combined action of two unlinked loci with incomplete dominance- *Cr* (for the yellow band) and *D* (for the red forewing band) (Mallet, 1989; Papa et al., 2008). It is also seen in crosses between rayed races of *H. melpomene,* where it is given by the interaction between *Yb* and *B.*

Linkage between orange elements. Both Yb and A segregate with ratios similar to 9:3:3:1 with all the orange elements of the pattern that appear controlled by Mendelian loci. Linkage is observed however between different orange spots. This implies that the orange elements form a separate linkage group from the white ones. The dennis, rays and hindwing bar all appear linked, as the recombination frequencies are lower than 50%. The calculated distances of 25.7cM between dennis and bar, 31.8cM between dennis and rays, and 27.7cM between bar and rays however appear very large compared to *H. melpomene*'s (Papa et al., 2008), and the relative distances are inconsistent. This is most likely due to errors in scoring the heterozygous phenotypes due to interference from the QTLs associated with black spots. The presence of recombinant individuals in the backcross is a clear sign of phenotyping errors. While pure *elevatus* phenotypes are very distinctive, distinguishing the heterozygous phenotype from the *pardalinus* one remains tricky (*Fig. 10*). With the knowledge that dennis and hindwing bar are given by extremely tightly linked genes in *H. melpomene* (Papa et al., 2008), one may assume that in truth the bar and the dennis are the same locus. This would solve both the inconsistency between distances and the problem of recombinants in the backcross: any phenotype scored as recombinant between dennis and hindwing bar could simply be given by quantitative variation affecting the shape of the bar. This hypothesis would be supported by the fact that in *H. hecale* and *H. ismenius* the entirety of the orange pattern is controlled by a single locus, *Br*, found on chromosome 18 and homologous to *H. melpomene*'s B/D locus (Huber et al., 2015).

Figure 10. A variety of phenotypes found in the backcrosses to *H. pardalinus*. Notice variations in the shape of the hindwing rays.

The other orange spots also appear linked to each other when they are scored as Mendelian loci. The orange band and forewing rays, originally scored as two separate elements, are instead given by a single locus hereby referred to as Fr, which shows evident linkage with the dumbbells locus. It is possible that the two dumbbells, the forewing rays and the orange band are actually all part of the same locus, with the recombinant phenotypes that lack the dumbbells being the result of epistatic interactions with the black QTLs. Based on this it would appear that the *H. pardalinus* and *H. elevatus* patterns are controlled by 5-6 loci (depending on whether the forewing rays and the dumbbells are assumed to be the same locus): dennis/band and rays (D and R), forewing rays/orange band (Fr), dumbbells (Db), white band (Yb) and apical band (Ab).

4.2 Quantitative variation in the black spots.

According to Nijhout's groundplan the majority of basic black spots that make up the *Heliconius* pattern are located in the apical and central region of the forewing and in the upper portion of the hindwing (Nijhout, 1990). Papa et al. (2013) verified the presence of several QTL-associated melanic spots in these regions of the *H. erato* wings. Thus, the two dumbbells and the apical forewing rays, which are located in these two regions, are subject to the most disturbance. While all F1 individuals show incomplete forewing rays and incomplete dumbbells, it is likely that these loci are in fact Mendelian in nature, with complete dominance of the *pardalinus* allele, and the incomplete phenotype associated with heterozygous individuals is instead dependent on the action of separate QTLs, unlinked to the orange group and responsible for the expansion of black elements. This possibility seems supported by the frequencies of complete:incomplete:absent dumbbells in the F2s, which are significantly different from the expected 1:2:1 ratios of loci with incomplete dominance. It is also potentially supported by the appearance of the Fr locus: the

orange band and the forewing rays never appear independently from one another- there are no individuals that sport the orange band and no forewing rays or vice versa. Yet the orange band never shows intermediate phenotypes, only the rays do. This is an example of how quantitative variation in the black spots may make the phenotype and its inferred genetic structure appear more complex than it actually is. In addition, the possible epistatic interactions between black elements and orange spots means that there may be instances where the alleles responsible for presence of the dumbbells are expressed in the genotype, but have no visible effect on the phenotype. This would be an extreme example of black expansion.

Notably some *H. melpomene* races sport a spot, called the cell spot, that is located in the same region as *H. pardalinus*'s upper dumbbell and has a similar shape (Sheppard et al., 1985). In *H. melpomene,* depending on the race, the cell spot can appear whole, or just as its upper or lower half. Suppression of this spot is given by QTLs and the amount of it that is expressed in the phenotype is also given by different combinations of suppressor alleles that vary with the population, with the entire dumbbell only appearing in the absence of all suppressors (Sheppard et al., 1985).

The presence of such high levels of disturbance from black expansion on the forewing is likely to be the result of the very different patterns of melanization seen in *H. pardalinus* and *H. elevatus.* The latter show extensive amounts of black pigment in the central and apical portion of the forewing, with only the dennis area and the white band being spared. The former instead shows enlarged but distinct black spots in the central region, in the discal cell, and in the lower portion of the forewing- the discal spot's elongated shape invades the dennis area. Less variation is observed in the hindwing's pattern due to the relatively smaller number of differences between the two species, but this is still enough to make scoring of heterozygous phenotypes less straightforward. Presumably the individuals that sport the maximum amounts of black expansion had all their black elements enlarged, both those typical of *H. elevatus* and those typical of *H. pardalinus.* This phenomenon is observed in other silvaniform species as well. When crossed, different races of *H. hecale,* in particular *H. h. melicerta* and *H. h. clearei*, that also have very different black patterns, produce offspring that display continuous variation in the levels of melanization of their fore- and hindwings, with the parental phenotypes representing the two extremes of the spectrum (Huber et al., 2015). Even *H. numata*'s patterns, which are given by a single gene, are nonetheless subject

to varying degrees of quantitative variation given by small effect QTLs controlling the black spots (Joron et al., 2006).

4.3 Different developmental interpretations of the loci can affect phenotype scoring.

From a developmental perspective, the interpretation of these loci's action changes depending on the way the phenotype is interpreted. There are two options: either the phenotype is seen as the formation of coloured elements on a basic melanic background, or the basic background consists in the coloured elements themselves and the melanic spots are the variable pattern. While this distinction does not affect the process of phenotype scoring, it provides alternative explanations for the observed patterns. As an example, in postman races of *H. melpomene* and *H. cydno,* the alleles that produce the most melanic phenotypes and those that in general involve a reduction in size of the coloured or white elements often appear dominant over those that promote an expansion of the coloured elements (Nijhout, 1994). In *H. hecale* crosses this also seems to be the case: the K locus that produces the white band is Mendelian and the recessive phenotype involves production of a wide, full white band, which is instead broken into smaller white regions in the dominant phenotype, due to expansion of black elements (Huber et al., 2015). The dominant form of an allele is typically (albeit not always) the one that represents a gain of function, which means that the dominant black expansion validates the hypothesis that the black pattern is laid onto an orange background rather than vice versa. However since the dennis is a dominant trait in rayed races of *H. melpomene* (Turner, 1962), this may not always be the case.

Understanding whether black expansion represents a gain of function or a loss of function is important, albeit impossible to determine from phenotype observation alone. If expansion of the melanic spots is in fact a gained function, then the incomplete orange pattern elements should be counted among the phenotypes where the orange elements are missing altogether (Sheppard et al., 1985), which would create different ratios. The explanation is likely to be more complex: observation of the *pardalinus-elevatus* hybrids suggests that there are actually two different sets of loci, one for the orange pattern elements and one for the QTLs controlling melanization. This is impossible to determine with certainty with no genomic data, but it is plausible because *H. hecale* and *H. ismenius* have two separate linkage groups, one for the orange/red elements and one for the melanic ones (Huber et al., 2015).

4.4 *H. elevatus* and *H. pardalinus*'s colour patterns share their genetic architecture with other silvaniforms.

This interpretation places the genetic architecture of *H. elevatus* and *H. pardalinus'*s colour pattern closer to that observed in *H. hecale* and *H. ismenius* than to the polymorphic single-gene system observed in *H. numata.* Huber et al. (2015) suggested that the single-locus architecture may be unique to *H. numata* among the silvaniforms, and that it may have evolved from the multilocus system seen in *H. hecale* and *H. ismenius,* a concept that had also been suggested by Jones et al. (2012). *H. numata'*s *P* locus is located on chromosome 15 (LG15), homologous to *H. melpomene* and *H. cydno*'s *N-Yb-Sb* group, and to *H. erato'*s *Cr* locus that controls the appearence of the yellow hindwing bar (Huber et al., 2015, Mallet 1989). In *H. hecale* and *H. ismenius* the locus on LG15 is called *N* and it controls the shape of the white band in *ismenius* and the presence of elements similar to *pardalinus*'s apical band in *hecale* (Huber et al., 2015). In *H. pardalinus* and *H. elevatus* the *Yb* and *a* loci are likely to be located in this linkage group.

Regardless of how many of the orange elements that were scored as separate loci are actually part of the same large effect locus, the presence of an orange linkage group is also consistent with the *H. hecale, H. ismenius* and *H. melpomene* architecture: in all these three species the orange/red portion of the pattern is controlled by a linkage group located on chromosome 18 (LG18) near the *optix* gene. LG18 consists in *D/B* in *melpomene* and in *Br* in the other two. It is likely that the *pardalinus* and *elevatus* orange LG corresponds to LG18, with *DBR/Fr/Db* being homologous to *D/B* and *Br.*

In *H. melpomene, ismenius* and *hecale,* as well as in *H. erato* and *H. cydno,* loci that control the shape of coloured pattern elements via the expansion or contraction of melanic spots are located on chromosome 10 (LG10) near the developmental gene *WntA*. These loci are all known as *Ac* save for *H. erato'*s, whose main QTL is *Sd,* that acts on melanic spots in the central region of the wing. *Sd*'s action is accompanied by *Ro'*s, a QTL for melanic spots in the apical region (Papa et al., 2013). In F2 hybrids between *H. elevatus* and *H. pardalinus* there appear to be two independent regions of black expansion on the forewing, in the central and apical region, much like in *H. erato*, so a similar organization is likely.

4.5 The white invasion dilemma: a re-activated *H. elevatus* allele?

White invasion, scored as the presence of white scales outside of the white band region regardless of the extent of its manifestation, is a phenotype that is unique to hybrids. The phenotype it adopts appears dependent on the Yb phenotype: in *ybyb* individuals the spots formed by white invasion are fully white, while in *Yb*- individuals even the spots are faded, much like the white band itself. It is occasionally possible to observe the contours of white invasion as scales with a different reflectance on a black background (*Fig. 8*), occasionally accompanied by sparse white scales- as stated before, this phenotype is thought to be typical of *Ybyb* individuals. In *H. melpomene* and *H. erato,* the yellow bar given by *Yb* takes on this phenotype in heterozygous *Ybyb* individuals, is complete in *ybyb* individuals and absent in *YbYb* (Mallet, 1989; Naisbit, 2003). Due to the appearance of white invasion depending on *Yb,* it is possible that the existence of such phenotypes is proof that *Yb* truly has incomplete dominance- it just cannot be detected in most cases.

While white invasion is apparently linked to orange elements of the pattern as well as the white band, it cannot be genetically linked to both. The apparent linkage between white invasion and the dumbbells is a product of the fact that white invasion always occurs within the shapes of pattern elements that would ordinarily be present on the wing. Thus, the presence of white dumbbells as a result of white invasion, even with no orange background, was scored as presence of the dumbbells rather than absence. If such spots are not expressed, white invasion is not observable. This suggests that white invasion does not have epistatic interactions with the black QTLs in the same way as other pattern elements do. This characteristic is useful because it may reveal whether the two dumbbells are expressed even in a background of full black expansion, and white expansion in the dumbbells was in fact interpreted as such. In an individual with black in its central forewing region, if white scales are present in the areas normally occupied by the dumbbells, it might mean that *Db* is being expressed but the excess of black scales is masking their normal effects.

White invasion is never observed in the F1 or in the backcrosses to *H. pardalinus*, meaning it must be an allele that originated in *H. elevatus* (*Fig. 11*). While the parental population of *H. elevatus* used for these crosses only has a simple white band, there are other races wherein the white elements have a similar configuration (Turner, 1966). For example, Guianese races of *H. elevatus* such as *H. elevatus bari* have several white spots in the central region of the wing, and such

phenotypes are also seen in *H. erato* (Papa et al., 2013) and in *H. melpomene* (Sheppard et al., 1985). The parental population of *H. elevatus* in these crosses sports a small white dot in the tornal region of the wing that may itself be a very restricted form of white expansion. It is possible that the phenomenon of white invasion is due to an ancestral gene that originally introgressed from *H. melpomene* into *H. elevatus* and was subsequently silenced in some *H. elevatus* races. If this is the case, its expression in F2s means that hybridization is removing the silencing mechanism.

Figure 11. Comparison between white invasion in an F2 hybrid (left) and H. elevatus bari (right). Picture of H. elevatus bari from *http://www.heliconius.net/.*

5. Conclusion.

The traits and associations characterized in this study seem to validate the hypothesis that *H. elevatus* may have formed via gene flow from *H. melpomene* into the silvaniform clade, as it seems to carry several alleles that are also present in the *melpomene/cydno* clade, and both *H. elevatus* and *H. pardalinus* appear to have linkage groups homologous not only to *melpomene'*s LG10, LG15 and LG18, but also to their close relatives *H. hecale* and *H. ismenius.* By contrast, this study confirmed that *H. numata*'s supergene is an acquired trait and that ancestral silvaniform *Heliconius* probably had a multilocus architecture.

Mate choice experiments (Queste, 2015) revealed that *H. elevatus* shows preference for individuals sharing its same colour pattern and discriminates against the *H. pardalinus butleri* pattern. Merrill (2011) showed that assortative mating in *H. melpomene* is a behavioural trait that can be mapped onto the genome and appears associated with the *B/D* locus, which means that *H.*

elevatus may have inherited this behaviour alongside the *melpomene* colour pattern loci. This would make the colour pattern in *H. elevatus* and *H. pardalinus* a magic trait: ecologically divergent, under strong disruptive selection, and at the same time associated with mate choice. There is no doubt that the colour pattern and the introgression event in *H. elevatus*'s evolutionary history have played a large role in the separation between these species. QTL mapping of the identified loci will reveal their location in the genome, confirming whether or not the two studied species' colour pattern loci are distributed in the same linkage groups as *H. melpomene*'s. In particular, QTL mapping of the melanic spots is expected to reveal important information about their organization (whether it is, in fact, similar to *H. erato*), their relationship with *H. melpomene'*s *Ac* locus and their interactions with other loci such as B/D and Yb, which could not be inferred from phenotype analysis.

Chapter 3

Characterizing the pheromone blend of *H. pardalinus* and *H. elevatus.*

1. Introduction

1.1 Insect pheromones: important means of communication

Pheromones are one of the most important means of conspecific and heterospecific communication in the living world, employed by bacteria, plants and animals alike to convey information to other individuals (Tillman et al., 1999). Broadly speaking, pheromones are highly species-specific blends of chemicals (or single chemicals) that have the function of altering another individual's behaviour, or in some cases, its physiology (Symonds & Elgar, 2008). Pheromone components are volatile molecules that are picked up by olfactory receptors and affect the recipient's nervous system: their range of action depends on factors such as the pheromone's chemical composition and volatility, the rate at which they diffuse, the mode of emission, the target individual's receptivity and environmental factors such as air currents. The immediacy of the behavioural response varies, with some taking effect instantly and others priming the recipient for a long-term behavioural change (Regnier & Law, 1968). Insects are among the best studied organisms in the field of chemical ecology due to their pheromone blends being particularly varied both in composition and in purpose. For insects, pheromones represent a method to understand the environment and to receive information about other individuals in the area, as a chemical signal may contain information about, for instance, sex, species, distance, mating status, age or relatedness of the emitter in relation to the recipient (Howard & Blomquist, 2005). The functions of insect pheromones as signals thus range from dispersal to aggregation and recruitment, from alarm to reproduction, and different species employ them for different tasks and to varying degrees (Howard & Blomquist, 2005).

These functions are especially evident in social insects, where communication and coordination are vital for the colony's success. Ants for example produce a range of pheromones that include

alarm and recruitment signals (Bossert & Wilson, 1963). A classic example of the latter are pheromone trails, commonly used by worker ants to communicate the location of a foraging spot to their nest mates. For instance, *Solenopsis saevissima* workers release a continuous recruitment signal in the air around them as they travel from their colony to a food source, and this prompts other workers to follow the trail, and as the followers release the same continuous emission, the signal is reinforced for every ant that follows the same route (Bossert & Wilson, 1963). There are species of ants whose trails can persist for several days (Bossert & Wilson, 1963). Ant alarm pheromones on the other hand are emitted in brief puffs and they have an alerting effect. These signals are often emitted as a result of injuries or death. At low concentrations, alarm pheromones act as attractants to other workers, but at high concentrations they become repellent (Moser et al., 1968). In honeybees (*Apis mellifera*), alarm pheromones are released by workers upon stinging a target, and signal to other bees to attack the same spot (Regnier & Law, 1968).

While the alarm and recruitment pheromones have been mainly studied in eusocial insects, the most well-known type of pheromones, the mating signals, as well as the aggregation signals, have been thoroughly studied on a variety of insect taxa, with particular focus on Blattodea, Diptera, Coleoptera and especially Lepidoptera (Tillman et al., 1999). Compared to other insect taxa, Lepidopterans are particularly well-studied as sex pheromones were originally characterized in moths due to many species of caterpillar being economically important agricultural pests (Symonds & Elgar, 2008).

1.2 Lepidopteran pheromone biosynthesis: a general look

Lepidopterans have been crucial in the study of pheromone chemical composition and in research concerning the biosynthetic pathways that lead to pheromone production. In fact, the first pheromone component ever identified was an alcohol isolated in 1959 from the secretions of the silkworm *Bombyx mori* (a commercially important species), subsequently named bombykol (Butenandt et al., 1961; Ando & Yamakawa, 2011). Since then, the blends of hundreds of species of moths have been characterized in-depth. In moths, female individuals produce long-distance sex pheromones that are picked up by the male antennae, providing potential mates with accurate information as to the female's position. As most moths are nocturnal, this represents their main method of mate localization- this is the reason why male moths sport very elaborate antennae, able to derive subtle information from a female's chemical signal (Roelofs et al., 2002)

Across all insect orders and regardless of the type of signal, there are two possible starting points to pheromone biosynthesis: *de-novo* synthesis and sequestration of precursors from the larval host plant. In the latter's case, the host plant precursor can be transformed into a behaviourally active molecule by either a few simple chemical reactions or more complex novel biosynthetic mechanisms evolved for that specific purpose (Tillman et al., 1999). These two starting points do not exclude each other and the same species can often make use of both. While pheromone blends can be very variable, this does not always imply a large number and a wide range of genes involved in their production (Symonds & Elgar, 2008). Many moths for example use simple hydrocarbons as their chemical signals, including simple and methylated alkanes or alkenes (Howard & Blomquist, 1982), which require the action of few families of enzymes, mainly desaturases, reductases and oxidases that act on fatty acid precursors (Symonds & Elgar, 2008). These hydrocarbons serve as the base of several derivatives including esters, alcohols and aldehydes (Roelofs & Rooney, 2003). Pheromones that consist of saturated molecules 10-18 carbons in length (including any derivative thereof with an oxidized functional group, such as aldehydes, alcohols and esters) are known as Type I pheromones and are most common in moths, while unsaturated molecules usually 17-23 carbons in length are Type II pheromones and they are rarer (Matsumoto, 2010). A single species' blend can include both types. Most important to the production of pheromones are the fatty acyl-CoA reductases, also known as fatty acid reductases (FARs), which replace the -COOH functional group with an -OH group, and the fatty acyl-CoA desaturases (also known as fatty acid desaturases or FADs) which carry out one of the most important steps: the transformation of a saturated chain into an unsaturated one via introduction of a double bond (Matsumoto, 2010) (*Fig. 12*).

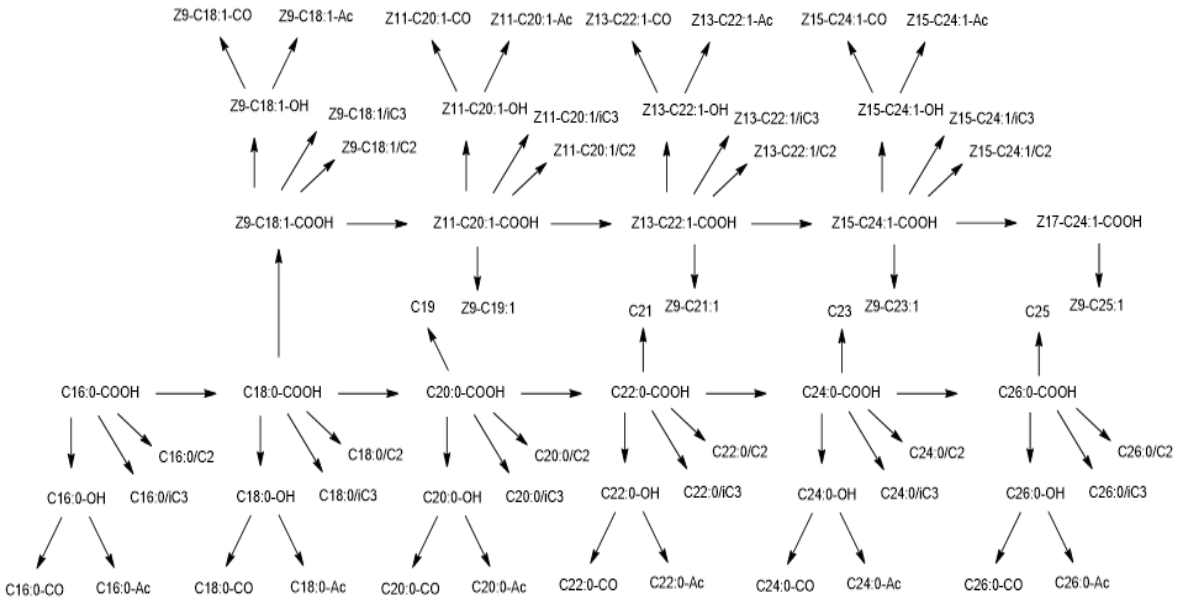


Figure 12. Fatty acid metabolism (Mann et al., unpublished). C(n) indicates the number of carbons in the main chain, Z(n) indicates the position of a double bond, if present (the number of double bonds is given after the number of carbons: C(n):0 for no double bonds, C(n):1 for one double bond). –COOH denotes fatty acids. Notice the desaturation event occurs early on, as C18:0-COOH (stearic acid) is transformed into Z9-C18:1-COOH (oleic acid). Other relevant compound classes represented here include alcohols (-OH), acetates (-Ac) and aldehydes (-CO). C(n) on its own represents alkanes, while Z(m)-C(n):1 represents alkenes.

For example, the red-banded leafroller moth *Argyrotaenia velutinana* has a sex pheromone blend mostly based on esters, with (Z)-11-tetradecenyl acetate and (E)-11-tetradecenyl acetate being the two main components (Bjostad & Roelofs, 1981; Roelofs & Rooney, 2003). These two esters are almost identical, sporting 14 carbons and a double bond in position 11 from the functional group, with the only difference being the orientation of the CH_2 groups adjacent to the double bond. In fact they are both derivatives of tetradecanoic acid (myristic acid): this precursor first undergoes desaturation, carried out by Δ11-desaturase which introduces the double bond; then, reduction, turning the fatty acid into an alcohol; lastly, acetylation, wherein the acetate group is added (Bjostad & Roelofs, 1981). Similarly, *Trichoplusia ni,* the cabbage looper moth, uses (Z)-7-dodecenyl acetate, which derives from hexadecanoic (palmitic) acid via Δ11-desaturation, followed by chain shortening (which turns the 16-carbon chain of palmitic acid into a 12-carbon chain), reduction and acetylation (Bjostad & Roelofs, 1983).

Unsaturated chains are common in nature and Δ11-desaturation is far from the only mechanism by which they can be formed. Several different families of desaturases exist, all of which specialize in introducing the double bond at different positions from the functional group (Roelofs & Rooney, 2003, Rodriguez et al., 2004). Δ9-desaturases are among the most common, being ubiquitous in animals as components of a cell's organelle membranes (Shanklin & Cahoon, 1998), and due to their fundamental role in metabolism, the double bond between C9 and C10 introduced by this family of enzymes is a very common structure in nature (Liénard et al., 2014).

Δ14, Δ5 and Δ7-desaturases also exist, and sometimes closely related species can differ dramatically in their active families of FA desaturases (Roelofs & Rooney, 2003). Evidence presented by Roelofs et al. (2002) showed that FADs evolved before the divergence between Lepidoptera and Diptera occurred, since both orders employ them for pheromone production, and different FAD families seem to have acquired this function independently from one another. In total, Roelofs lists the two kinds of Δ9 desaturases (the ones that favour 16C substrates and the ones that favour 18C ones), Δ10-desaturases/Δ11-desaturaes and Δ14-desaturases as the four main groups of FADs based on their genetic structure (Roelofs et al., 2002). Δ9 desaturases evolve slowly, presumably due to their ancestral metabolic function (Knipple et al., 2002); Δ10/11 desaturases evolve relatively faster and have completely lost metabolic function, being involved exclusively in sex-pheromone production; Δ14 desaturases, still poorly studied, evolve at the fastest rate (Roelofs et al., 2002). Interestingly, while it appears that the majority of characterized moth Δ9 FADs specialize on molecules within the 16C-18C range, a type of Δ9-desaturase was found in *Choristoneura parallela,* the spotted fireworm moth, that can act on any substrate between 14C and 26C in length (Liu et al., 2004). Its sequence does not cluster with the other two groups of Δ9-desaturases, suggesting it could probably be a cluster of its own (Liu et al., 2004).

Another crucial enzyme family in pheromone biosynthesis consists in fatty-acyl CoA reductases (FARs). The role of these enzymes is to transform fatty acids into alcohols, that can then be transformed into other classes of compounds such as aldehydes and esters, as seen in *Argyrotenia velutinana* and *Trichoplusia ni.* FARs are not very substrate-specific: usually, a single type of FAR can reduce a range of fatty-acyl precursors to alcohols, regardless of the presence of double bonds, with the only restriction coming from the length of the substrate molecule, which has to fall within the acceptable range for that FAR (Liénard et al., 2010). For instance, in the moth genus

Yponomeuta, pgFAR enzymes (pheromone gland FARs) accept substrates between 14 and 16 carbons in length. Thus the diversity in blends shown by different *Yponomeuta* species does not depend so much on the action of fatty acyl reductases as it does on the upstream biosynthetic machinery that determines the production of intermediates (Liénard et al., 2010). The fact that FARs do not specialize on a single substrate and that the end product depends largely on the available fatty acyl compounds is important, because it means that these enzymes can effectively facilitate the evolution of a diverse range of pheromone blends across different species. In fact, they will be able to carry out the reduction step regardless of the intermediates they are provided with (Hagström et al., 2012).

Like FADs, FARs are very diversified in insects, and as in FADs, changes in the genetic structure of FARs can cause significant modifications in the pheromone blend composition that can drive species divergence (Lassance et al., 2010). An example of this is seen in the corn borer moth *Ostrinia nubilalis,* which exists in two races: E and Z. Their blends are comprised of the same two compounds, (E)-11-tetradecenyl acetate and (Z)-11-tetradecenyl acetate, but the ratios are different between the two races. However, the unsaturated fatty acyl- intermediates to these two compounds are present in similar ratios in both races, which suggests that the important switch is occurring at the reduction stage (Lassance et al., 2010). In fact, a study on F1 hybrids between the two revealed that the change in ratios between the E and Z populations is due to a single allele of the gene *pgFAR* that encodes for the reductase (Lassance et al., 2010). The two alleles, $pgFAR^{E}$ and $pgFAR^{Z}$, are fixed in the E and Z population respectively, and the different results they produce in the pheromone blend are enough to cause reproductive isolation between the otherwise interfertile races (Lassance et al., 2010).

1.3 Pheromones in butterflies

Butterflies and moths are often culturally regarded as separate groups of Lepidopterans, but this is unlikely to be the case. In fact, butterflies (Rhopalocera) are currently believed to be phylogenetically nested within the moths (Heterocera), essentially having descended from moths ancestors and having acquired a number of traits that are unusual in other moth families, including bright colours, a diurnal lifestyle and enhanced visual perception (Liénard et al., 2014). While in moths females emit pheromones to attract males, in butterflies it is males. Due to butterflies relying heavily on their colour patterns to detect potential mates and conspecifics, it

seems reasonable to expect a loss in chemosensory genes compared to the moths that have to rely on chemical signals for those tasks. Yet certain butterfly genera (namely *Danaus* and *Heliconius*) actually show expansions in the chemosensory gene families (*Heliconius* Genome Consortium, 2012). Moreover, while it is true that butterflies appear to have lost the long-range signals observed in moths, male butterflies appear to still be able to convey large amounts of information to the females in the form of short range "scent-bouquets" (Liénard et al., 2014). One of the most important species in the study of butterfly pheromones, *Bicyclus anynana,* had its male sex pheromone blend characterized and the study in question revealed a composition not unlike that of female moths' sex pheromones, including (Z)-9-tetradecenol (Z9-14:OH) and hexadecanal (16:Ald), both derived from the metabolism of palmitic acid (C16) (Nieberding et al., 2008). The similarity of *B. anynana* males' blend with that of several female moths made it a good model for researching the evolutionary origin of butterfly biosynthetic pheromone production pathways in relation to moths. A study by Liénard et al. (2014) revealed that the production of Z9-14:OH is given by a three-step pathway of Δ11-desaturation of hexadecanoic acid into (Z)-11-hexadecenoic acid followed by β-oxidation into (Z)-9-tetradecenoic acid and finally reduction into (Z)-9-tetradecenol. 16:Ald is instead produced via a two-step pathway that begins with reduction of hexadecanoic acid to hexadecanol and then oxidation into hexadecanal. These two reactions involve a Δ11-desaturase and two pheromone gland FARs whose genes are orthologous to moth FADs and FARs (Liénard et al., 2014).

Butterflies rely on their vision to a much greater extent than moths do, which may be problematic in groups with mimicry. Several species of butterflies, even outside of the *Heliconius* genus, engage in Müllerian or Batesian mimicry such that multiple species converge on similar colour patterns. For example, African species of Danainae (the milkweed butterflies) form mimicry rings corresponding to several colour pattern models, meaning that several species cannot recognize their conspecifics visually. A study on 10 species of African Danainae butterflies revealed that males of each species have a bouquet made up of different classes of compounds (with complexity varying among species from 12 to 59 different molecules), including saturated and unsaturated hydrocarbons, aldehydes, alcohols, aromatic compounds and other plant derivatives (Schulz et al., 1993). While they share several common molecules, each species has at least one unique compound that distinguishes its blend from other species- in other words, an evolutionary autapomorphy: a feature that is only found in a specific taxon (Schulz et al., 1993). In such cases, it

is plausible and likely that chemical communication may become important in species recognition (Schulz et al., 1993).

1.4 *Heliconius* pheromones: a relatively unexplored topic

In spite of the importance of *Heliconius* butterflies as Lepidopteran models in the study of evolution, relatively little attention has been given to their pheromones. Most studies concerning *Heliconius* chemical communication have focused on two specific types of signals: chemicals produced by female pupae and antiaphrodisiacs (Schulz et al., 2008; Estrada et al., 2010; Estrada et al., 2011).

Heliconius males often engage in a unique behaviour known as pupal mating, wherein they form an association with a mature female chrysalis and remain to its side until it ecloses so as to secure mating with that specific individual (Estrada et al., 2010). A study on *H. charithonia* revealed that while male and female pupae produce the same cuticular hydrocarbons and the same volatiles, some of the plant derivatives (terpenoids) are sex specific (Estrada et al., 2010). Mature male pupae produce linalool, which repels males, while mature female pupae produce linalool oxide (found in immature female pupae as well, but only in trace amounts), which attracts them (Estrada et al., 2010). Antiaphrodisiacs on the other hand have a very different function: they are donated from the male to the female during mating and their function is to enforce monogamy and reduce male-male competition by signalling that the female is already mated, making her less attractive to other males that may wish to attempt courtship and reducing harassment (Schulz et al., 2008). Like other classes of pheromones in other organisms, the antiaphrodisiac bouquet's chemical composition is very diverse in terms of number of compounds and their concentrations, and different compounds and compound classes have been gained and lost several times during the *Heliconius* evolutionary radiation (Estrada et al., 2011). The effect of this is so strong that except in the case of recently diverged species pairs like *H. cydno* and *H. pachinus*, the antiaphrodisiac composition carries little phylogenetic signals (Estrada et al., 2011).

It is known that antiaphrodisiacs and sex pheromones released by the males during courtship are different in a number of ways. Antiaphrodisiacs are released from the male abdominal glands, while sex pheromones are released from a specific region of the male hindwing that is covered in live brush-like scales (androconia) which have been hypothesized to have a secretory function

based on their morphology, as opposed to the rest of the wing which is covered in dead scales (Vanjari et al., 2015). Antiaphrodisiacs are passed to the females by contact, while sex pheromones are likely dispersed during courtship: males will approach and hover the female before attempting to mate, and the hovering phase is probably the time male sex pheromones are released. In *H. melpomene,* antiaphrodisiacs and male sex pheromones have been characterized by Schulz et al. (2008) and Vanjari et al. (2015) respectively, and the two blends have different compositions. Similarly, in a study on *H. melpomene* and *H. timareta*'s sex pheromones (Meròt et al., 2015) it was shown that the androconia and the male abdominal tips produce different blends, with the abdominal tip's blend being more diverse and consisting in a concoction of several esters and fatty acids. Thus albeit the pheromones identified in *H. melpomene* and *H. timareta* are only putative, even in other *Heliconius* species, sex pheromones cannot be assumed to have the same chemical composition as antiaphrodisiacs.

In light of this, the courtship emissions of *Heliconius* require more attention, as they are likely to play an extremely important part in species divergence, due to their role in mate choice and identification of conspecifics. In the study of *Heliconius* evolution and speciation, it is a priority not only to characterize the blends' composition, but also to understand through what mechanisms different species have developed different blends. While the pheromone blend components may be expected to be quantitative rather than Mendelian in nature, the F2 hybrid phenotypes can still be used to deconstruct the parental species' blends, offering important insights as to the genetic organization of this complex trait. Using gas chromatography-mass spectrometry the pheromone composition of *H. pardalinus*, *H. elevatus* and their hybrids was characterized. In addition to this, a novel method involving correlation plots between compounds was employed in an attempt to determine the genetic structure underlying the production of different blend components using the phenotype of F2 hybrids.

2. Materials and methods

Androconia were extracted from the hindwings of 189 adult *Heliconius* males (~21 days old, as this is when they reach sexual maturity). These included 133 F2 hybrids, 19 F1 hybrids, 14 backcrosses to *H. pardalinus*, 10 *H. elevatus* and 13 *H. pardalinus.* Of these, Jake Morris (Ph.D) had previously analyzed the 17 F1 individuals, as well as 5 *elevatus* and 5 *pardalinus*. Among the 133 F2 hybrids

are 5 individuals that were fed exclusively on *Passiflora edulis* as caterpillars and 5 that were fed exclusively on *P. serratodigitata,* included in the study to check whether the butterflies pheromone blend was affected by larval food plant. Otherwise, larvae were reared on a mixture of *Passiflora* species; *P. caerulea* and *P. edulis* in York, and *P. edulis* and *P. serratodigitata* in Peru. Of the F2 individuals, 18 descended from F1 parents bred in captivity in York, and the other 115 were bred in Peru (the York and the Peru population they came from are the same mentioned in Chapter 2 concerning colour pattern analysis, though not all individuals used for colour pattern scoring were used for pheromone analysis).

All individuals were sacrificed and their androconia were excised from the hindwings and suspended in dichloromethane (CH_2Cl_2). Dichloromethane was chosen due to being a weakly polar molecule, making it the ideal solvent for large organic molecules such as pheromone components, which tend to be non-polar. Pheromone analysis was carried out via GC/MS in the Department of Organic Chemistry of Braunschweig Technische Universität.

The dichloromethane suspensions containing the androconia were separated from the wing fragments and their volume reduced with a nitrogen evaporator (this process only evaporates dichloromethane and the least heavy molecules- the compounds that make up the pheromone blend remain unchanged). 212.8 ng of tridecyl acetate was added to each suspension to serve as the internal standard. Tridecyl acetate was prepared via tridecane esterification in a dichloromethane medium.

The resulting solutions were run on a 7890A GC-System coupled with a MSD 5975C mass analyzer (Agilent Technologies, Santa Clara, USA) instrument fitted with a HP-5MS column (50 m, 0.25 mm i.d., 0.25 µm f.t.; Agilent Technologies). The ionization-method was electron impact with a collision-energy of 70 eV. Conditions were as follows: inlet pressure 9.79 psi, He 20 mL min^{-1}, injection volume 1 µL. The GC was programmed as follows: starting at 50°C increasing at 5°C min^{-1} to 320°C and hold that temperature for 5 min. The carrier gas was He at 1.2 mL min^{-1}. For all identified compounds the concentration was calculated from the peak's area, as reported by the AMDIS software. Each compound's chromatogram was interpreted by the software via the NIST databases and the additional databases compiled at the Institute of Organic Chemistry of Braunschweig TU by F. Mann (Ph.D), who had also originally identified most of the compounds. All

identifiable compounds between undecane and nonacosanal were scored, including any potential contaminants or extraneous compounds, which were eliminated at a later point alongside compounds that appeared very sporadically (less than 10 times across the entire dataset).

Pure species analysis. The 10 previously analyzed pure individuals (5 *H. elevatus* and 5 *H. pardalinus*) were pooled with the newly analyzed individuals, for a total of 10 *elevatus* and 13 *pardalinus.* The two datasets were pooled together as conservatively as possible, eliminating all compounds that were found in only one dataset and all compounds that appeared in fewer than four individuals. Forty compounds in total were retained. The logged concentrations (ln(conc. +1)) of all these compounds were analyzed using R(corrplot). The expectation for this investigation was that the correlated compounds should show two clusters, one for the *H. pardalinus* compounds and one for the *H. elevatus* ones, and the non species-specific compounds should either not fall into either cluster, or appear correlated to both. Kruskal-Wallis tests were run to check for differences between the compounds' median concentrations (not logged) in the two species. This test was favoured over a t-test due to the concentrations not being normally distributed. A principal component analysis was then performed on the compounds that were significantly different to check for consistency with past analyses.

Whole-dataset analysis. The entire dataset (including both the York and the Peru F2 broods, the backcrosses, the F1s and the pure species) was logged (ln +1 to control for the presence of zeros) and then normalized using the concentrations of a single compound that had appeared consistently across all samples and in past samples of the same kind. The amounts of this compound would presumably act as proxies for the general amount of blend produced by an individual, while being unaffected by any other factor. 11-Methylpentacosane was chosen as a normalising compound due to appearing in all but 3 individuals (which had to be eliminated from the analysis). This normalization method results in the chosen standard compound's concentrations always being reduced to 1, thus rendering the compound impossible to analyze, and a methylated alkane was considered an acceptable loss of information compared to the other possible candidates which were alkenes (Z-9-Henicosene and Z-9-Tricosene). This method was chosen over percentage contribution of each compound to the overall blend because transforming the concentrations into percentages causes the values of each compound to be affected by others;

this effect is particularly pronounced when the overall amount of compounds in the blend varies due to a few abundant compounds. This normalized dataset will be referred to as C25Me dataset.

A separate attempt at normalizing the dataset was made with the averaged concentrations of heptacosane and nonacosane. The reason for this choice was that due to their size and weight heptacosane and nonacosane are likely to be cuticular hydrocarbons, and were in fact documented as cuticular hydrocarbons in *Heliconius melpomene* the past (Vanjari et al., 2015). Due to this, they are unlikely to be involved in chemical signalling. The size range for cuticular hydrocarbons in *Heliconius* is thought to be C26-C30 but hexacosane and octacosane were not used as they were often absent from individuals. This dataset will be referred to as C27-29 dataset.

Statistical analysis of the results was carried out in R-Studio 0.99.491 using mainly the ggplot2, reshape2 and corrplot packages. The ggplot2 and reshape2 packages were used to produce histograms for the distribution of concentrations of every compound across all individuals. The reason why this was done was to show whether any compound showed a bimodal or trimodal distribution of its frequencies, with most individual falling into either of two or three clusters. This situation may indicate simple Mendelian control for the production of a compound, as the two main peaks would represent the dominant and the recessive phenotype, or in the case of three peaks, incomplete dominance. On the other hand, corrplot was used to produce a correlation matrix between all the identified compounds. The order of the compounds in the matrix was manipulated with the *hclust* algorithm (part of corrplot itself) to so as to show clear clusters of correlated molecules. This was done in order to detect associations between compounds that may hint at a shared biosynthetic origin (and may thus be indicative of genetic structure). Compounds that appeared less than 4 times in the F2 were excluded from the analysis, and 3 individuals where 11-methylpentacosane and heptacosane/nonacosane were not found (these were the same individuals). Obvious contaminants such as plasticizers were eliminated.

Correlation matrices were produced for a dataset including the two pure species and for one including F2 hybrids and backcrosses. In the case of the two F2 datasets (the C25Me and the C27-29 ones), 100 90% bootstrap replicates were extracted from the data, each one selecting 117 individuals out of the 133 that made up the full dataset. This was used as a method to reduce the

impact of outliers that may affect correlations. Correlations were calculated for all 100 of them, then averaged to produce a single final correlation matrix. Correlations for both datasets were calculated three times, using Pearson, Spearman and Kendall's coefficients to check for robustness of the associations. This was also a method to control for the effect of outliers: Pearson's correlation coefficient is calculated from the actual concentrations of the compounds, which make it a powerful test for detecting linear relationships but also particularly sensible to outliers, whereas Spearman and Kendall's coefficients are calculated based on ranks and are thus less subject to the outliers' effect.

Clusters that were consistent across both standardization methods and regardless of the correlation coefficient used in R(corrplot) were noted and all compounds within each cluster were averaged. Histograms were produced with *ggplot* to check whether any of them behaved like a single locus, much like what was done for the single compounds.

A principal component analysis was carried out on the whole dataset including all F1, F2, BC and pure individuals, in order to check whether the F2 clustered closer to any of the other groups (pure, F1 and BC). A PCA was also run on the two pure species on their own to check for consistency of this dataset with previous pure species analysis (this was necessary as the two pure species datasets were later joined together), and on a dataset including just the 10 individuals from the feeding experiment. The dataset used for PCA was not normalized by the concentration of 11-methylpentacosane, but it was logged.

3. Results

3.1 Preliminary analysis of the pure species

Past analysis. In previous analyses by Jake Morris (Ph.D), 56 compounds overall were scored in the pure species, but only 31 were shown by Kruskal-Wallis tests to be significantly different between the species, and only 20 were used in the PCA as rare compounds and compounds whose identity was unconfirmed were excluded. These 20 were: heneicosadiene, homovanillylalcohol, oleyl acetate, eicosene, icosyl acetate, (Z)-11-icosenylpropionate, phytol, ?-docosene, (Z)-9-tricosene, (Z)-11-eicosenylacetate, (Z)-9-heneicosane, hexahydrofarnesylacetone, octadecyl acetate, hexacosanal, tricosane, heneicosene, icosane, hexacosane, 11-methylpentacosane and ?-

henicosenyl acetate. Importantly, for the purpose of this analysis, the concentrations of each compound were converted to percentages by individual. The resulting PCA sorted *H. elevatus* and *H. pardalinus* individuals into two clear clusters, as expected from the two pure species given their very different pheromone blends. This dataset will be referred to as the "J" dataset.

Joining the J dataset with new data. For this analysis concentrations were favoured over percentages, and because the original concentrations were not available for all of the J dataset, some individuals had to be left out of the joint analysis. The 10 J dataset individuals (5 *H. elevatus* and 5 *H. pardalinus butleri*) for whom the concentration could be rescued were analyzed alongside newly acquired individuals, for a total of 13 *H. pardalinus* and 10 *H. elevatus*. The new individuals were analyzed alongside F2 individuals, thus 93 compounds were scored, but for the sake of joining the two datasets, most of them (including contaminants, compounds that appeared less than 4 times across both species, and compounds that were only detected in one of the two datasets) were eliminated, bringing the number of compounds down to 40. The discrepancy in the number of compounds detected in this preliminary analysis compared to the current one stems from the fact that the J dataset individuals were run on a different GC column, however most of the parental compounds that did not appear in the current dataset were present at low concentrations and were not species-specific.

A Kruskal-Wallis test was run on the compounds' concentration to check for differences between species and 28 compounds were found to differ significantly in median concentration between *H. elevatus* and *H. pardalinus* (*Table 3*). A PCA was run on these compounds and it revealed two

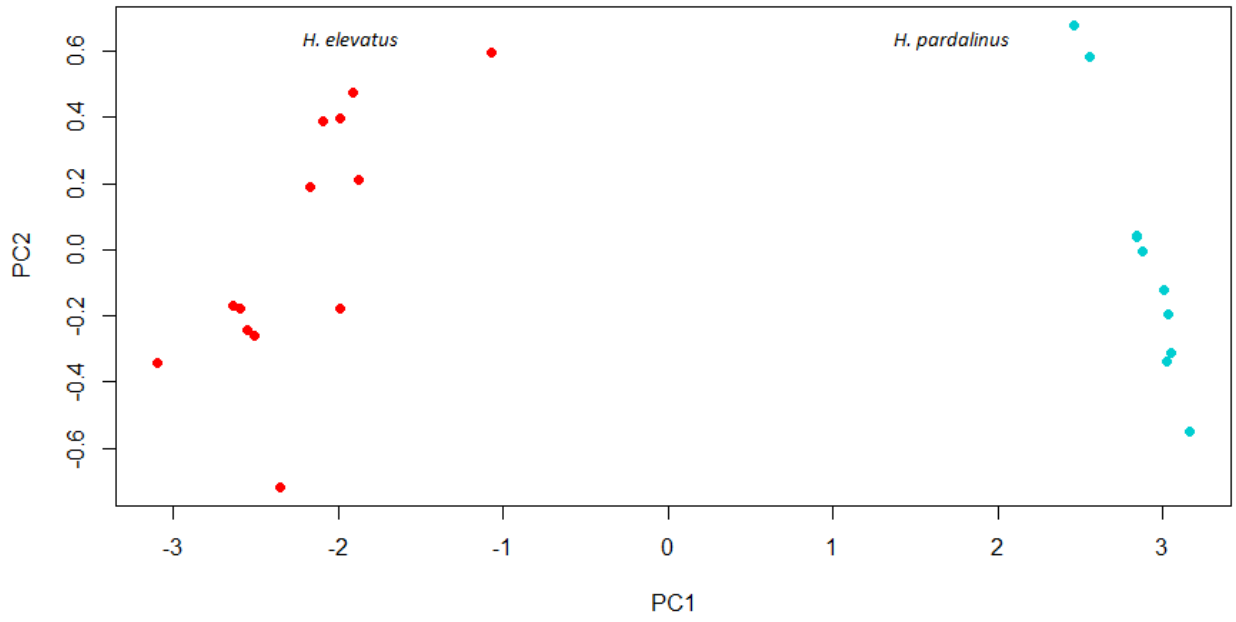


Figure 13. Principal component analysis of the pure species. Red = *H. elevatus,* cyan = *H. pardalinus.*

neatly separated clusters corresponding to *H. pardalinus* and *H. elevatus*, consistent with previous results from the J dataset in spite of the difference in the number of compounds used for the analysis (20 for the J dataset, 28 for the current one) (*Fig. 13*).

Table 3. Median, 25% (Q1) and 75% (Q3) quartiles for the concentrations (nmol) of compounds from H. elevatus and H. pardalinus. Also shown are the results of the Kruskal-Wallis test between the two species.

	H. elevatus			*H. pardalinus*			Kruskal-Wallis test	
Compound	$\bar{x}$	Q1	Q3	$\bar{x}$	Q1	Q3	p-value	Chi-square
Homovanillylalcohol	0.04	0.02	0.09	0.09	0.07	0.15	0.10	2.716
Syringaaldehyde	0.00	0.00	0.00	0.00	0.00	0.00	0.97	0.002
Diterpen	0.04	0.02	0.08	0.03	0.02	0.05	0.53	0.392
Hexadecanal	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.02**	**0.03**	**4.984**
Hexahydrofarnesylacetone	**0.00**	**0.00**	**0.00**	**5.74**	**4.54**	**6.91**	**0.00**	**17.702**
U-icosen	**0.00**	**0.00**	**0.00**	**0.03**	**0.02**	**0.06**	**0.00**	**17.769**
icosane	**0.04**	**0.01**	**0.10**	**0.00**	**0.00**	**0.01**	**0.01**	**5.974**
UU-Henicosadiene	**0.00**	**0.00**	**0.01**	**0.03**	**0.01**	**0.04**	**0.01**	**6.363**
(Z)-9-Henicosen	**0.01**	**0.00**	**0.04**	**8.70**	**7.52**	**12.74**	**0.00**	**16.429**
1-henicosene	0.00	0.00	0.00	0.00	0.00	0.05	0.52	0.412
1-Octadecanol	0.00	0.00	0.01	0.00	0.00	0.00	0.18	1.779
Henicosane	**12.99**	**8.08**	**15.51**	**0.08**	**0.05**	**0.12**	**0.00**	**16.266**
U-Docosene	**0.00**	**0.00**	**0.00**	**0.10**	**0.06**	**0.16**	**0.00**	**17.750**
oleyl_acetate	**0.00**	**0.00**	**0.00**	**0.08**	**0.06**	**0.10**	**0.00**	**15.569**
(Z)-11-icosenal	**0.00**	**0.00**	**0.00**	**0.02**	**0.00**	**0.06**	**0.04**	**4.329**
Docosane	**0.12**	**0.07**	**0.21**	**0.01**	**0.00**	**0.02**	**0.00**	**10.216**
Octadecyl_acetate	**0.00**	**0.00**	**0.00**	**0.05**	**0.04**	**0.06**	**0.00**	**15.630**
Phytol	**0.00**	**0.00**	**0.00**	**0.09**	**0.02**	**0.13**	**0.00**	**14.369**
icosanal	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.881
(Z)-11-icosenol	**0.05**	**0.00**	**0.08**	**0.62**	**0.43**	**0.76**	**0.00**	**9.397**
(Z)-9-Tricosene	**0.02**	**0.02**	**0.06**	**0.83**	**0.44**	**1.17**	**0.00**	**16.536**
Tricosane	**0.91**	**0.58**	**1.24**	**0.03**	**0.03**	**0.06**	**0.00**	**16.569**
11-Methyltricosane	0.01	0.00	0.02	0.00	0.00	0.02	0.76	0.095
(Z)-11-icosenylacetate	**0.00**	**0.00**	**0.00**	**20.63**	**17.03**	**22.13**	**0.00**	**17.274**
Tetracosane	**0.10**	**0.07**	**0.15**	**0.03**	**0.00**	**0.05**	**0.00**	**8.835**
icosyl_acetate	**0.00**	**0.00**	**0.00**	**0.12**	**0.07**	**0.18**	**0.00**	**13.573**
(Z)-11-icosenylpropionate	**0.00**	**0.00**	**0.00**	**2.08**	**1.60**	**2.70**	**0.00**	**17.692**
Pentacosane	**0.35**	**0.27**	**0.42**	**0.11**	**0.05**	**0.13**	**0.00**	**10.441**
11-Methylpentacosane	**0.64**	**0.26**	**0.70**	**0.12**	**0.08**	**0.21**	**0.00**	**8.337**
Icosyl_butyrate	0.00	0.00	0.00	0.00	0.00	0.00	0.80	0.064
(Z)-13-Docosenylacetate	**0.00**	**0.00**	**0.00**	**0.04**	**0.02**	**0.08**	**0.00**	**15.569**
Hexacosane	**0.12**	**0.08**	**0.14**	**0.04**	**0.03**	**0.07**	**0.00**	**8.240**
docosyl_acetate	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.769
11-Methylhexacosane	**0.03**	**0.00**	**0.04**	**0.00**	**0.00**	**0.00**	**0.01**	**6.619**
Heptacosane	**0.74**	**0.62**	**0.83**	**0.16**	**0.12**	**0.24**	**0.00**	**16.290**
11-Methylheptacosane	0.11	0.05	0.14	0.05	0.04	0.06	0.05	3.778
octacosane	0.06	0.04	0.11	0.01	0.01	0.08	0.10	2.750
Hexacosanal	**0.01**	**0.00**	**0.06**	**0.00**	**0.00**	**0.00**	**0.04**	**4.031**
Nonacosane	**0.52**	**0.46**	**0.61**	**0.25**	**0.18**	**0.28**	**0.00**	**14.583**
Octacosanal	**0.21**	**0.08**	**0.60**	**0.01**	**0.00**	**0.07**	**0.00**	**8.324**

R(corrplot) generates correlation plots wherein each compound is tested against itself (complete correlation with a coefficient always = 1) and every other compounds in the dataset. Each square of the matrix represents a correlation test and the colour of the squares indicates the correlation coefficient and thus the strength and direction of the relationship, with red representing strong negative correlations and blue representing strong positive ones. Shown in colours are only the significant correlations ($p<0.05$): blank squares represent pairs of compounds that had a correlation coefficient of 0 and showed no significant correlation with one another ($p>0.05$). The correlation matrix obtained from the compounds' concentrations in the two pure species (*Fig. 14*) shows two clusters of positively correlated compounds alongside a few unspecific compounds that appear correlated to both. The positive correlations are shown in blue, and the negative ones in red. These clusters represent the pheromone blends of *H. elevatus* and *H. pardalinus*. Compounds belonging to either cluster are generally negatively correlated to those belonging to the other cluster. The larger cluster, corresponding to *H. pardalinus*, includes alkenes and alkene derivatives, alcohols, aldehydes, esters, homovanillyl-alcohol and syringaldehyde (albeit syringaldehyde only correlates with homovanillyl-alcohol, docosyl acetate and (Z)-11-icosenal) phytol and hexahydrofarnesyl-acetone, while the other cluster, corresponding to *H. elevatus*, includes alkanes and methylated alkanes, and an unidentified diterpene as the only plant derivative.

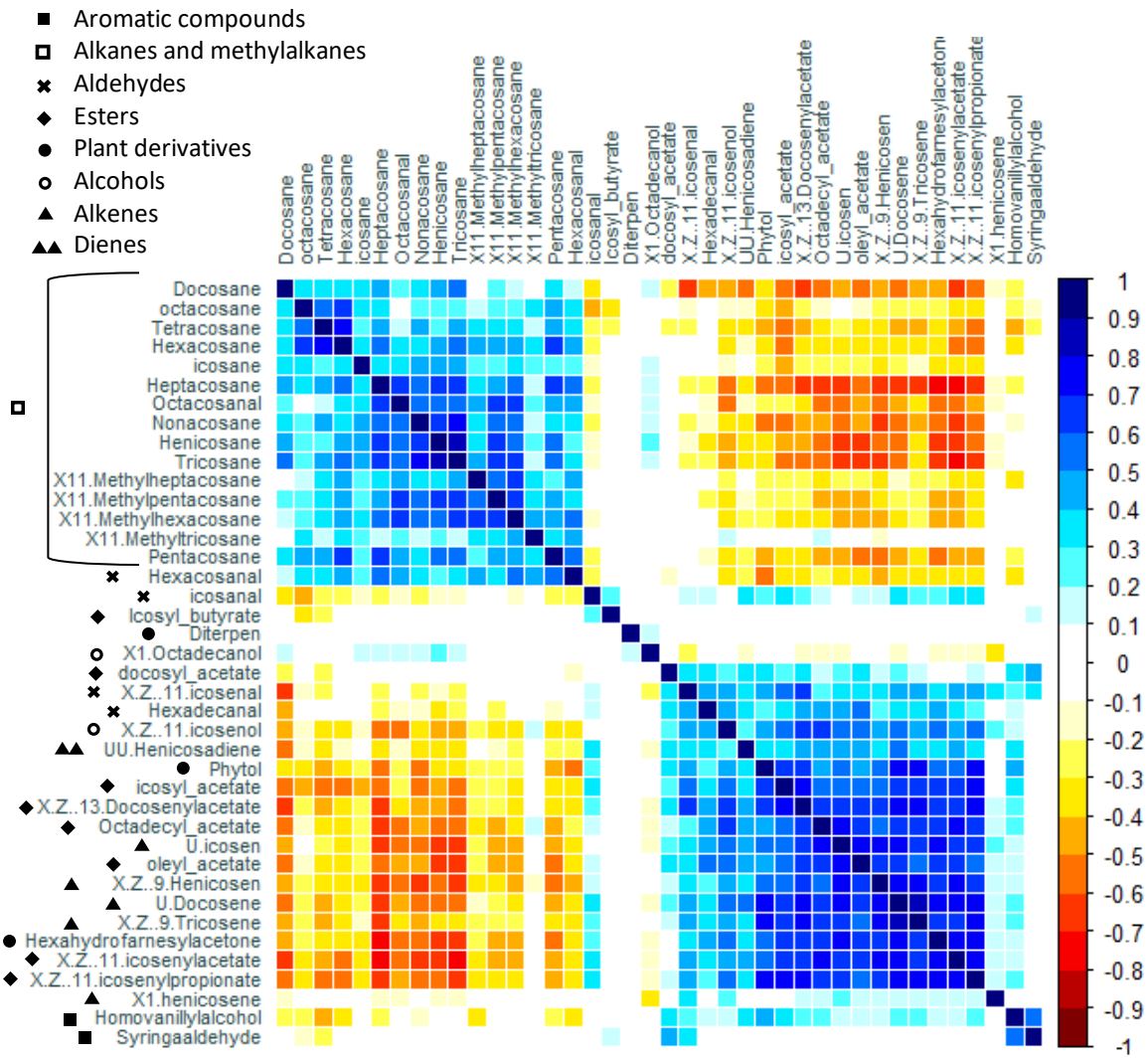


Figure 14. Correlations between compounds in the pure species. Colours represent Pearson's correlation coefficient. Only significant correlations ($p<0.05$) are shown. Different families of compounds are indicated by the small shapes.

While the presence of alkenes (especially those with a double bond in position 9) is obviously an important difference between the two species, it was not possible to determine the position of the double bond for even-numbered alkenes such as ?-icosene and ?-docosene, as well as the position of the functional groups on several compounds derived from even-numbered alkanes, and the position of the two double bonds in ??-henicosadiene. Regardless of this, they were used for the analysis due to showing obvious differences between the two species.

3.2 F2, F1 and BC analysis

Detected compounds. In total, 93 compounds were detected in the F2 but only 40 were actually used in the analysis: contaminants, unidentified molecules and compounds that appeared 4 times or less and compounds that appeared only as trace amounts were eliminated. Analyzed compounds include aromatic plant derivatives (homovanillyl-alcohol and syringaldehyde), diterpenes (mainly phytol and its derivative hexahydrofarnesyl-acetone), long-chained alkanes starting from 18C in length (albeit octadecane and nonadecane were later eliminated from the analysis due to appearing very infrequently), methylated alkanes starting from 23C, long-chained alkenes starting from 20C, including a diene (??-Henicosadiene), esters both saturated and unsaturated, including acetates, propionates and butyrates, alcohols and aldehydes (both saturated and unsaturated). An entire list of used compounds is reported in *Table 4*.

F1 and backcrosses to *H. pardalinus*. In terms of their location in the PCA scatterplot (*Fig. 17*), backcrosses to *H. pardalinus* behave as one might expect: their phenotypes are usually intermediate between the two pure species but closer to that of *H. pardalinus*. In terms of concentrations, the BC were also more *pardalinus*-like, with the compounds' concentrations being usually smaller than those observed in *pardalinus.* The main exception to this is (Z)-9-Henicosene, whose median concentration is much higher in BC than in *H. pardalinus* (*Table 4*). The BC correlation matrix shows clusters of strongly correlated compounds.

F1 hybrids on the other hand do not have intermediate phenotypes as one might expect. While in the PCA they cluster at an intermediate point between the pure species (albeit closer to *H. pardalinus*), several compounds both saturated and unsaturated appear strongly upregulated in the F1s. These are usually, but not exclusively, *H. pardalinus* compounds. They include homovanillyl-alcohol, (Z)-9-henicosene, (Z)-9-Tricosene, tetracosane, (Z)-11-icosenylacetate and propionate, and heptacosane (*Table 4*).

Table 4. List of all compounds included in the whole dataset analysis. Median, 25% quartile (Q1) and 75% quartile (Q3) of the compounds' concentrations (nmol) are reported for comparisons between the different broods.

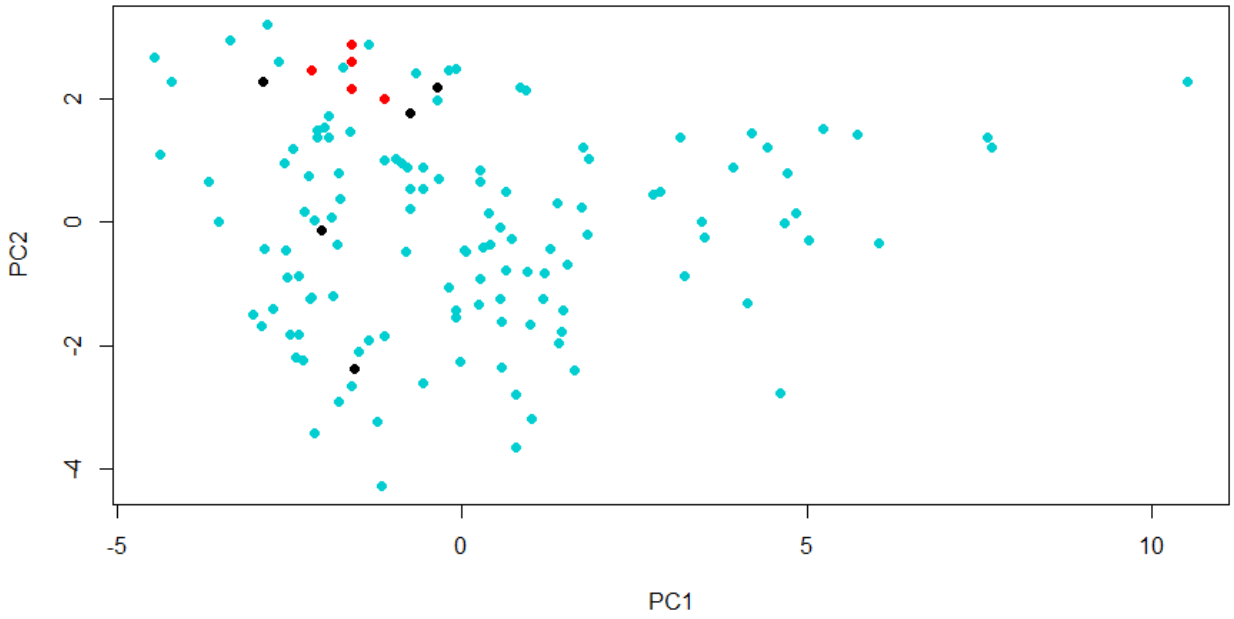
	H. elevatus			*H. pardalinus*			F1			BC			York F2			Peru F2		
Compound	$\bar{x}$	Q1	Q3	$\bar{x}$	Q1	Q3	$\bar{x}$	Q1	Q3	$\bar{x}$	Q1	Q3	$\bar{x}$	Q1	Q3	$\bar{x}$	Q1	Q3
Homovanillylalcohol	0.04	0.02	0.09	0.09	0.07	0.15	1.88	0.20	2.85	0.12	0.00	0.31	0.30	0.15	0.54	1.02	0.27	5.10
Syringaaldehyde	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	3.98
Diterpen	0.04	0.02	0.08	0.03	0.02	0.05	0.04	0.00	0.06	0.01	0.01	0.03	0.01	0.01	0.02	0.03	0.00	0.06
Hexadecanal	0.00	0.00	0.00	0.01	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hexahydrofarnesylacetone	0.00	0.00	0.00	5.74	4.54	6.91	0.18	0.00	0.36	0.17	0.12	0.54	0.07	0.02	0.28	0.09	0.00	1.27
U-icosen	0.00	0.00	0.00	0.03	0.02	0.06	0.13	0.07	0.19	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00	0.00
icosane	0.04	0.01	0.10	0.00	0.00	0.01	0.07	0.04	0.10	0.00	0.00	0.01	0.01	0.00	0.02	0.01	0.00	0.03
UU-Henicosadiene	0.00	0.00	0.01	0.03	0.01	0.04	0.09	0.07	0.13	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.00
(Z)-9-Henicosen	0.01	0.00	0.04	8.70	7.52	12.74	49.11	43.80	53.25	20.04	12.18	39.31	2.71	1.16	4.31	2.66	0.52	11.10
1-henicosene	0.00	0.00	0.00	0.00	0.00	0.05	0.16	0.08	0.21	0.02	0.00	0.07	0.01	0.00	0.02	0.00	0.00	0.05
1-Octadecanol	0.01	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Henicosane	12.99	8.08	15.51	0.08	0.05	0.12	6.46	3.93	7.85	0.30	0.15	0.89	0.66	0.32	0.85	1.38	0.36	6.72
U-Docosene	0.00	0.00	0.00	0.10	0.06	0.16	0.90	0.42	1.25	0.08	0.05	0.12	0.06	0.02	0.09	0.01	0.00	0.07
oleyl_acetate	0.00	0.00	0.00	0.08	0.06	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(Z)-11-icosenal	0.00	0.00	0.00	0.02	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Docosane	0.12	0.07	0.21	0.01	0.00	0.02	0.14	0.07	0.18	0.03	0.01	0.04	0.03	0.01	0.04	0.03	0.00	0.09
Octadecyl_acetate	0.00	0.00	0.00	0.05	0.04	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Phytol	0.00	0.00	0.00	0.09	0.02	0.13	0.07	0.02	0.12	0.07	0.03	0.08	0.02	0.00	0.05	0.00	0.00	0.04
(Z)-11-icosenol	0.05	0.00	0.08	0.62	0.43	0.76	0.39	0.00	0.92	0.17	0.08	0.25	0.13	0.04	0.38	1.33	0.29	6.75
(Z)-9-Tricosene	0.02	0.02	0.06	0.83	0.44	1.17	8.78	4.29	12.64	0.83	0.66	1.22	0.74	0.22	1.25	0.40	0.06	1.32
Tricosane	0.91	0.58	1.24	0.03	0.03	0.06	0.61	0.49	0.85	0.08	0.05	0.12	0.10	0.05	0.15	0.13	0.04	0.48
11-Methyltricosane	0.01	0.00	0.02	0.00	0.00	0.02	0.06	0.00	0.11	0.02	0.01	0.03	0.02	0.00	0.03	0.00	0.00	0.02
(Z)-11-icosenylacetate	0.00	0.00	0.00	20.63	17.03	22.13	11.56	8.67	22.41	4.22	1.13	7.06	0.41	0.02	2.89	0.04	0.00	3.45
Tetracosane	0.10	0.07	0.15	0.03	0.00	0.05	0.16	0.05	0.25	0.06	0.06	0.11	0.03	0.02	0.05	0.06	0.02	0.14
Icosyl_acetate	0.00	0.00	0.00	0.12	0.07	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(Z)-11-icosenylpropionate	0.00	0.00	0.00	2.08	1.60	2.70	3.39	1.57	7.16	0.69	0.12	1.81	0.09	0.01	0.17	0.02	0.00	0.64
henicosenyl_acetate	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pentacosane	0.35	0.27	0.42	0.11	0.05	0.13	0.34	0.20	0.53	0.19	0.12	0.23	0.14	0.10	0.19	0.15	0.08	0.38
(Z)-11-icosenylisobutyrate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
11-Methylpentacosane	0.64	0.26	0.70	0.12	0.08	0.21	0.31	0.23	0.43	0.17	0.11	0.23	0.14	0.10	0.34	0.19	0.09	0.42
Icosyl_butyrate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(Z)-13-Docosenylacetate	0.00	0.00	0.00	0.04	0.02	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hexacosane	0.12	0.08	0.14	0.04	0.03	0.07	0.03	0.00	0.09	0.03	0.01	0.04	0.02	0.02	0.03	0.05	0.02	0.12
11-Methylhexacosane	0.03	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.02	0.01	0.00	0.02	0.00	0.00	0.01
Heptacosane	0.74	0.62	0.83	0.16	0.12	0.24	0.76	0.47	1.11	0.27	0.15	0.40	0.31	0.21	0.46	0.55	0.28	1.16
11-Methylheptacosane	0.11	0.05	0.14	0.05	0.04	0.06	0.03	0.00	0.08	0.07	0.03	0.12	0.04	0.03	0.11	0.06	0.01	0.12
octacosane	0.06	0.04	0.11	0.01	0.01	0.08	0.03	0.00	0.06	0.02	0.01	0.03	0.01	0.01	0.02	0.04	0.01	0.10
Hexacosanal	0.01	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.01	0.00	0.00	0.00
Nonacosane	0.52	0.46	0.61	0.25	0.18	0.28	0.73	0.48	1.53	0.21	0.13	0.26	0.21	0.17	0.31	0.64	0.36	1.22
Octacosanal	0.21	0.08	0.60	0.01	0.00	0.07	0.18	0.09	0.37	0.12	0.06	0.22	0.03	0.02	0.07	0.12	0.02	0.34

3.3 In-depth analysis of F2 hybrids

Principal component analysis. PCA on the dataset with all F2 individuals shows no clusters and high amounts of variation, albeit the two groups from the feeding experiment clustered relatively close to each other (*Fig. 15*). No clusters were detected between the 10 individuals used for the feeding experiment when they were analyzed on their own (*Fig. 15*). With the exception of two individuals of the *P. serratodigitata* treatment, all individuals that were part of this experiment clustered together (*Fig. 15*). Due to the small sample size it is difficult to tell whether this by

chance, or if actually limiting the caterpillars to a single host plant reduces overall variance in the pheromones. When a PCA is run on the feeding experiment individuals alone, they do not form separate clusters, but the two *serratodigitata* outliers remain (*Fig. 16*).

Figure 15. PCA of the logged concentrations of the pheromone blend in F2 individuals with different colours showing the individuals from the feeding experiments. Red = fed on *P. edulis,*



black = fed on *P. serratodigitata,* cyan = F2 that were not part of the feeding experiment.

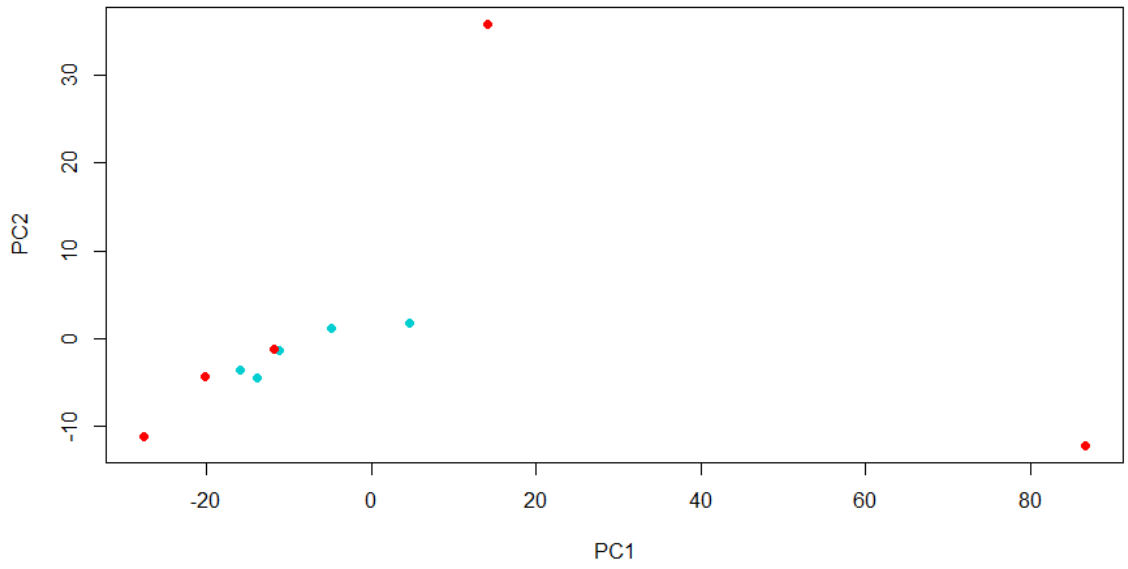


Figure 16. PCA of logged concentrations of the pheromone blend in the 10 individuals from the feeding experiment. Red = fed on *P. serratodigitata,* cyan = *P. edulis.*

PCA on the full dataset including the pure species, the F1 and F2 hybrids and the backcrosses shows that the two pure species form two well-separated clusters. All the F1 hybrids cluster together closer to *H. pardalinus* than to *H. elevatus,* with the exception of an outlier that appears very different from the rest and closer to the *H. elevatus* cluster; backcrosses show the same patterns, albeit they did cluster separately from the F1 hybrids. The F2 hybrids do not form their own cluster and instead appear scattered across the entire plot, covering the entire range of values from *H. elevatus* to *H. pardalinus.* While the 18 York F2 individuals are overall less variable than the Peru ones, they do not form separate clusters: the phenotypes of the York F2 hybrids still cover the entire range of values between the two pure species (*Fig. 17*).

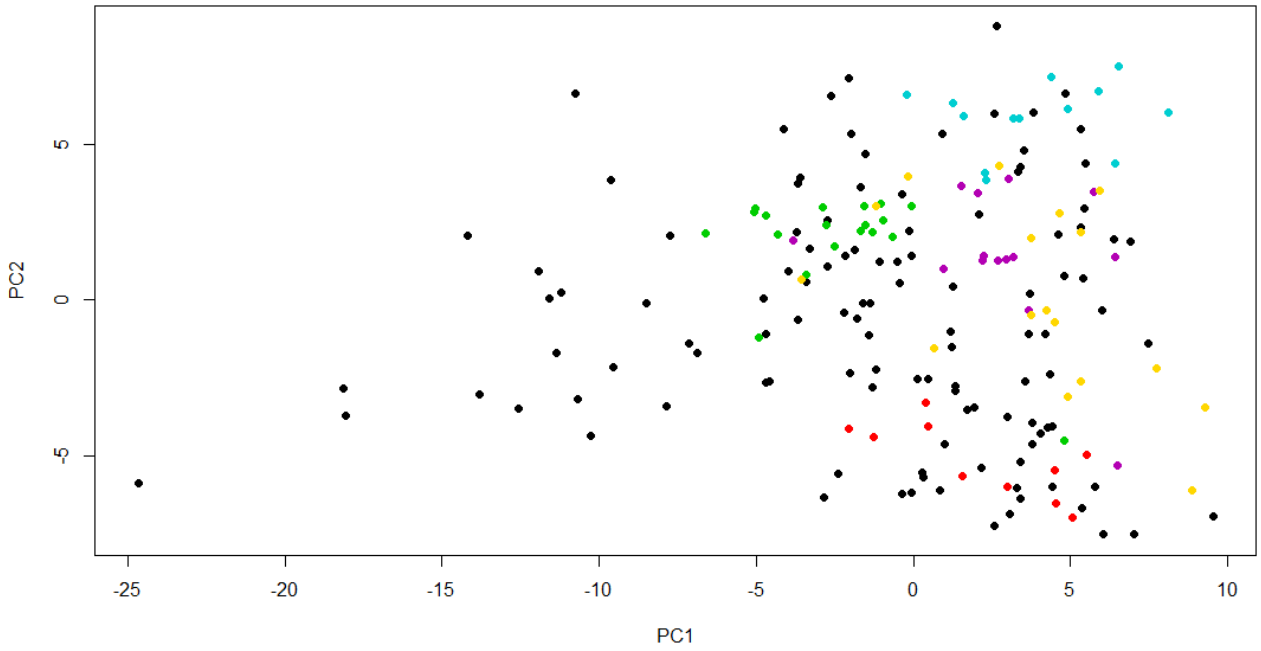
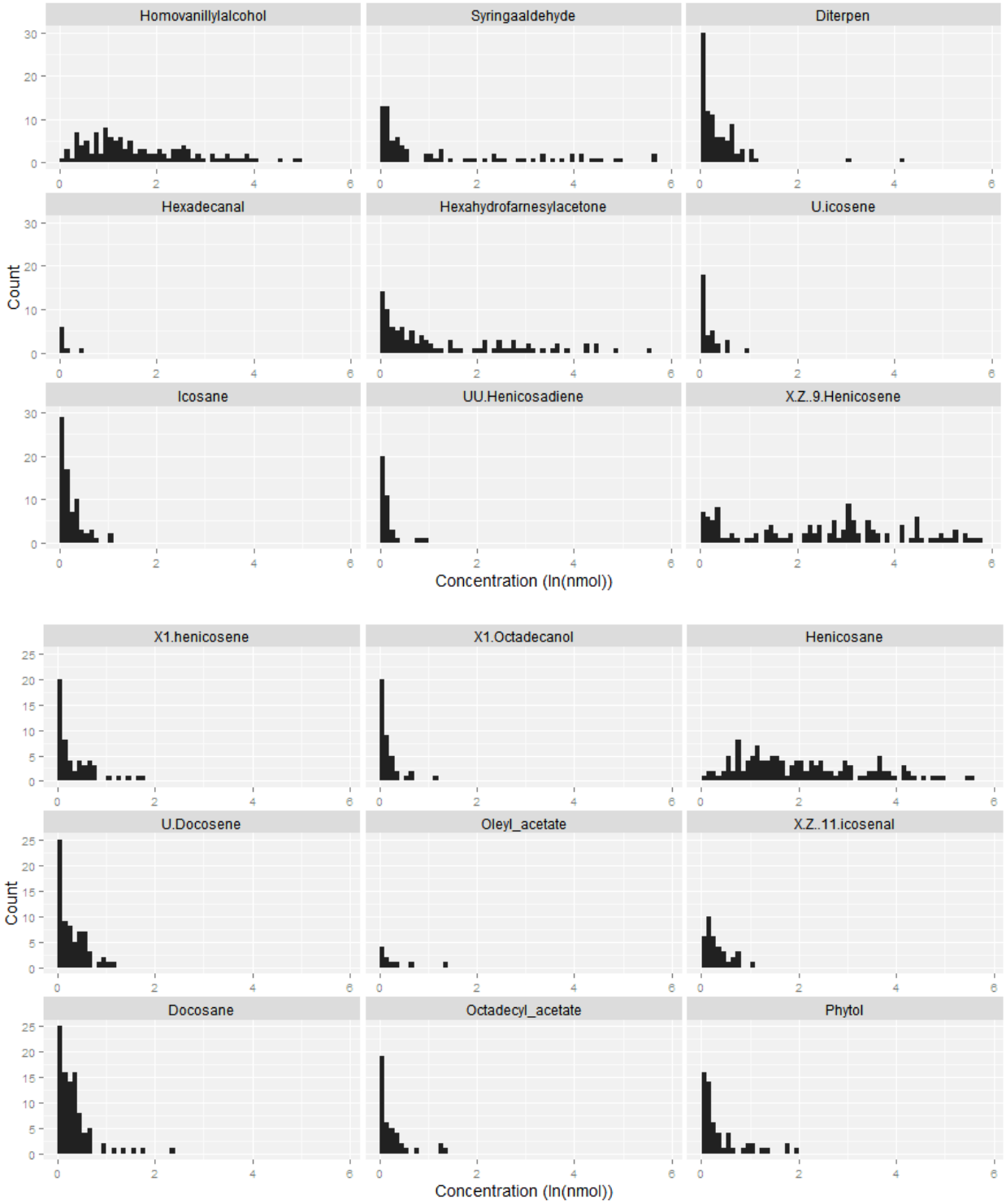


Figure 17. *PCA on the logged pheromone blends of the full dataset. Black = F2 (Peru), orange = F2 (York), cyan =* H. pardalinus, *red =* H. elevatus, *purple = backcrosses to* H. pardalinus*, green = F1.*

F2 hybrids show tremendous variability in their phenotype. No compound appears to be given by a single gene with Mendelian properties (i.e. distributions are not bimodal or trimodal), and no compound is normally distributed. Most compounds show an evident skew towards low concentrations even when logged, regardless of the normalization method, as shown by histograms produced with R(ggplot) (*Fig. 18*), though 11-methylpentacosane normalization yielded less skewed distributions in general. The most common compound is henicosane, which appeared in all 133 individuals. The two rarest compounds were hexacosanal and oleyl acetate, both of which only appeared 11 times. In the York F2 individuals the compounds with highest median values were (Z)-9-henicosene (x=2.71), henicosane (x=0.66) and (Z)-9-tricosene (x=0.74). In the Peru F2 individuals these compounds were (Z)-9-henicosene (x=2.66), henicosane (x=1.38) and (Z)-11-icosenol (x=1.33) (*Table 2*).

F2 concentration histograms (C25Me)



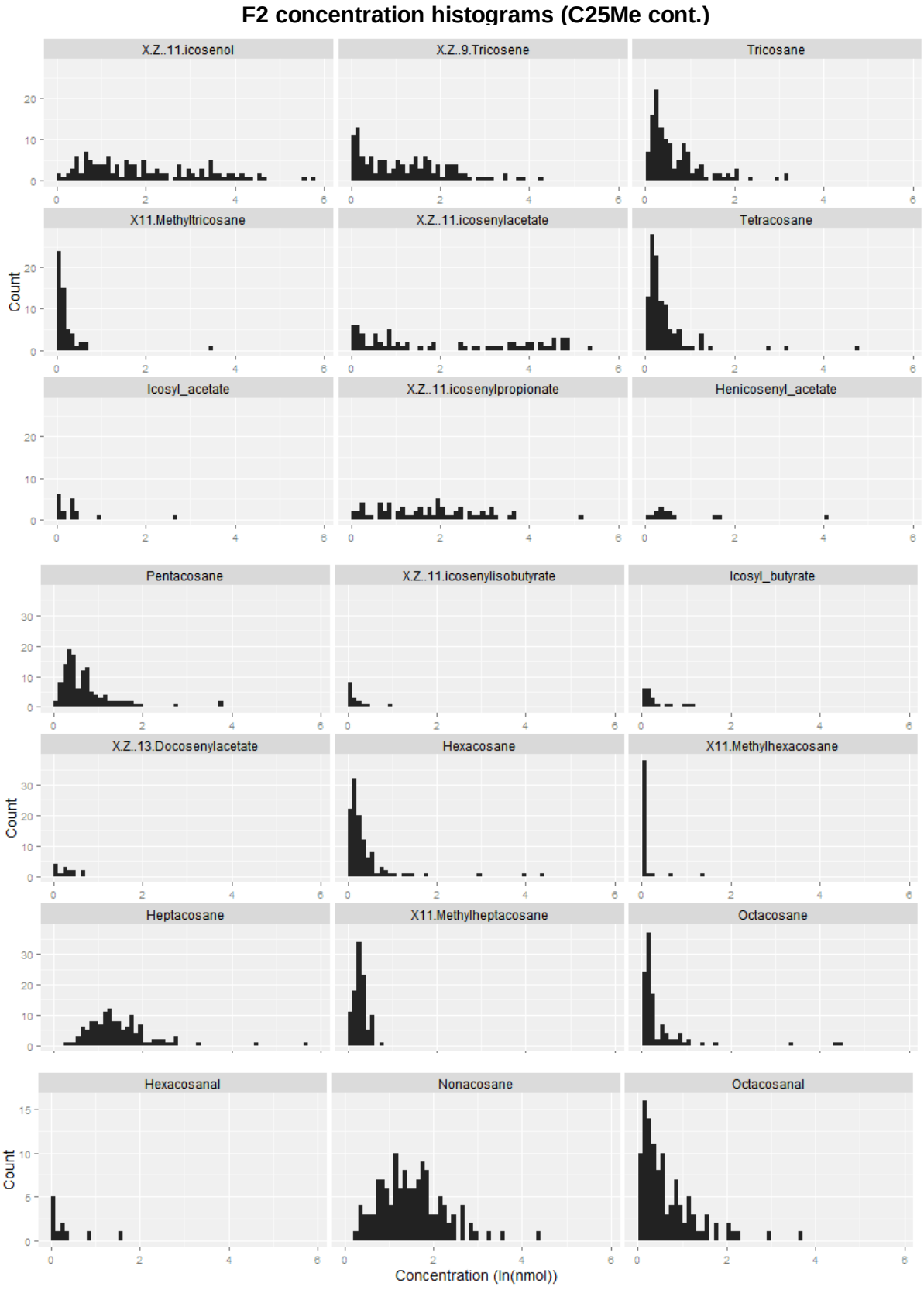


Figure 18. F2 pheromone blend components. Concentration histograms for the 11-methylpentacosane normalized dataset. Each graph represents a compound. None show bimodal or trimodal distributions expected of control by a single Mendelian locus. On each plot, the x-axis represents the compound's concentration, the y-axis is a simple count of individuals. Each bar reports the number of individuals whose concentration falls within a specific range.

The genomes of each of the F2 individuals comprise a different mosaic of contributions from *H. pardalinus* and *H. elevatus*. Correlations between compounds were used as a method to investigate the genetic structure of the hybrids' biosynthetic machinery, the assumption being that compounds controlled by the same section of the genome and those that share a common biosynthetic step in their production should be correlated. Using R(corrplot), six correlation matrices were generated: three (using Pearson, Kendall and Spearman's correlation coefficient respectively) for the C25Me dataset and three for the C27-29 dataset. The resulting correlation matrices ordered by the *hclust* algorithm on R shows several small clusters (*Fig. 19-20*). Most of these are in common between the two normalization methods (11-methylpentacosane and heptacosane/nonacosane) and they are present regardless of the three types of correlation

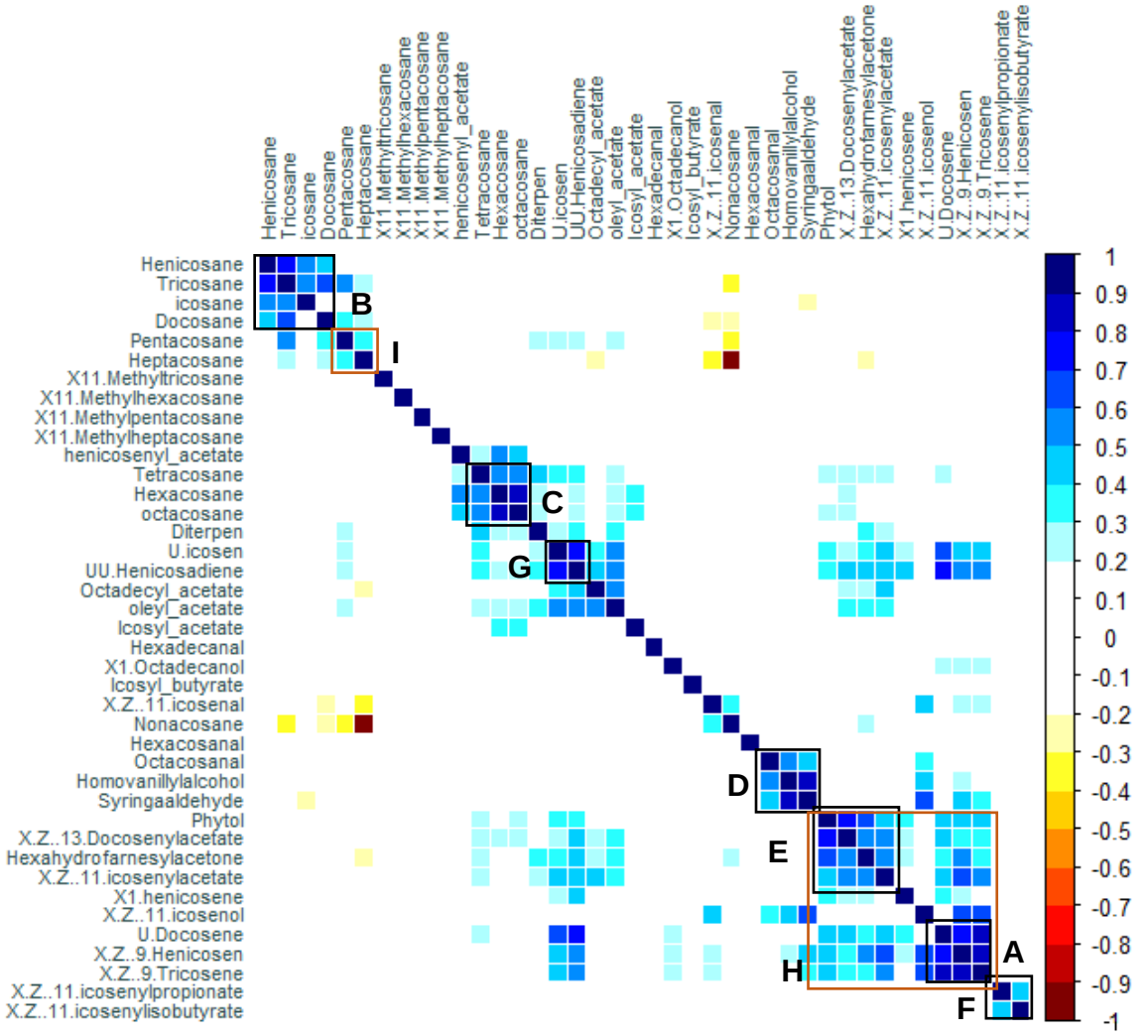


Figure 19. Correlations between compounds in the F2 standardized by C27-29. Colours represent Pearson's correlation coefficient. Only significant values ($p<0.05$) are shown.

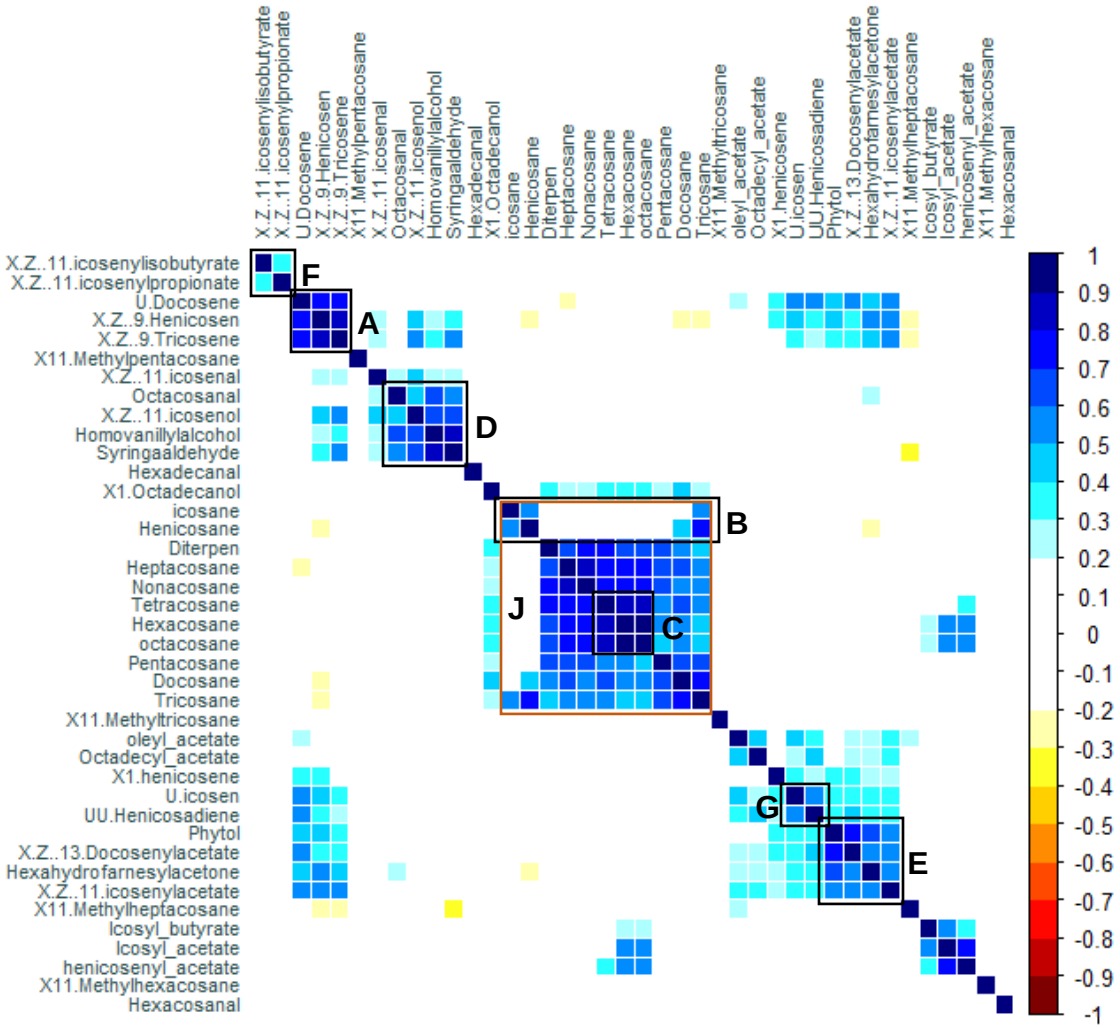


Figure 20. Correlations between compounds in the F2 standardized by C25Me. Colours represent Pearson's correlation coefficient. Only significant values ($p<0.05$) are shown.

Table 5. The main clusters of compounds detected by R(corrplot).

List of all detected clusters			
Cluster	**Common clusters**		
A	Z9-henicosene		
	?-docosene		
	Z9-tricosene		
B	henicosane		
	tricosane		
	icosane		
	docosane		
C	tetracosane		
	hexacosane		
	octacosane		
D	octacosanal		
	homovanillylalcohol		
	syringaldehyde		
E	phytol		
	HHFA		
	Z13-docosenylacetate		
	Z11-icosenylacetate		
F	Z11-icosenylpropionate		
	Z11-icosenylisobutyrate		
G	?-icosene		
	??-henicosadiene		
	Unique to C27-29		**Unique to C25Me**
H	phytol	**J**	Henicosane
	HHFA		tricosane
	Z13-docosenylacetate		icosane
	Z11-icosenylacetate		docosane
	Z9-henicosene		pentacosane
	U-docosene		heptacosane
	Z9-tricosene		Nonacosane
I	heptacosane		Tetracosane
	pentacosane		Octacosane
			Hexacosane

Seven clusters of associated compounds appeared in both the 11-methylpentacosane and the hepta/nonacosane dataset regardless of the correlation coefficient (Table 5). The first includes the main C21-C23 alkenes: (Z)-9-henicosene, ?-docosene and (Z)-9-tricosene. The second includes C20-C23 alkanes (icosane, henicosane, docosane, tricosane); it is worth mentioning that occasionally these compounds appeared as part of a more inclusive alkane cluster in the 11-methylpentacosane dataset, but this cluster was not robust across all methods. The third includes large even-numbered alkanes in the C24-C28 range. The fourth includes homovanillyl-alcohol, syringaldehyde and the large aldehyde octacosanal- this cluster consistently includes (Z)-11-icosenol in the heptacosane/nonacosane dataset. The fifth cluster includes phytol, hexahydrofarnesyl acetone, (Z)-11-icosenyl acetate and (Z)-13 docosenyl acetate, whose correlation with the other three compounds is often weaker than the other correlations in the cluster. The sixth cluster consists of two complex esters, (Z)-11-icosenyl propionate and (Z)-11-icosenylisobutyrate. icosenyl isobutyrate. Lastly, the seventh cluster includes two minor alkenes, ?-icosene and ??-henicosadiene. When averaged, no cluster behaved like a single-locus trait (*Fig. 22-23*). Structures of all the compounds belonging to these seven clusters are in Fig. 21.

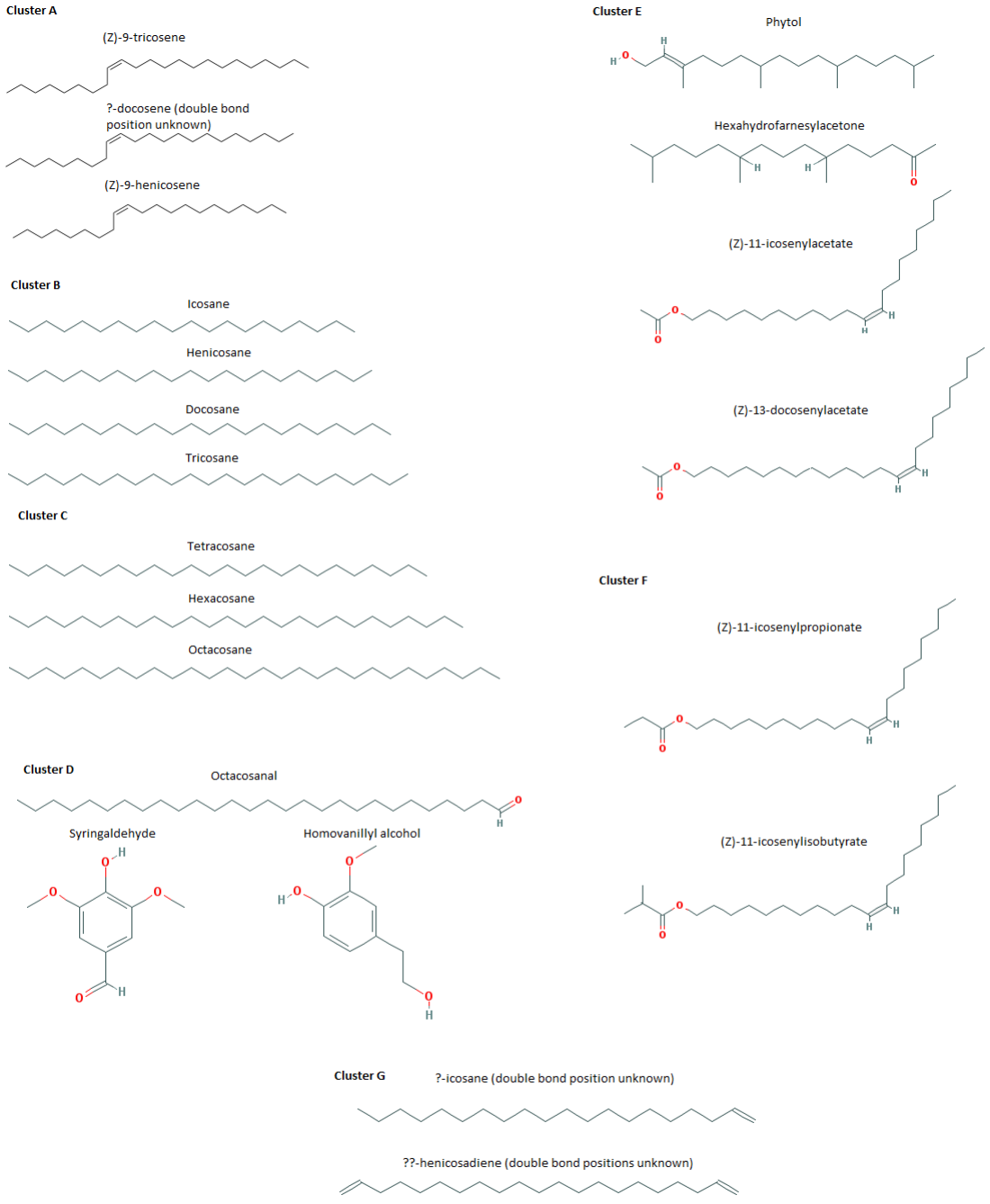


Figure 21. Chemical structures of the compounds belonging to the main clusters A-G.

In the 11-methylpentacosane dataset most alkanes formed a single cluster, which was always broken up into smaller clusters in the heptacosane/nonacosane dataset. Conversely, heptacosane

and pentacosane always formed their own separate cluster in the hepta/nonacosane dataset, and phytol, hexahydrofarnesyl acetone, (Z)-11-icosenyl acetate and the main C21-23 alkenes formed a large cluster, further subdivided into smaller blocks based on the strength of the association (*Table 3)*.

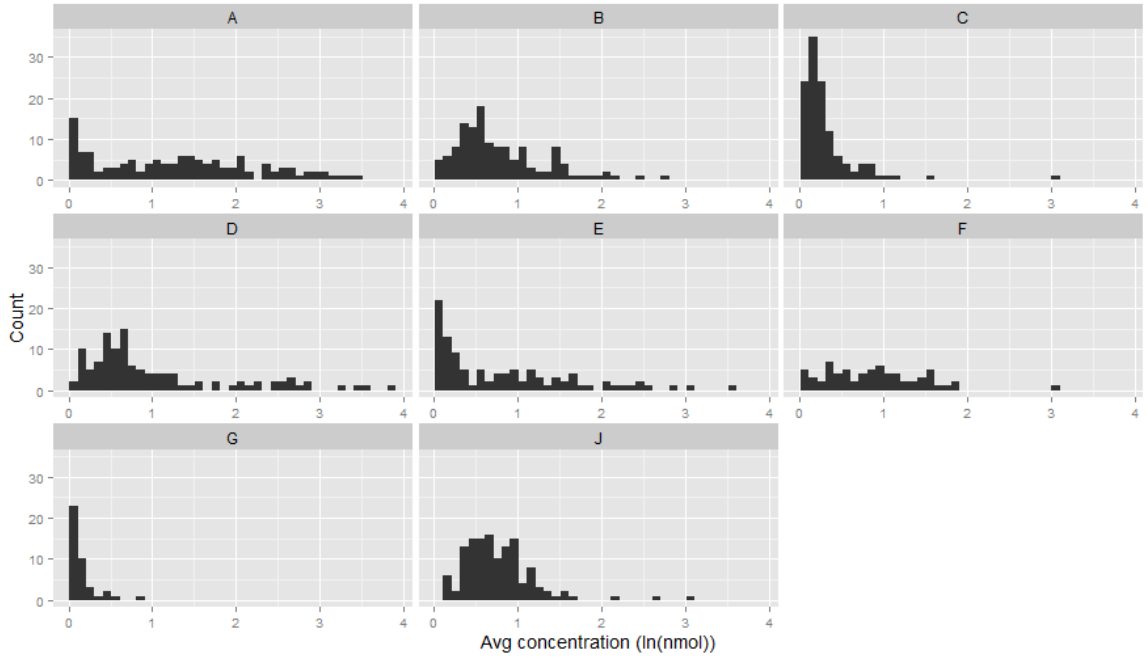


Figure 22. Histograms for the averaged concentrations of compounds belonging to each cluster in the C25Me dataset. Cluster names are indicated at the top of each graph. Cluster A= C21-23 alkenes; B= C20-23 alkanes; C= large even numbered alkanes; D= vanillin derivatives; E= phytol derivatives; F= complex esters; G= minor alkenes; J= all alkanes. For a complete list of compounds in each cluster, see *Table 5.*

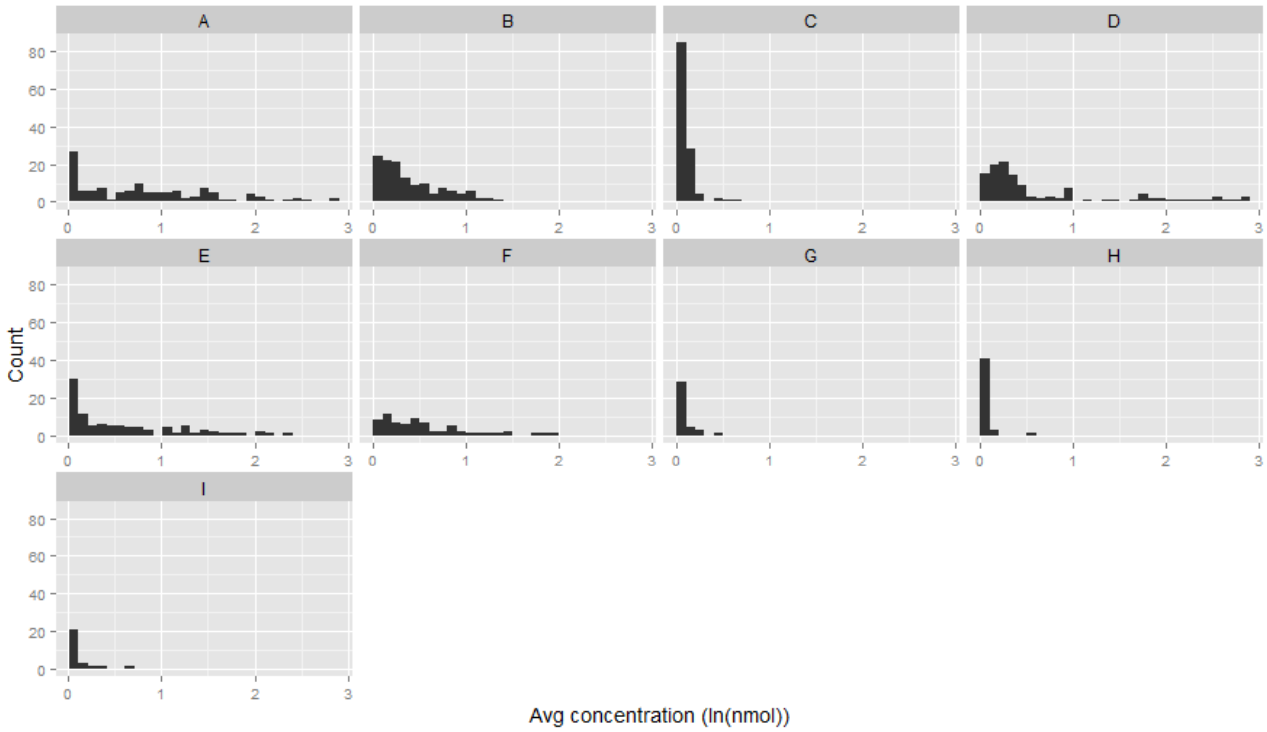


Figure 23. Histograms for the averaged concentrations of compounds belonging to each cluster in the C27-29 dataset. Cluster names are indicated at the top of each graph. Cluster A= C21-23 alkenes; B= C20-23 alkanes; C= large even numbered alkanes; D= vanillin derivatives; E= phytol derivatives; F= complex esters; G= minor alkenes; H= alkenes and phytol derivatives; I= large odd-numbered alkanes. For a complete list of compounds in each cluster, see *Table 5.*

4. Discussion

The results from the pure species analysis were consistent with previous analysis of the male sex pheromone blend in *H. elevatus* and *H. pardalinus,* as discussed below. F1 individuals were expected to have concentrations intermediate between the two pure species, but in fact several of their compounds were strongly upregulated. Backcrosses had concentrations similar to *H. pardalinus* as expected. F2 individuals were highly variable, but correlation matrices revealed clusters of correlated compounds, potentially hinting at the underlying genetic structure.

4.1 Male sex pheromone differences between *H. elevatus* and *H. pardalinus*.

Preliminary PCA analysis carried out by Jake Morris (PhD) on the percentages of compounds found on 24 pure individuals showed the different species as clearly separated into neat clusters. Previous GC/MS analysis revealed that *H. elevatus* and *H. pardalinus* have very different

pheromone blend components: the former has a relatively simple blend comprising highly concentrated long-chained alkanes starting from icosane (20C), accompanied by methylated alkanes starting from 11-methyltricosane (23C) and very low concentrations of alkenes (20-23C). Conversely the latter has a complex blend that includes high concentrations of alkenes (C20-23), esters, phytol derivatives, the aromatic compounds homovanillyl-alcohol and syringaldehyde, alcohols and aldehydes, with alkanes being found in very low concentrations. In light of this, the pure species correlation matrix behaved as expected: the compounds that were previously found to be typical of *H. elevatus* appeared all positively correlated to one another, and negatively correlated to the typical *H. pardalinus* compounds, and vice versa (*Fig. 14*). With no recombination, the pure species phenotype behaves as a single block of co-varying elements. In *H. elevatus,* the most prominent compounds are henicosane (C21) and tricosane (C23), while in *H. pardalinus* they are the corresponding alkenes (Z)-9-henicosene and (Z)-9-tricosene. Increasing the sample size of individuals analysed and using concentrations rather than percentages does not change these results. Of all these compounds, homovanillyl-alcohol was the only one that gave contrasting results between the percentage dataset and the actual concentration dataset: in the former, its relative amounts appeared significantly different between the two pure species, while in the latter they do not. However, the pure species correlation matrix still revealed that the concentrations of homovanillyl-alcohol co-vary with those of other *pardalinus* compounds.

4.2 Fatty acid derivatives in *H. elevatus* and *H. pardalinus.*

Most compounds detected in *H. elevatus* and *H. pardalinus* are fatty acid derivatives that range in size from C20 to C29 (*Fig. 24*). Considering most moths use molecules below C20 in size, this may appear unusual but it is not the first time such large compounds have been detected in a butterfly's pheromones. *Idea leuconoe,* a Danaine butterfly, also shows long chain hydrocarbons as part of its blend, including, most prominently, tricosane and (Z)-9-tricosene, alongside other alkanes and alkenes (as well as several other classes of large FA derivatives) (Nishida et al., 1996). *H. melpomene* itself has been shown to use compounds in the C18-C22 size range, mainly aldehydes both saturated and unsaturated (Vanjari et al., 2015).

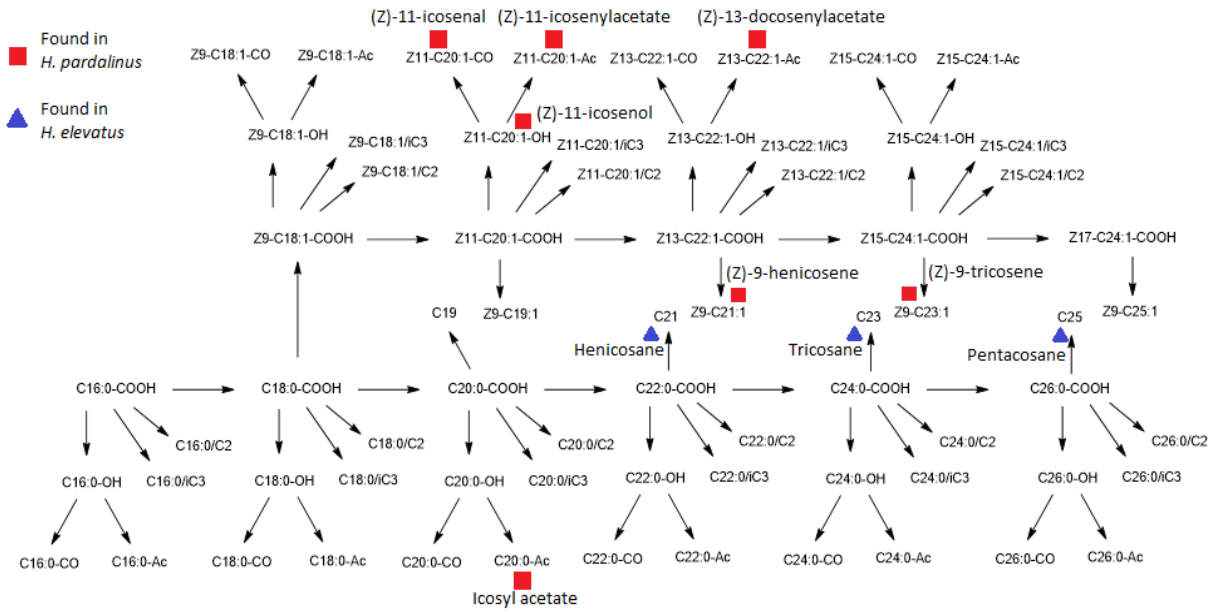


Figure 24. The FA metabolic pathway showing the most prominent H. pardalinus and H. elevatus specific compounds. Even-numbered alkanes and alkenes, as well as their derivatives, minor compounds, and compounds of unclear structure are not included in this figure.

In *H. elevatus,* the high number and high concentrations of alkanes relative to other compounds suggest a relatively simple mechanism for pheromone production. Alkanes have no side groups or double bonds, and they are simply created via the elimination of the carboxy- group from their fatty acid precursors, a reaction initiated by FARs that convert the fatty acid into its derivative alcohol (Matsumoto, 2010). The different lengths displayed by alkanes are given not by subsequent elongation of the alkane chain itself, but by fatty acid elongation prior to their reduction. Fatty acid elongation is a reaction that proceeds two carbons at a time: that is because the new units are provided by a molecule, malonyl-CoA (Voet & Voet, 2011), which consists in a two-carbon chain. It is itself a fatty acid derivative, with malonic acid being its precursor. Thus, even-numbered FAs are always present in higher amounts than odd-numbered FAs.

Because alkanes (and alkenes) are ultimately formed via loss of the fatty acid's carboxyl group, including its single carbon atom, an alkane always sports one less carbon than its fatty acid precursor, leading to an odd-numbered chain (Voet & Voet, 2011). Thus one would expect the most abundant alkanes in the *H. elevatus* blend to be the odd-numbered ones, and this is

confirmed by the data: henicosane (C21) and tricosane (C23) are not only the most abundant alkanes, they are also the most abundant compounds of the *H. elevatus* blend overall. All odd-numbered alkanes up to nonacosane, the largest odd-numbered alkane scored in this analysis, have higher concentrations than the even-numbered ones, not only in *H. elevatus,* but also in *pardalinus,* where all alkanes have low concentrations overall. The odd-numbered alkanes' averaged concentrations are also more variable. While alkanes of any size between C20 and C29 are produced in *H. elevatus,* any alkane larger than pentacosane (C25) is more likely to be a cuticular hydrocarbon than a pheromone component (Vanjari et al., 2015).

While in *H. elevatus* the alkanes make up the majority of the blend, in *H. pardalinus* alkenes and unsaturated esters are the most abundant compounds (*Fig. 24*). The production of both these classes of compounds primarily depends on the introduction of a double bond within a fatty acid precursor, performed by fatty acid desaturases (Roelofs & Rooney, 2003). This process begins with octadecanoic acid, also known as stearic acid (C18-COOH): it is transformed into (Z)-9-octadecenoic (oleic) acid (Z9-C18:1) by a Δ9-desaturase, and its length is then adjusted by elongation (Mann F. et al, unpublished), by the same process as saturated fatty acids. Thus, unsaturated fatty acids with an odd carbon number are once again expected to be present at higher concentrations than even-numbered ones. The detectable alkenes, unlike the alkanes, are restricted to the C20-C23 range, and the medians of (Z)-9-henicosene and (Z)-9-tricosene were higher than those that of ?-icosene and ?-docosene (wherein the double bond position could not be determined). The exception to this is 1-henicosene, but this compound's concentration was not significantly different between the two species, and it was usually very small (on average, 0.01 nmol in *H. elevatus* and 0.02 nmol in *H. pardalinus*).

The production of esters begins with the reduction of the FA precursor by FARs, into an alcohol. This alcohol can then be transformed into either an aldehyde or an ester (Liénard et al., 2010). This process does not involve the loss of one carbon from the main chain, unlike the production of alkenes and alkanes. Thus, a FA-derived ester will have the same number of carbons as its FA precursor, and esters with an even number of carbons are tendentially more common. *H. pardalinus*'s main ester, (Z)-11-icosenyl acetate (Z11-C20:Ac), derives from (Z)-11-icosenoic acid with (Z)-11-icosenol as an intermediate, which can also be transformed into (Z)-11-icosenal. Like all fatty acid (Z)-11-icosenoic acid is undetectable and its alkene derivative, (Z)-9-nonadecene, is

not part of the *pardalinus* blend. The other even-numbered unsaturated ester found in *H. pardalinus,* (Z)-13-docosenyl acetate (Z13-C22:Ac), derives from (Z)-13-docosenoic acid by the same reaction. *H. pardalinus* also produces other esters: icosyl acetate (C20:Ac), from the saturated icosanoic acid, and oleyl acetate (?-octadecenyl acetate) from oleic acid. It is worth noting that both species can produce both alkanes and alkenes: the vast differences in the abundances of these two classes of compounds are what distinguishes the two species.

4.3 Species-specific non FA-derivatives: compounds sequestered from the food plants.

Other than FA derivatives both species show unique compounds that derive from plants. In *H. elevatus* the only compound of this type is a diterpene, but lack of info about its structure made it impossible to use in the analysis. In *H. pardalinus* there are two groups of plant derivatives: the aromatic compounds, which include two vanillin-like molecules, homovanillyl-alcohol and syringaldehyde, both derived primarily from lignin, and the terpenoids phytol and hexahydrofarnesyl acetone, both derived from chlorophyll. Both of these were probably acquired at the larval stage given that lignin and chlorophyll are found in the stem and leaves of the host plants primarily. Because these are ubiquitous compounds, *H. elevatus* caterpillars also acquire them while feeding, but it appears that only *H. pardalinus* has evolved a mechanism to produce pheromones from them. Whichever their origin is, it is clear that *Heliconius* are capable of synthesizing their compounds both by sequestration of plant derivatives and by *de novo* production, as reported for other species in the past (Tillman et al., 1999).

4.4 Several biosynthetic reactions are upregulated in F1 but not in BC.

The high median concentrations of compounds found in the F1 individuals may be indicative of an interesting phenomenon: by hybridizing, *H. pardalinus* and *H. elevatus* are apparently loosening their regulatory mechanisms, leading to a general increase in the concentrations of products. This does not only involve the FA biosynthetic pathway however, since it also affects homovanillyl-alcohol. However, these F1 results are currently difficult to interpret without further knowledge of the genetic architecture controlling *Heliconius* male sex pheromones. In the backcrosses this is not the case. The backcrosses' correlation matrix shows several associations that are present in the F2 as well, but also many cases where compounds appear correlated that were not significantly correlated in the F2. This implies that the BC blend is not under the same control mechanisms as

the F1 and the F2 blend- a finding that justified the exclusion of backcrosses from the correlation analysis of F2 hybrids (*Fig. 25*).

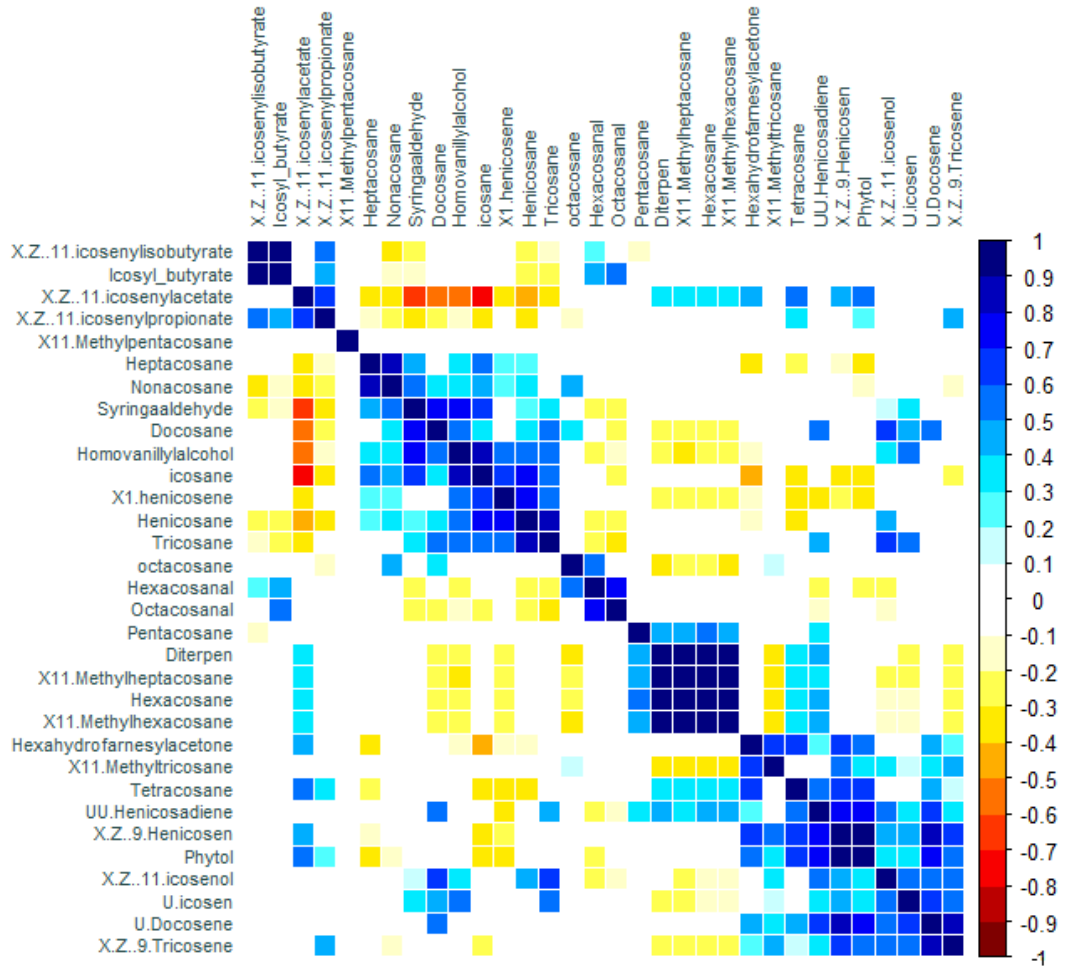


Figure 25. Correlation matrix for logged compound concentrations in the backcross. Colours represent Pearson's correlation coefficient. Only significant correlations ($p<0.05$) are shown.

4.5 F2 analysis: no compound acts like a Mendelian locus.

Single-compound concentration analysis with *ggplot2* was not deemed useful to understand the genetic mechanisms underlying pheromone production. The histograms had two purposes: 1) to show whether each compound's concentrations were normally distributed, and 2) to find out whether any of the compounds showed a bimodal distribution, potentially hinting at simple genetic control. It is unlikely that any compound's production would be controlled by a single

gene, but if a single Mendelian locus were having a large effect on the concentration, one might expect to see a bimodal distribution with the two peaks representing two different concentration phenotypes (subject to quantitative variation), or potentially a trimodal distribution in a scenario of incomplete dominance. Instead, no histograms showed any such patterns. Standardizing by 11-methylpentacosane removed more skew than standardizing by heptacosane/nonacosane. The PCA similarly revealed no subgroups within the F2: individuals are continuously distributed between the two parental phenotypes. This makes it difficult to infer how many genes underlie the production of a single compound.

4.6 Clusters partially resolve the biosynthetic machinery.

Clusters detected in the correlation matrices can offer insight into the problem. The F2 correlation matrices, as expected, do not show the two clear pure species blocks: recombination deconstructed the parental phenotype. Those detected with all three correlation methods and across both datasets appear to have biological significance, though others may be an artefact of the standardization method. As 7 clusters were detected, one may infer that there are at least 7 unlinked genomic region controlling pheromone production.

Cluster A: C21-23 alkenes. A prominent cluster contains (Z)-9-henicosene (Z9-C21:1), ?-docosene (?-C22:1) and (Z)-9-tricosene (Z9-C23:1) (*Fig. 26)*. This cluster contains two of the most important compounds of the *H. pardalinus* blend, Z9-C21:1 and Z9-C23:1. These alkenes are the simple product of Δ9-desaturation of stearic acid into (Z)-9-oleic acid, followed by its elongation and reduction. This suggests the presence of a Δ9-desaturase that acts on C18 substrates similarly to those seen in *Ostrinia nubilalis* (Roelofs et al., 2002). Z9-C21:1 and Z9-C23:1 are present in both species, meaning that they must necessarily share the gene encoding this FAD, yet the average

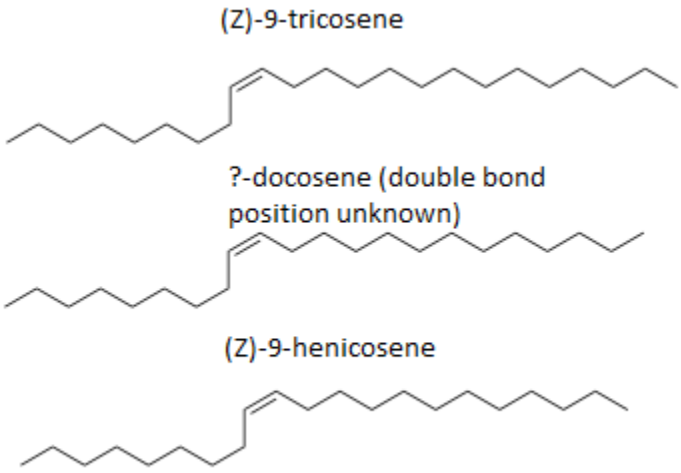


Figure 26. Cluster A. Note that in ?-docosene, neither the double bond's position nor its orientation (E or Z) are known.

unlogged concentration of (Z)-9-henicosene in *H. pardalinus* and in *H. elevatus* are approximately in a 500:1 ratio, and the ratio is 24:1 for (Z)-9-tricosene. This must mean that the species-specific switch did not involve the FAD itself, but rather its control system, leading to upregulation in Δ9-desaturase production in *H. pardalinus*. This does not necessitate a complex evolutionary history: a single mutation can have large results on pheromone production, as seen in different *Ostrinia* races (Roelofs et al., 2002). ?-docosene is never observed in *H. elevatus* which would seem to pose a problem to this hypothesis. However it may simply be a side product: its unlogged concentrations in *H. pardalinus* average to 0.12 nmol against (Z)-9-henicosene's average of 10.9 nmol and (Z)-9-tricosene's average of 0.94 nmol, yet the three are correlated. It may be undetectable in *H. elevatus* simply because the concentrations of Z9-C21:1 (on average 0.02 nmol) and Z9-C23:1 (0.04 nmol) are too low for its biosynthesis as a side product (or it could be that ?-docosene is produced, but its concentrations are undetectable). The same phenomenon is seen with ?-icosene, which occasionally appears as part of this cluster (its concentration in *H. pardalinus* being even lower than ?-docosene, at 0.04 nmol).

Clusters B and C: alkanes. The presence of different alkane clusters (*Fig. 27*) confirms the presence of FARs with different substrate specificities. The main cluster is C20-23, which includes henicosane and tricosane, the main compounds of the *H. elevatus* pheromones, and the even-numbered icosane and docosane. Once again, the low concentrations of even-numbered alkanes and their correlation with henicosane and tricosane seem to

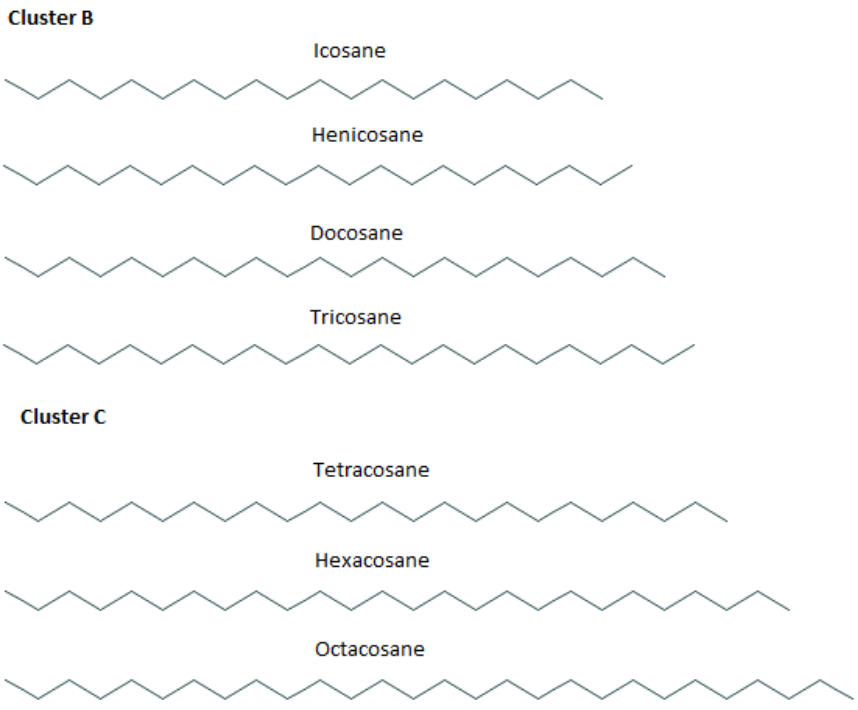


Figure 27. The two most robust alkane clusters, B (C20-23 in length) and C (C24-28 in length).

suggest that they are side products. Within the heptacosane/nonacosane dataset it is possible to observe an association between pentacosane and heptacosane- these two compounds are always strongly correlated but only in the heptacosane/nonacosane dataset they form a cluster separately from other alkanes. However out of this two, heptacosane is unlikely to play a major role in chemical signalling due to its large size. This potentially suggests the presence of two FARs, one that acts on FAs of a maximum length of C24, and one that acts on larger compounds, regardless of their function. The third alkane cluster, comprising the even-numbered low-concentration alkanes tetracosane, hexacosane and octacosane, is peculiar in that it contains no odd-numbered alkanes, and it includes non species-specific compounds (hexacosane and octacosane), which may be cuticular hydrocarbons. This suggests the presence of three FARs. All organisms are expected to have several types of FARs and the evolution of FARs with different substrate specificities has been shown in the past to require only a few amino acid substitutions (Chacòn et al., 2013). Once again, their presence in both parental species suggests that *H. elevatus* has only upregulated its alkane production rather than evolving novel FARs.

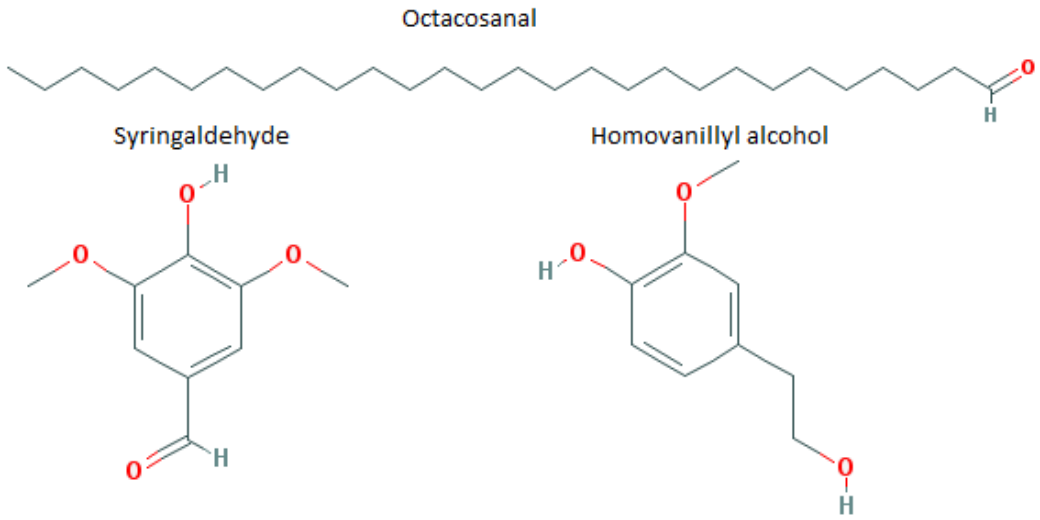


Figure 28. Cluster D. Notice the similarity in structure between homovanillyl-alcohol and syringaldehyde.

Cluster D: aromatic vanillin derivatives. The two aromatic compounds detected in *H. elevatus* and *H. pardalinus*, homovanillyl-alcohol and syringaldehyde, are correlated to each other and to octacosanal (*Fig. 28*), occasionally showing correlations with other alcohols and aldehydes as well.

Homovanillyl-alcohol and syringaldehyde belong to the same class of compounds, methoxyphenols, which are plant derivatives. Homovanillyl-alcohol, also known as vanillic alcohol, is a vanillin derivative whose biosynthetic mechanism in animals is unstudied. It is known however that in yeast it can be produced via the simple reduction of vanillin (Loscos et al., 2007). Syringaldehyde and vanillin are not derivatives of one another, but they both derive from the degradation of lignin in weakened plants (Meyer & Norris, 1967). They are behaviourally active as have attractant effects on *Scolytus* bark beetles that lay their eggs within damaged bark (Meyer & Norris, 1967). So both homovanillyl-alcohol and syringaldehyde are ultimately lignin derivatives, which makes it is highly unlikely that *H. elevatus* and *H. pardalinus* are producing these compounds. It is far more plausible that they are obtaining the compounds from their food plant, and it is probable that this uptake occurs at the larval stage of their life cycle (as adults feed on nectar and pollen), when they feed on the leaves and stems of *Passiflora,* and that the compounds are stored, only to be transported in the androconia when the butterfly reaches adulthood. This kind of behaviour is not unheard of in Lepidoptera since it is also observed in *Pieris brassicae,* wherein the larvae take up plant compounds by feeding and later they are transformed into pheromone components (Schulz et al., 2011). If this is the case, the different levels of HVA and SA detected in the two species (albeit the difference is not significant) are probably due to the activity of a transporter. The significance of octacosanal in relation to these two compounds is unknown, but it could be a component of the food plant's cuticular wax (Schmid & Bandi, 1969).

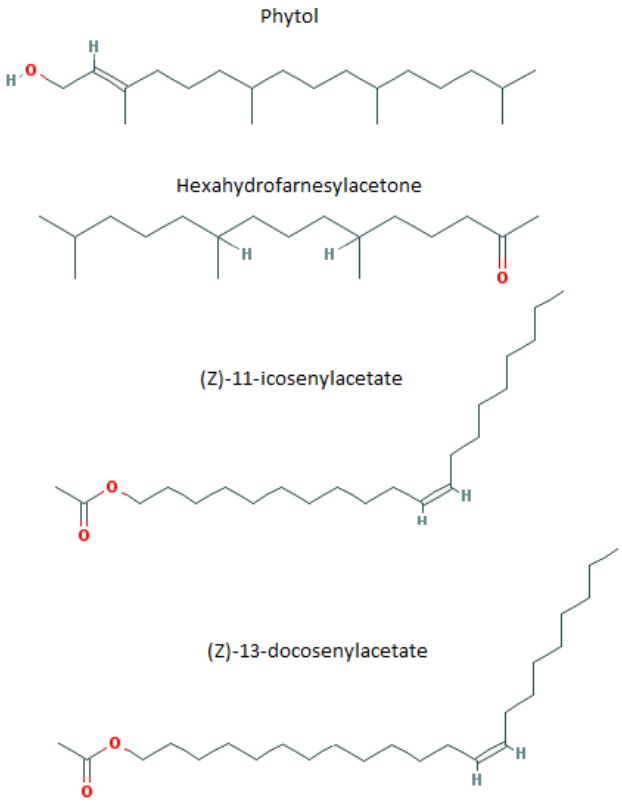


Figure 29. Cluster E. The similarity in structure between phytol and HHFA is evident. The acetate groups potentially derived from phytol degradation are circled.

Cluster E: phytol, its degradation product and unsaturated esters. The cluster comprising phytol, hexahydrofarnesylacetone (HHFA) and the esters also appears to have biological significance (*Fig. 29*). These compounds only appear in *H. pardalinus:* they do appear one time in *H. elevatus*. Phytol and HHFA are both plant derivatives: the former, in fact, is a component of chlorophyll and is thus ubiquitous in plants. HHFA is a product of the oxidative degradation of phytol and it is used as a pheromone by *Pieris brassicae*, phytol being the aforementioned compound that they can assimilate during their larval stage (Schulz et al., 2011). Thus it makes sense that these two compounds would be correlated. The production of HHFA also releases acetates (Schulz et al., 2011), that may be incorporated into the production of unsaturated esters such as (Z)-11-icosenylacetate and (Z)-13-docosenylacetate.

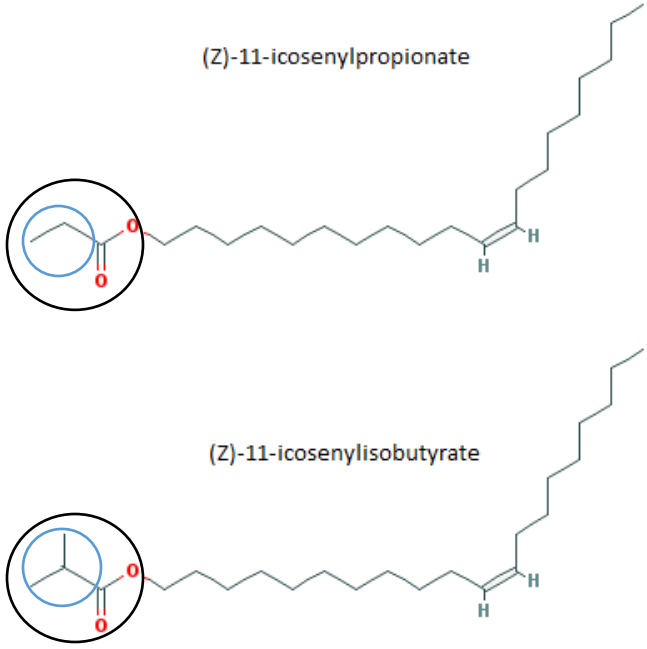


Figure 30. Cluster F. The main chain's length is the same in both esters, but the functional group's length changes. The black circles indicate the propionate and butyrate group, the blue ones the acetate portion of the functional group.

Cluster F: propionates and butyrates. This cluster includes complex esters, that is, esters derived from the elongation of (Z)-11-icosenylacetate's acetate chain (*Fig. 30*). These consist in (Z)-11-icosenylpropionate and (Z)-11-icosenylisobutyrate. Both (Z)-11-icosenylacetate and (Z)-11-icosenylpropionate are commonly found in *H. pardalinus* but (Z)-11-icosenylisobutyrate is very rare, only appearing in samples with the highest concentration of (Z)-11-icosenypropionate. This makes sense because the latter is built upon the former (Blomquist & Vogt, 2003): upon the acetate base, a new carbon unit is added forming a propionate, to which a new carbon unit is added to form the butyrate. This process might be performed by the same enzyme.

Cluster G: minor alkenes. This cluster consists in ?-icosene and the poly-unsaturated ??-henicosadiene (*Fig. 31*). The absence of information on the double bonds' positions makes it difficult to tell what the relationship between these two compounds is. They often appear

correlated with cluster A, and due to their basic carbon chain being longer than C18, it is likely that they both derive from the desaturation of stearic acid into (Z)-9-oleic acid by a Δ9-desaturase just like all the other alkenes. But the second double bond of ??-henicosadiene must be introduced by a different enzyme, and in the possibility that ?-icosene may not be derived from the same Δ9-desaturation event as alkenes from cluster A, one may hypothesize that its double bond is

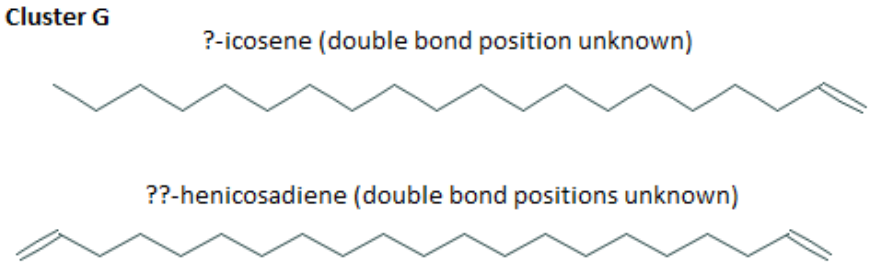


Figure 31. Cluster G. The positions of the double bonds remain unknown, the ones represented here are just some of the possibilities.

introduced by the same FAD that confers ??-henicosadiene its second double bond. Thus, important information lies in the double bond's position and the impossibility to identify it in this study means that this cluster cannot be fully explained. Since some noctuid moths like *Achaea janata* use (Z6,Z9)-6,9-henicosadiene (Francke & Schulz, 2010), it is plausible that this double bond would be located on carbon 6.

5. Conclusions.

Overall, determining the genetic architecture of the *H. pardalinus* and *H. elevatus* pheromone blend from this preliminary phenotypic analysis is difficult. However patterns are visible within the F2 which help with resolving the biosynthetic machinery of these butterflies. Certainly the activity of FADs and FARs represents the main difference. At least one FAD, a Δ9-desaturase, had its activity upregulated in *H. pardalinus.* Both species share 2-3 FARs with different specificities, but they seem to be more active in alkane production in *H. elevatus*- this is isn't however necessarily due to actual differences in the activity of FARs. Rather, due to the fact that its FADs are very downregulated compared to *H. pardalinus, H. elevatus* may simply have more saturated substrate for the FARs to act on and produce alkanes.

H. pardalinus is able to produce esters, both saturated and unsaturated, yet the lack of correlation between all of them suggests that they are given by different mechanisms. It is interesting that the production of HHFA may promote ester biosynthesis by releasing acetates- it means that acquiring the ability to sequester and transform plant compounds may promote the development of novel compounds for the pheromone blend. This is not necessarily the case however, since the aromatic plant derivatives in cluster D appear to have no effect on anything else.

The F2 pheromone results suggest that male sex pheromone differences between *H. pardalinus* and *H. elevatus* are controlled by 7 unlinked genomic regions each corresponding to the 7 clusters discussed above. *H. pardalinus* and *H. elevatus* are very closely related sister species, and genomic data suggest that current gene flow between the two is rampant across ~95% of their genomes (K. Dasmahapatra pers. Comm). Under the "islands of divergence" scenario of speciation with gene flow, genes responsible for reproductive isolation between taxa are expect to be clustered in the genome. However, the preliminary insight into the genomic architecture of the pheromone differences obtained here suggest that there is no such clustering of reproductive barrier genes between these two taxa.

Overall, the correlation method revealed a large amount of information on the *H. pardalinus* and *H. elevatus* pheromone biosynthetic machinery and paved the way to many more potential studies to further explore the topic. For example, it is unlikely that all analyzed compounds actually have an active role as sex pheromones in *H. pardalinus* and *H. elevatus.* Behavioural experiments and analysis of olfactory stimulation via an electroantennographic assay may reveal which ones are the most important. Experiments on incorporation of deuterium-labelled FA precursors would reveal more about the biosynthetic pathways themselves. QTL mapping of these compounds may help validate the hypotheses outlined above about each of the 7 clusters.

Chapter 4

Concluding remarks

The study of speciation is closely dependent on the study of the genetic mechanisms underlying species divergence, especially when it comes to the elusive sympatric speciation. *Heliconius pardalinus* and *Heliconius elevatus* are an extremely useful system for researching the genetics of species divergence in sympatry. A genomic island of divergence that originated from an adaptive introgression event with *H. melpomene* is observable in the *H. elevatus* genome, centered around a colour locus homologous to *H. melpomene'*s B/D (*Heliconius* Genome Consortium, 2012). Moreover, the colour pattern itself acts as a magic trait, since mate preference maps to the same region of the genome in *H. melpomene* (Merrill et al., 2011). This means that in a simple system of two sister species it is possible to observe the effects of sympatric speciation and three of the main phenomena associated with it: adaptive introgression, the formation of islands of divergence, and magic traits. Yet the silvaniform clade, to which they belong, remains poorly studied in comparison to the *melpomene-cydno* clade. The aim of this study was to resolve the genetic structure of two features of *H. elevatus* and *H. pardalinus,* both important in mating isolation. One is the colour pattern, a topic that has been studied in *Heliconius* for more than a century (Merrill et al., 2015), which is already known to be a magic trait. The other, suggested by the expansion of chemosensory and olfactory gene families in *Heliconius* (*Heliconius* Genome Consortium, 2012), is the male sex pheromone blend, a topic that in contrast to the colour pattern is almost completely unexplored.

In Chapter 2 I focused on the genetic structure of the colour pattern. Even with a lack of genomic data, the phenotype can be used to infer how many loci are involved in colour pattern formation and whether they are linked. It appears that these two species' colour pattern is controlled by the action of 5-6 large effect loci that control the general appearance of orange and white spots, and several QTLs associated with melanic spots. Importantly, the loci associated with orange elements of the pattern and those associated with white/yellow elements appear to belong to two separate linkage groups. If the black QTLs do in fact map to a third linkage group, this means the genetic organization of colour pattern loci in *H. pardalinus* and *H. elevatus* is not only similar to their close

relatives *H. hecale* and *H. ismenius,* but also to *H. melpomene* (Huber et al., 2015). Aside from the main pattern elements, the white invasion phenotype is particularly intriguing because it may offer some insight on the control mechanism for colour pattern formation. It appears that this locus has been silenced in the parental *H. elevatus* population used for this study, and hybridization removed the inhibitory mechanism.

The investigation of pheromones detailed in Chapter 3 did not yield such clear results, as no compound's production appeared controlled by a Mendelian locus. However an impressive amount of information could be derived from the species' phenotype, and especially from correlations between compounds. The *pardalinus*-specific upregulation of a Δ9-desaturase and the *elevatus*-specific upregulation of at least 3 fatty acid reductases were observed, as well as the ability, in *H. pardalinus,* to sequester plant compounds and use their derivatives as pheromones. The fact that one such plant derivative (HHFA) appears to affect the production of esters may hint at the fact that a single change in the pheromone biosynthetic network can affect the production of multiple compounds. As the first time a *Heliconius* sex pheromone blend has been investigated at this depth, this study posed several problems that need to be addressed as pheromones become the focus of more *Heliconius*-centered studies, namely the best way to normalize such a complex dataset while minimizing loss of information. The pheromone study also served as an important basis for future research, potentially focusing on behavioural responses or on investigating the putative biosynthetic mechanisms via deuterium-labelling. Both the pheromone study and the colour pattern one need to be complemented by QTL mapping to investigate the actual genetic structure of these traits, to verify the accuracy of inferences based on the phenotype, and most importantly to check whether any of these traits maps to known islands of divergence. When this data will be joined with data on other potential prezygotic barriers such as host plant preference, wing shape and flight dynamics, the result will be a powerful analysis of divergence in the *H. pardalinus-H. elevatus* system that will advance our general understanding of the mechanisms behind sympatric speciation and speciation with gene flow.

References.

Ando, T., & Yamakawa, R. (2011). Analyses of Lepidopteran sex pheromones by mass spectrometry. *TrAC Trends in Analytical Chemistry*, *30*(7), 990–1002.

Bates, H. W. (1862). XXXII. Contributions to an insect fauna of the Amazon valley. Lepidoptera: Heliconidae. *Transactions of the Linnean Society of London*, *23*(3), 495–566.

Baxter, S. W., Papa, R., Chamberlain, N., Humphray, S. J., Joron, M., Morrison, C., Jiggins, C. D. (2008). Convergent evolution in the genetic basis of Müllerian mimicry in *Heliconius* butterflies. *Genetics*, *180*(3), 1567-77.

Bjostad, L. B., & Roelofs, W. L. (1981). Sex pheromone biosynthesis from radiolabeled fatty acids in the redbanded leafroller moth. *The Journal of Biological Chemistry*, *256*(15), 7936–40.

Bjostad, L. B., & Roelofs, W. L (1983). Sex pheromone biosynthesis in *Trichoplusia ni*: key steps involve delta-11 desaturation and chain-shortening. *Science*, *220*(4604), 1387-9.

Blomquist, G. J., & Vogt, R. (2003). *Insect pheromone biochemistry and molecular biology : the biosynthesis and detection of pheromones and plant volatiles*. Elsevier/Academic Press.

Bolnick, D. I., & Fitzpatrick, B. M. (2007). Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*, *38*(1), 459–487.

Bossert, W. H., & Wilson, E. O. (1963). The analysis of olfactory communication among animals. *Journal of Theoretical Biology*, *5*(3), 443–469.

Butenandt, A., Beckmann, R., & Hecker, E. (1961). Über den Sexuallockstoff des Seidenspinners, I. Der biologische Test und die Isolierung des reinen Sexuallockstoffes Bombykol. *Hoppe-Seyler´s Zeitschrift Für Physiologische Chemie*, *324*(Jahresband), 71–83.

Chacón, M. G., Fournier, A. E., Tran, F., Dittrich-Domergue, F., Pulsifer, I. P., Domergue, F., & Rowland, O. (2013). Identification of amino acids conferring chain length substrate specificities on fatty alcohol-forming reductases FAR5 and FAR8 from Arabidopsis thaliana. *The Journal of Biological Chemistry*, *288*(42), 30345–55.

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland: Sinauer Associates, Inc.

Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., Nadeau, N. J., Jiggins, C. D. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487,* 94-98.

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, *491,* 756-760.

Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., & Holzapfel, C. M. (2010). Resolving postglacial phylogeography using high-throughput sequencing. *PNAS*, *107*(37), 16196–200.

Estrada, C., Schulz, S., Yildizhan, S., & Gilbert, L. E. (2011). Sex-specific chemical cues from immatures facilitate the evolution of mate guarding in *Heliconius* butterflies. S. *Evolution*, *65*(10), 2843–2854.

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*(7), 342–350.

Feder, J. L., Berlocher, S. H., Roethele, J. B., Dambroski, H., Smith, J. J., Perry, W. L., Aluja, M. (2003). Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(18), 10314–10319.

Feder, J. L., Gejji, R., Yeaman, S., Nosil, P., Coyne, J. A., Orr, H. A., Kim, J. F. (2012). Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1587), 461–474.

Fitzpatrick, B. M., Fordyce, J. A., & Gavrilets, S. (2008). What, if anything, is sympatric speciation? *Journal of Evolutionary Biology*, *21*(6), 1452–9.

Fordyce, J. A., Nice, C. C., Forister, M. L., & Shapiro, A. M. (2002). The significance of wing pattern diversity in the Lycaenidae: mate discrimination by two recently diverged species. *Journal of Evolutionary Biology*, *15*(5), 871–879.

Francke, W., & Schulz, S. (2010). 4.04 – Pheromones of terrestrial invertebrates. In Mander L & Liu H (Eds.), *Comprehensive Natural Products II* (pp. 153–223).

Gilbert, L. (1976). Postmating female odor in *Heliconius* butterflies: a male-contributed antiaphrodisiac? *Science*, *193*(4251), 419–420.

Hagström, Å. K., Liénard, M. A., Groot, A. T., Hedenström, E., Löfstedt, C., Bardwell, L., Wang, J. (2012). Semi–selective fatty acyl reductases from four Heliothine moths influence the specific pheromone composition. *PLoS ONE*, *7*(5), e37230.

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *The Journal of Heredity*, (S1), 795–809.

Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, *22*(18), 4606–4618.

Howard, R. W., & Blomquist, G. J. (1982). Chemical ecology and biochemistry of insect hydrocarbons. *Annual Review of Entomology*, *27*(1), 149–172.

Howard, R. W., & Blomquist, G. J. (2005). Ecological, behavioural, and biochemical aspects of insect hydrocarbons. *Annual Review of Entomology*, *50*(1), 371–393.

Huber, B., Whibley, A., Poul, Y., Navarro, N., Martin, A., Baxter, S., Joron, M. (2015). Conservatism and novelty in the genetic architecture of adaptation in *Heliconius* butterflies. *Heredity*, *114*, 515–524.

Jiggins, C. D., Naisbit, R. E., Coe, R. L., & Mallet, J. (2001). Reproductive isolation caused by colour pattern mimicry. *Nature*, *411*(6835), 302–305.

Jiggins, C. D. (2008). Ecological speciation in mimetic butterflies. *BioScience*, *58*(6), 541–548.

Jiggins, C. D., Mavarez, J., Beltrán, M., McMillan, W. O., Johnston, J. S., & Bermingham, E. (2005). A genetic linkage map of the mimetic butterfly *Heliconius melpomene*. *Genetics*, *171*(2), 557–570.

Jiggins, C. D., Salazar, C., Linares, M., & Mavarez, J. (2008). Hybrid trait speciation and *Heliconius* butterflies. *Phil. Trans. R. Soc. B*, *363*(1506), 3047-54.

Jiggins, C. D., Wallbank, R. W., Hanly, J. J. (2017). Waiting in the wings: what can we learn about gene co-option from the diversification of butterfly wing patterns? *Phil. Trans. R. Soc. B, 372*(1713):20150485.

Jones, R. T., Salazar, P. A., Ffrench-Constant, R. H., Jiggins, C. D., & Joron, M. (2012). Evolution of a mimicry supergene from a multilocus architecture. *Proceedings of the Royal Society, B*, *279*(June 2011), 316–325.

Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., Constant, R. H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, *477*(7363), 203–206.

Joron, M., Papa, R., Beltrán, M., Chamberlain, N., Mavárez, J., Baxter, S., Jiggins, C. D. (2006). A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biology*, *4*(10), e303.

Joron, M., Wynne, I. R., Lamas, G., & Mallet, J. (1999). Variable selection and the coexistence of multiple mimetic forms of the butterfly *Heliconius numata*. *Evolutionary Ecology*, *13*(7–8), 721–754.

Kawecki, T. J. (2004). Genetic theories of sympatric speciation. In U. Dieckmann, M. Doebeli, J. A. J. Metz, & D. Tautz (Eds.), *Adaptive Speciation.* (pp. 36–53). Cambridge: Cambridge University Press.

Knipple, D. C., Rosenfield, C.-L., Nielsen, R., You, K. M., & Jeong, S. E. (2002). Evolution of the integral membrane desaturase gene family in moths and flies. *Genetics*, *162*(4), 1737-52.

Kozak, K. M., Wahlberg, N., Neild, A. F. E., Dasmahapatra, K. K., Mallet, J., & Jiggins, C. D. (2015). Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius butterflies*. *Systematic Biology*, *64*(3), 505–524.

Kronforst, M. R. & Papa, R. (2015). Kronforst, M. R., & Papa, R. (2015). The Functional Basis of Wing Patterning in Heliconius butterflies. *Genetics*, *200*, 1–19.

Kronforst, M. R., Young, L. G., Kapan, D. D., McNeely, C., O'Neill, R. J., & Gilbert, L. E. (2006). Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(17), 6575–80.

Lassance, J.-M., Groot, A. T., Liénard, M. A., Antony, B., Borgwardt, C., Andersson, F., Löfstedt, C. (2010). Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. *Nature*, *466*(7305), 486–489.

Le Poul, Y., Whibley, A., Chouteau, M., Prunier, F., Llaurens, V., Joron, M. (2014). Evolution of dominance mechanisms at a butterfly mimicry supergene. *Nature Communications, 5*(5644).

Lewis, J. J., van der Burg, K. R. L., Mazo-Vargas, A., Reed, R. D., Ahola, V., Lehtonen, R., … Andolfatto, P. (2016). ChIP-seq-annotated *Heliconius erato* genome highlights patterns of cis-regulatory evolution in Lepidoptera. *Cell Reports*, *16*(11), 2855–2863.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Liénard, M. A., Hagström, A. K., Lassance, J.-M., & Löfstedt, C. (2010). Evolution of multicomponent pheromone signals in small ermine moths involves a single fatty-acyl reductase gene. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(24), 10955–60.

Liénard, M. A., Wang, H. L., Lassance, J.M., & Löfstedt, C. (2014). Sex pheromone biosynthetic pathways are conserved between moths and the butterfly *Bicyclus anynana*. *Nature Communications*, *5*(3957).

Liu, W., Rooney, A. P., Xue, B., & Roelofs, W. L. (2004). Desaturases from the spotted fireworm moth (*Choristoneura parallela*) shed light on the evolutionary origins of novel moth sex pheromone desaturases. *Gene*, *342*(2), 303–311.

Löfstedt, C. (1990). Population variation and genetic control of pheromone communication systems in moths. *Entomologia Experimentalis et Applicata*, *54*(3), 199–218.

Loscos, N., Hernandez-Orte, P., Cacho, J., & Ferreira, V. (2007). Release and formation of varietal aroma compounds during alcoholic fermentation from nonfloral grape odorless flavor precursors fractions. *Journal of Agricultural and Food Chemistry*, *55*(16), 6674–6684.

Mallet, J. (1989). The genetics of warning colour in Peruvian hybrid zones of *Heliconius erato* and *H. melpomene*. Proc. Roy. Soc. Lond. B 236: 163-185 (1989). *Proc. R. Soc. Lond. B*, *236*, 163–185.

Mallet, J. (2007). Hybrid speciation. *Nature*, *446*(7133), 279–283.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229–237.

Mallet, J., Beltrán, M., Neukirchen, W., Linares, M., Berlocher, S., Feder, J., Linares, M. (2007). Natural hybridization in Heliconiine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology*, *7*(28).

Mann F, Vanjari S, Morris J, Mann S, Dasmahapatra K, Linares M, Schulz S. (2015). *Scent compound metabolomics of Heliconius wing androconia. (Unpublished)*. Institute of Organic Chemistry, Technische Universität Braunschweig.

Martin, A., Papa, R., Nadeau, N. J., Hill, R. I., Counterman, B. A., Halder, G., Reed, R. D. (2012). Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(31), 12632–7.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, *23*(11), 1817–28.

Matsumoto, S. (2010). Molecular mechanisms underlying sex pheromone production in moths. *Bioscience, Biotechnology, and Biochemistry*, *74*(2), 223–231.

Mavárez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., & Linares, M. (2006). Speciation by hybridization in *Heliconius* butterflies. *Nature*, *441*(7095), 868–871

Mayr, E. (1962). *Animal species and evolution*. Harvard University Press, Harvard.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303

Mérot, C., Frérot, B., Leppik, E., & Joron, M. (2015). Beyond magic traits: multimodal mating cues in *Heliconius* butterflies. *Evolution*, *69*(11), 2891–2904.

Merrill, R. M., Dasmahapatra, K. K., Davey, J. W., Dell'Aglio, D. D., Hanly, J. J., Huber, B., … Yu, Q. (2015). The diversification of *Heliconius* butterflies: what have we learned in 150 years? *Journal of Evolutionary Biology*, *28*(8), 1417–38.

Merrill, R. M., Van Schooten, B., Scott, J. A., & Jiggins, C. D. (2011). Pervasive genetic associations between traits causing reproductive isolation in *Heliconius* butterflies. *Proceedings. Biological Sciences*, *278*(1705), 511–18.

Merrill, R. M., Wallbank, R. W. R., Bull, V., Salazar, P. C. A., Mallet, J., Stevens, M., & Jiggins, C. D. (2012). Disruptive ecological selection on a mating cue. *Proceedings. Biological Sciences*, *279*(1749), 4907–13.

Meyer, H. J., & Norris, D. M. (1967). Vanillin and syringaldehyde as attractants for *Scolytus multistriatus* (Coleoptera: Scolytidae). *Annals of the Entomological Society of America*, *60*(4), 858-859.

Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C., & Forejt, J. (2009). A mouse speciation gene encodes a meiotic histone H3-methyltransferase. *Science*, *323*(5912), 373-375.

Moser, J. C., Brownlee, R. C., & Silverstein, R. (1968). Alarm pheromones of the ant *Atta texana*. *Journal of Insect Physiology*, *14*(4), 529–535.

Nadeau, N. J., Ruiz, M., Salazar, P., Counterman, B., Medina, J. A., Ortiz-Zuazaga, H., Papa, R. (2014). Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene and H. erato*. *Genome Research*, *24*(8), 1316–1333.

Nadeau, N. J., Pardo-Diaz, C., Whibley, A., et al. (2016) The gene *cortex* controls mimicry and crypsis in butterflies and moths. *Nature, 534,* 106-110.

Naisbit, R. E., Jiggins, C. D., & Mallet, J. (2003). Mimicry: developmental genes that contribute to speciation. *Evolution and Development*, *5*(3), 269–280.

Nieberding, C. M., de Vos, H., Schneider, M. V., Lassance, J.-M., Estramil, N., Andersson, J., Brakefield, P. M. (2008). The male sex pheromone of the butterfly *Bicyclus anynana*: towards an evolutionary analysis. *PLoS ONE*, *3*(7), e2751.

Nijhout, H. F. (n.d.). The development and evolution of wing patterns. Retrieved October 30, 2016, from http://sites.biology.duke.edu/nijhout/patterns2.html

Nijhout, H. F. (1994). Developmental Perspectives on Evolution of Butterfly Mimicry. *BioScience*, *44*(3), 148–157.

Nijhout, H. F., Wray, G. A., & Gilbert, L. E. (1990). An analysis of the phenotypic effects of certain colour pattern genes in *Heliconius* (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society*, *40*(4), 357–372.

Nijhout, H. F. (2001). Elements of butterfly wing patterns. *Journal of Experimental Zoology*, *291*(3), 213–225.

Nishida, R., Schulz, S., Fukami, H., Kuwahara, Y., Honda, I. K., & Hayashi, N. (1996). Male sex pheromone of a giant Danaine butterfly, *Idea leuconoe*. *Journal of Chemical Ecology*, *22*(5), 949-972.

Noor, M., Feder, J. (2006). Speciation genetics: evolving approaches. *Nature Reviews Genetics, 7*(11), 851-861.

Norris, L. C., Main, B. J., Lee, Y., Collier, T. C., Fofana, A., Cornel, A. J., & Lanzaro, G. C. (2015). Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proceedings of the National Academy of Sciences*, *112*(3), 815-820.

Nosil, P., & Schluter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology & Evolution*, *26*, 160–167.

Orr, H. A., & Presgraves, D. C. (2000). Speciation by postzygotic isolation: forces, genes and molecules. *BioEssays*, *22*(12), 1085–1094.

Papa, R., Kapan, D. D., Counterman, B. A., Maldonado, K., Lindstrom, D. P., Reed, R. D., … Churchill, G. (2013). Multi-allelic major effect genes interact with minor effect QTLs to control adaptive color pattern variation in *Heliconius erato*. *PLoS ONE*, *8*(3), e57033.

Papa, R., Morrison, C. M., Walters, J. R., Counterman, B. A., Chen, R., Halder, G., … Gentleman, R. (2008). Highly conserved gene order and numerous novel repetitive elements in genomic regions linked to wing pattern variation in *Heliconius* butterflies. *BMC Genomics*, *9*(345).

Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., Jiggins, C. D. (2012). Adaptive introgression across species boundaries in Heliconius butterflies. *PLoS Genetics*, *8*(6).

Queste, L. M. (2015). *Characterising reproductive barriers between three closely related Heliconius butterfly taxa.* MSc by Research thesis, University of York.

Reed, R. D., Papa, R., Martin, A., Hines, H. M., Counterman, B. A., Pardo-Diaz, C., McMillan, W. O. (2011). optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science*, *333*(6046), 1137-41.

Regnier, F. E., & Law, J. H. (1968). Insect pheromones. *Journal of Lipid Research*, *9*(5), 541–551.

Rieseberg, L. H., Van Fossen, C., & Desrochers, A. M. (1995). Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature*, *375*(6529), 313–316

Rodríguez, S., Hao, G., Liu, W., Piña, B., Rooney, A. P., Camps, F., Fabriàs, G. (2004). Expression and evolution of Δ9 and Δ11 desaturase genes in the moth *Spodoptera littoralis*. *Insect Biochemistry and Molecular Biology*, *34*(12), 1315–1328.

Roelofs, W. L., Liu, W., Hao, G., Jiao, H., Rooney, A. P., & Linn, C. E. (2002). Evolution of moth sex pheromones via ancestral genes. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(21), 13621–13626.

Roelofs, W. L., & Rooney, A. P. (2003). Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(16), 9179-9184.

Rosser, N. (2012). *Speciation and biogeography of Heliconiine butterflies.* PhD thesis, University College London.

Rosser, N., Phillimore, A. B., Huertas, B., Willmott, K. R., & Mallet, J. (2012). Testing historical explanations for gradients in species richness in Heliconiine butterflies of tropical America. *Biological Journal of the Linnean Society*, *105*, 479–497.

Rundle, H., Nosil, P. (2005). Ecological speciation. *Ecology Letters, 8,* 336-352.

Schmid, H. H., & Bandi, P. C. (1969). n-Triacontanal and other long-chain aldehydes in the surface lipids of plants. *Hoppe-Seyler's Zeitschrift Fur Physiologische Chemie*, *350*(4), 462–466.

Schulz, S., Boppre, M., & Vane-Wright, R. I. (n.d.). Specific mixtures of secretions from male scent organs of African milkweed butterflies (Danainae). *Philosophical Transactions: Biological Sciences*, *342*(1300), 161–181.

Schulz, S., Estrada, C., Yildizhan, S., Boppré, M., & Gilbert, L. E. (2008). An antiaphrodisiac in *Heliconius melpomene* butterflies. *Journal of Chemical Ecology*, *34*(1), 82–93.

Schulz, S., & Van Loon, J. J. A. (2011). The biosynthesis of hexahydrofarnesylacetone in the butterfly *Pieris brassicae*. *Journal of Chemical Ecology*, *37*(4), 360–363.

Servedio, M. R., Doorn, G. S. Van, Kopp, M., Frame, A. M., & Nosil, P. (2011). Magic traits in speciation: "magic" but not rare? *Trends in Ecology & Evolution*, *26*(8), 389–397.

Shanklin, J., & Cahoon, E. B. (1998).Desaturation and related modifications of fatty acids. *Annual Review of Plant Physiology and Plant Molecular Biology*, *49*(1), 611–641.

Sheppard, P. M., Turner, J. R. G., Brown, K. S., Benson, W. W., & Singer, M. C. (1985). Genetics and the evolution of Mullerian mimicry in *Heliconius* butterflies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *308*(1137).

Song, Y., Endepols, S., Klemann, N., Richter, D., Matuschka, F.-R., Shih, C.-H., Kohn, M. H. (2011). Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Current Biology*, *21*(15), 1296–301.

Supple, M. A., Papa, R., Hines, H. M., McMillan, W. O., Counterman, B. A., Harrison, R., Rieseberg, L. (2015). Divergence with gene flow across a speciation continuum of *Heliconius* butterflies. *BMC Evolutionary Biology*, *15*(204).

Supple, M. A., Hines, H. M., Dasmahapatra, K. K., Lewis, J. J., Nielsen, D. M., Lavoie, C., Counterman, B. A. (2013). Genomic architecture of adaptive color pattern divergence and convergence in *Heliconius* butterflies. *Genome Research*, *23*(8), 1248–57.

Symonds, M. R. E., & Elgar, M. A. (2008). The evolution of pheromone diversity. *Trends in Ecology & Evolution*, *23*(4), 220–228.

Tillman, J. A., Seybold, S. J., Jurenka, R. A., & Blomquist, G. J. (1999). Insect pheromones—an overview of biosynthesis and endocrine regulation. *Insect Biochemistry and Molecular Biology*, *29*(6), 481–514.

Turner, J. R. G. (1966). A little-recognized species of *Heliconius* butterfly (Nymphalidae). *Journal of Research on the Lepidoptera*, *5*, 97–112.

Turner, J. R. G., & Crane, J. (1962). The genetics of some polymorphic forms in *Heliconius melpomene* Linnaeus and *Heliconius erato* Linnaeus. I. Major genes. *Zoologica*, (47), 141–152.

Vanjari, S., Mann, F., Merrill, R. M., Schulz, S., & Jiggins, C. D. (2015). Male sex pheromone components in the butterfly *Heliconius melpomene*. *bioRxiv*.

Via, S., Rice, W. R., Hostert, E. E., Via, S., Coyne, J. A., Orr, H. A., … Donnelly, P. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1587), 451–60.

Voet, D., & Voet, J. G. (2011). *Biochemistry*. New York: John Wiley & Sons.

Wallbank, R. W. R., Baxter, S. W., Pardo-Diaz, C., Hanly, J. J., Martin, S. H., Mallet, J., … Gernhard, T. (2016). Evolutionary novelty in a butterfly wing pattern through enhancer shuffling. *PLOS Biology*, *14*(1), e1002353.

Wu, C.-I., & Ting, C.-T. (2004). Genes and speciation. *Nature Reviews Genetics*, *5*(2), 114–122.

Zhang, W., Dasmahapatra, K. K., Mallet, J., Moreira, G. R. P., & Kronforst, M. R. (2016). Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biology*, *17*(25)