



# From wing pattern genes to the chemistry of speciation : an integrative dissection of the early stages of diversification in mimetic butterflies

Bárbara Huber

## ► To cite this version:

Bárbara Huber. From wing pattern genes to the chemistry of speciation : an integrative dissection of the early stages of diversification in mimetic butterflies. Animal biology. École pratique des hautes études - EPHE PARIS, 2015. English. NNT : 2015EPHE3057 . tel-02099663

**HAL Id: tel-02099663**

**<https://theses.hal.science/tel-02099663>**

Submitted on 15 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



École Pratique  
des Hautes Études

**École Pratique des Hautes Études**

Mention « Systèmes intégrés, environnement et biodiversité »

École doctorale de l'École Pratique des Hautes Études

*Laboratoire Biologie Intégrative des Populations*

---

**From wing pattern genes to the chemistry of  
speciation : an integrative dissection of the early stages  
of diversification in mimetic butterflies**

---

présentée par

**Bárbara HUBER**

THÈSE DE DOCTORAT DE BIOLOGIE

Sous la direction de :

**M. Thierry WIRTH**

Directeur d'études

Soutenue le 25 novembre 2015 devant le jury composé de :

Mme. Claude WICKER-THOMAS	DR	CNRS	Rapporteur
M. Walter SALZBURGER	Prof.	Université de Bale	Rapporteur
Mme. Sophie MONTUIRE	DE2	EPHE	Examinatrice
M. Mathieu JORON	DR	CNRS	Co-encadrant
M. Thierry WIRTH	DE1	EPHE	Directeur
M. Nicola NADEAU	DR	University of Sheffield	Examinatrice





*Para mis tres amores, mamá, papá y Ana*  
*Para mis dos hombres, Cristóbal y Pablo*

## Remerciements

*Gracias a la vida, que me ha dado tanto...*

After all the support I have got to arrive to the end of this long way, I can barely say that what I am finishing here is *my* PhD. I learnt that science is and should be a communal work, and that is precisely what this PhD is: a *minga*. I would therefore like to thank so many people...

I must start by thanking Mathieu for offering me a very nice subject and the possibility to work in his team, for supporting me and our project unconditionally, for his patience and generosity during biological discussions, and especially for trusting me and my abilities (sometimes more than myself) to explore and conduct new projects. Thanks as well for investing a portion of your grants in this project.

I am thankful to Thierry Wirth for accepting being my supervisor, and to the EPHE and its people for providing a formal ground to this work and for the travel allowances.

I was so lucky to have fallen in a Lab harbouring so precious and welcoming people. Thank you Yann for sitting back-to-back with me day after day during these years, slowly weaving a precious companionship. Violaine, Lise, Florence, Marianne, Nicolas, Rob, all of you have been so wonderful companions! I am most thankful to Annabel for her great advice and unlimited help, and for always having the right word. Thank you Claire and Mónica for your friendship, your smiles and your optimism.

I wish to thank warmly Walter Salzburger and Claude Wicker-Thomas for taking on the work of carrying out such a careful and *in record time* reading of this thesis. I would like to thank Nicola Nadeau and Sophie Montuire for accepting being a part of my dissertation defense panel. Also, thanks to Virginie Orgogozo, Guillaume Achaz, Brigitte Frérot and Stefano Mona for giving good advice at my *comité de thèse*. Thank you Christophe Haag for that good week of work on a project that we still need to accomplish, but that taught me a lot.

I would like to thank the *Heliconius* Consortium for strengthening this feeling of laboring in a real community. Special thanks to John Davey, Richard Merrill, Simon Martin, Chris Jiggins and the rest of their team in Cambridge. Also, I am thankful to MD Sharma and to Jenny Barna for handling the clusters so efficiently.

I am deeply thankful to STRI in Panama for providing the perfect working ambiance during my field work. I spent a whole year in Gamboa and I had a beautiful time, thanks to the warmth of Adriana and Liz Liz Liz, to the supervision and kindness of Owen, to the support of Moisés, and to the fruitful dialogues with Catalina. I am happy to have shared this work with three interns in Gamboa (Darha Solano, Benjamin Rice and Robert Orpet) and one in Paris (Valentin Rineau). I thank them for their help and great ideas. I would also like to thank Andrés Orellana, Patricio Salazar, Nicola Nadeau and Tim Thurman for collecting *Very Important Butterflies*.

Thanks to Brigitte Frérot for expanding my horizons towards a topic I am now most interested in. Thanks also to Ene and Centina for helping me running tens of samples at their lab in Versailles.

At the MNHN I am thankful to Alice for always solving my bureaucratic problems and to Monsieur Gilles for cheering us up every day. Thank you Simone and Axelle for sharing refreshing moments.

I wish to thank those who accompanied us during these years and who made the weight of the thesis lighter and the harshness of living in a foreign land kinder. These include our friends from *El Trío*. Anne-Lise, you have been a light to our soul. Claudia, you have been a source of strength. I also thank you, Lucie and Jean, for your friendship.

Abroad the frontiers, there was my godmother Evelyne and my family and friends in Switzerland, with whom special moments along these years made me feel like at home. Thank you Juan and Judi for treating me as family. Thank you Violeta and Antoine, for so much tenderness. Thanks to Antoine's family for lending us a place where to stay in Paris. Thank you Gaby, Celeste, Gerard, Francisco, Raquel, Lirey, Jean Pierre, for always being there. I want to thank the uncles and aunts of Cristóbal for their hospitality in distinct parts of the globe. I am infinitely thankful to Olga and Ramón for their unconditional care and encouragement this last year.

Mamá, papá and Ana, you are of course to whom I owe the most, which is why this achievement is yours. And you, Cristóbal, I will never find the right words to express my gratitude to life for putting you in my way. Thanks for walking this way together, hand in hand, for sharing the joys and the sorrows of this adventure, for taking care of me, for our Pablo, who filled our lives with joy and gave me the last push. Thanks, Cristóbal, for filling me with peace *cuando miro al fondo de tus ojos claros...*

## Résumés

### **From wing pattern genes to the chemistry of speciation: an integrative dissection of the early stages of diversification in mimetic butterflies**

#### Abstract

How does biological diversification occur in the face of genetic exchange? How do reproductive barriers evolve and function? What is the role of adaptive traits in promoting diversification and speciation? These open questions in evolutionary biology are at the core of this project. In order to tackle them, we have focused on butterflies in the neo-tropical genus which are an important component of the diverse butterfly communities in the Neo-tropics. Butterflies in the genus *Heliconius* are unpalatable to predators, use warning colours to advertise their defences, and mimic other defended butterflies in their local communities. The genus has undergone an adaptive radiation in wing colour patterns as a result of natural selection for mimicry, and is also well known for assortative mating based on wing pattern. I have extended the current knowledge about the ecological function and the genetic basis of wing color patterns in these butterflies and explored the importance of wing coloration relative to chemical signaling in the early stages of diversification. To this aim, I have characterised the adaptive divergence between lineages at different stages of the speciation continuum, by integrating genomic, phenotypic, behavioural, chemical and ecological data. More precisely, I have studied the so-called silvaniform sub-clade of *Heliconius*, known for harbouring species with tiger patterns that participate in mimicry with large groups of other closely and distantly-related species. My work includes the comparative description of the genetic architecture of wing pattern adaptation in two species, *H. hecale* and *H. ismenius*, using crosses, genome-wide next-generation genotyping, and advanced morphometrics of colour patterns. I have also explored the importance of natural and sexual selection on wing-patterning loci at early stages of divergence in the genus. In particular, I have analysed the structure and maintenance of a hybrid zone between two distinctly coloured parapatric races of *H. hecale* by using a combination of population genetics and genomics, coupled to a phenotypic analysis of the clines and to behavioural assays on male-based mate choice. Finally, I have carried out genome-wide analyses of divergence and gene flow with whole genome sequencing data to look for evidence of introgression between coexisting, hybridising co-mimetic species. This was again coupled to experiments on mating preferences and behavior,

and yielded evidence for important differences in putative pheromone signals which may mediate species recognition and the maintenance of species boundaries. Overall, my results show that although selection on wing pattern divergence have been central to the diversification of the genus *Heliconius*, the accumulation of other barriers to gene flow may be important for the speciation process to be completed.

### Keywords

*Heliconius*, butterflies, reproductive isolation, speciation continuum, genomics, pheromones, genetic architecture

## Une étude intégrative des stades précoces de l'isolement reproducteur chez les papillons *Heliconius*

### Résumé

Comment la diversification biologique peut-elle avoir lieu malgré les échanges génétiques ? Comment les barrières reproductives entre espèces évoluent-elles et fonctionnent-elles ? Les changements adaptatifs de certains traits favorisent-ils la diversification et la spéciation ? Ces questions ouvertes en biologie évolutive constituent la base de ce projet. Pour y répondre, nous nous sommes intéressés aux papillons du genre néo-tropical *Heliconius* qui constituent une partie importante des communautés diversifiées de papillons néotropicaux. Les papillons de ce groupe sont immangeables pour les prédateurs, arborent des colorations d'avertissement qui signalent leur toxicité, et miment d'autres papillons toxiques dans leurs communautés locales. Ce genre a connu une radiation adaptative des motifs colorés soumis à la sélection naturelle favorisant le mimétisme de divers signaux locaux, mais ces motifs sont également connus comme signaux intraspécifiques favorisant les appariements homogames. Mes travaux ont permis d'approfondir les connaissances actuelles sur la fonction écologique et la base génétique de la couleur des ailes chez ces papillons, et d'explorer l'importance de la couleur des ailes par rapport aux signaux chimiques au cours des premières étapes de diversification. Dans cette optique, j'ai caractérisé la divergence adaptative entre les taxons à différents stades du continuum de spéciation, par une approche intégrative combinant des données génomiques, phénotypiques, comportementales, chimiques et écologiques. Plus précisément, j'ai étudié le sous-clade de *Heliconius* appelé sylvaniformes, contenant des espèces de papillons aux motifs tigrés, qui participent à des relations de mimétisme avec de nombreuses autres espèces fortement apparentées ou non. Mes travaux incluent la description comparative de l'architecture génétique des motifs colorés adaptatifs parallèlement chez les espèces *H. hecale* et *H. ismenius*, en utilisant des croisements, du génotypage génomique à haut débit, et de la morphométrie des motifs colorés. J'ai également exploré l'importance de la sélection naturelle et sexuelle sur les locus contrôlant ces motifs colorés aux stades précoces de divergence dans ce genre. En particulier, j'ai analysé la structure et le maintien de la zone d'hybridation entre deux races parapatricques de *H. hecale* montrant des colorations différentes, en combinant la génétique et la génomique des populations, ainsi que l'analyse phénotypique de clines et des tests comportementaux sur le choix de partenaire chez les mâles. Enfin, j'ai effectué des analyses génomiques de la divergence et du flux de gènes en me basant sur des données de re-séquençage de génomes complets afin de rechercher des traces d'introgession entre des espèces co-mimétiques et étroitement apparentées. Ceci a été également couplé à des expériences de préférence et de comportement sexuel, ainsi qu'à des analyses chimiques montrant d'importantes différences dans des composés qui pourraient intervenir dans la reconnaissance spécifique et le maintien des limites entre espèces. Dans l'ensemble, mes travaux montrent que bien que la sélection agissant sur les motifs colorés des ailes ait été centrale dans la diversification du genre *Heliconius*, l'accumulation d'autres barrières au flux de gènes peut jouer un rôle important dans l'aboutissement du processus de spéciation.

## Mots clés

*Heliconius*, papillons, isolement reproducteur, continuum de spéciation, génomique, phéromones, architecture génétique



## List of publications

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

Merrill RM, Dasmahapatra KK, Davey JW, Dell’Aglia DD, Hanly JJ, **Huber B**, *et al.* (2015). The diversification of *Heliconius* butterflies : what have we learned in 150 years ? *J Evol Biol* : **28** : 1417–1438.

# Contents

List of publications	10
Introduction	13
0.1 How do discrete biological entities arise? . . . . .	14
0.2 How does hybridisation affect diversification? . . . . .	31
0.3 What is the genetic architecture of adaptation? . . . . .	37
0.4 How does our work contribute to answer largely unresolved questions? . . . . .	47
1 Conservatism and novelty in the genetic architecture of adaptation	54
1.1 Introduction . . . . .	55
1.2 Materials and Methods . . . . .	57
1.3 Results . . . . .	60
1.4 Discussion . . . . .	66
1.5 Supplementary information . . . . .	71
2 The hybrid zone between two races of <i>Heliconius hecale</i>	104
2.1 Introduction . . . . .	105
2.2 Materials and Methods . . . . .	107
2.3 Results . . . . .	111
2.4 Discussion . . . . .	116
2.5 Supplementary information . . . . .	124
3 Coexisting closely-related species under the same disguise	129
3.1 Introduction . . . . .	130
3.2 Materials and methods . . . . .	132
3.3 Results . . . . .	138
3.4 Discussion . . . . .	148
3.5 Supplementary information . . . . .	159
4 Concluding remarks	171
4.1 It's all about selection . . . . .	171
4.2 Powerful, but not almighty . . . . .	174
4.3 Pervasive hybridisation is not the rule in the genus . . . . .	175
A Résumé français	176
A.1 Introduction . . . . .	176
A.2 Chapitre 1. Du conservatisme et de la nouveauté dans l'architecture génétique de l'adaptation . . . . .	179

---

A.3	Chapitre 2. La zone d'hybridation entre deux races de <i>Heliconius hecale</i> . . . .	181
A.4	Chapitre 3. Des espèces apparentées et sympatriques avec le même déguisement	184
A.5	Conclusions . . . . .	187
	List of Tables	193
	List of Figures	194
	Bibliography	196

# Introduction

## Foreword

Organisms display striking variation in phenotypic morphology and function, both of which are fundamental to biological diversification. How did this remarkable diversity arise is one of the central questions in evolutionary biology. More specifically, what are the forces driving the diversity of colours, shapes, behaviours and other traits we see in nature? How does evolution shape these traits through the selection of “types” adapted to particular environments? And finally, how does adaptation to particular environmental conditions originate the discrete entities we call species? Attempts to answer these questions have resulted from a fruitful combination of models with empirical data. Thanks to the large availability of data, the study of evolution has now become highly integrative and has focused on exploring how genes, organisms and environments interact for diversification to proceed. However, more than 150 years after the publication of *The Origin of Species* ([Darwin, 1859]), much remains to be understood about the factors and mechanisms contributing to evolution.

To solve open questions in evolutionary biology, attention has often been put on hotspots of biodiversity, since these enclose several taxonomic groups that have undergone adaptive radiations and are therefore perfect scenarios to disentangle the mechanisms responsible for diversification. The Neo-tropics are listed as one of the most biodiverse regions in the world. Thus, it is not surprising that the complexity of neo-tropical species communities inspired several naturalists to describe diverse phenomena and developing theories linked to evolutionary biology. The independent works of Henry Walter Bates and Fritz Müller in the late 1800s stand out in this respect ([Bates, 1862]; [Müller, 1879]). They explored communities of neo-tropical butterflies and observed that groups of species shared strikingly similar wing colour patterns, a resemblance that could hardly be explained by chance. Alternatively, these authors keyed out an explanation that was in consonance with Darwin’s precepts. Wing motifs were suggested to be involved in warning signalling to predators (in particular to birds), thus having a protective function. Mimicry, which is the adaptive resemblance in signal between several species in a locality, could consist either of edible species mimicking distasteful models (Batesian mimicry) or of several distasteful species displaying a common pattern and thus benefiting from predators learning about their toxicity (Müllerian mimicry; see Figure 1 for a visual description of the phenomenon). Mimicry, which has been observed in a varied array of animal taxa including insects, frogs, snakes and millipedes ([Brown, 1981]; [Pfennig et al., 2001]; [Marek and Bond, 2009]; [Alexandrou et al., 2011]; [Twomey et al., 2013]), drives the adaptation of

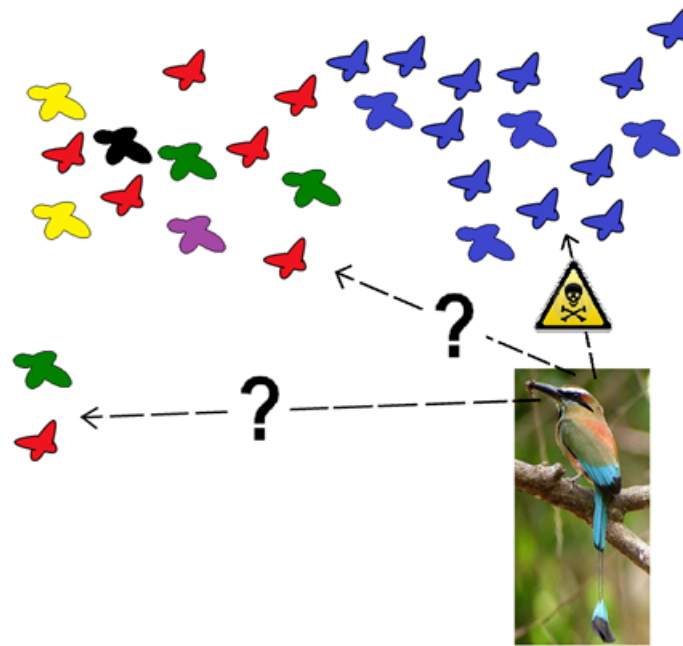


Figure 1 – **Schematic representation of Müllerian mimicry.** Different unpalatable species share the same conspicuous pattern (in this case the blue colour) which is distinguished from other patterns and associated to toxicity when memorized by predators.

lineages to “mimetic environments” known as mimicry rings (Figure 2).

Butterflies in the genus *Heliconius* (tribe Heliconiini) are a classical example of Müllerian mimicry. These butterflies, which are unpalatable to predators, are famous for their high diversity of wing colour patterns, which warn predators of their toxicity. The multiple and variable geographic races of the ~40 species in this clade show mimetic relationships with other co-occurring species in a given geographic region. This phenotypic convergence for mimicry is subjected to very strong selective pressures. Moreover, wing colour attributes are also implicated in mate recognition, which strengthens their importance as reproductive isolation barriers. Hence, *Heliconius* butterflies represent a great model for the study of speciation, providing a unique opportunity to understand how both convergent and diversifying evolution processes have taken place. They represent the natural laboratory in which we address questions related to phenotypic diversification and speciation.

### 0.1 How do discrete biological entities arise?

*There is hardly an ecological factor that does not affect speciation directly or indirectly, actually or potentially.*

Mayr 1963, *Animal Species and Evolution*, p. 556

Understanding how populations evolve differences, how species originate and how their distinctiveness is maintained are central questions in evolutionary biology. In

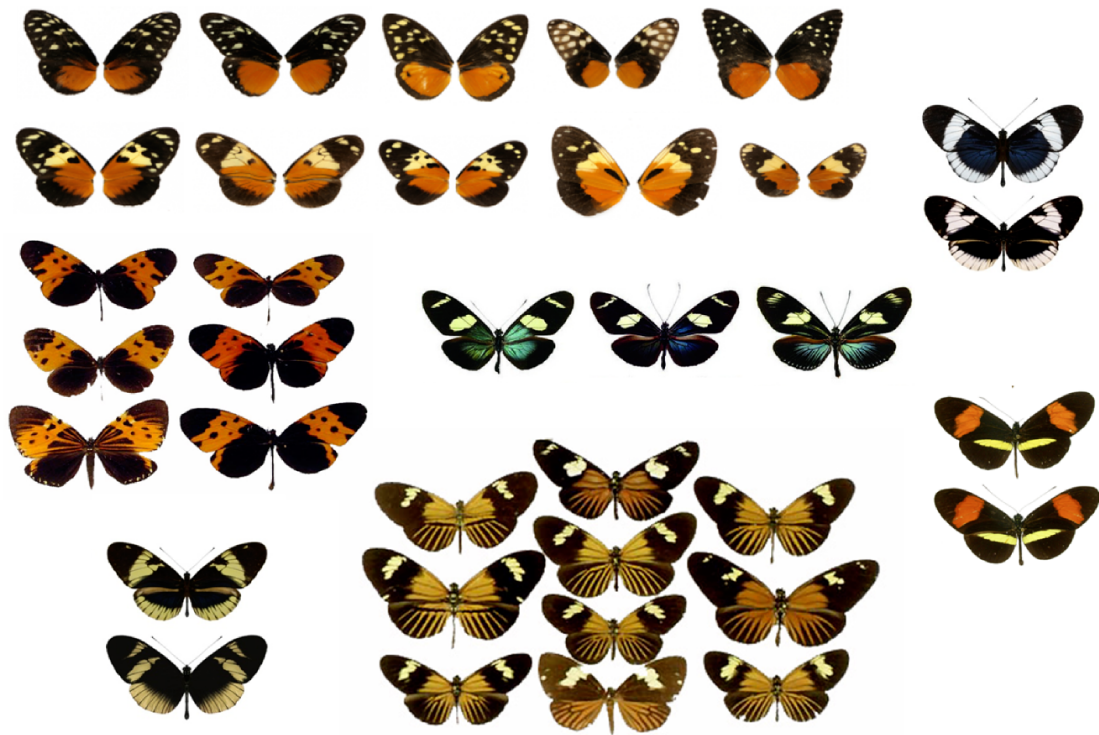


Figure 2 – Examples of mimicry rings formed by several toxic neo-tropical butterfly species, including butterflies in the genus *Heliconius*.

a context of habitat loss, climate change, species invasions, resistance to insecticides and antibiotics, among other anthropogenic disturbances, elucidating the evolutionary mechanisms of adaptive divergence might help in predicting the response of organisms to global changes.

Referring to “discrete biological entities” is here an alternative way to say “species”. Nowadays, the biological species concept introduced in [Mayr, 1942] is highly debated. According to Mayr, species are groups of interbreeding natural populations that are reproductively isolated from other such groups. Here, I adopt an extension of this concept that allows for some extent of permeability in the species boundaries ([Coyne and Orr, 2004]). Despite this permeability, species still constitute discrete categories that can be distinguished on a morphological, genetic, ecological or behavioural basis. “Discrete” here does therefore not imply complete reproductive isolation, but rather the maintenance of species-related identity in the face of potential or actual gene flow. This is the idea behind the genotypic cluster species concept, which defines species as distinguishable clusters of genotypes that have few or no intermediates when in contact with each other ([Mallet and Gilbert, 1995]). Although ecotypes or subspecies (i.e. closely-related taxa below the species level) are also discrete in some respects, they are generally highly continuous genetically. Additionally, their distinctiveness may lack permanence and may be more easily reversed than species identity, following a change in the pattern of divergent selection ([Coyne and Orr, 2004]).

The process by which discrete entities originate is called speciation. Nevertheless,

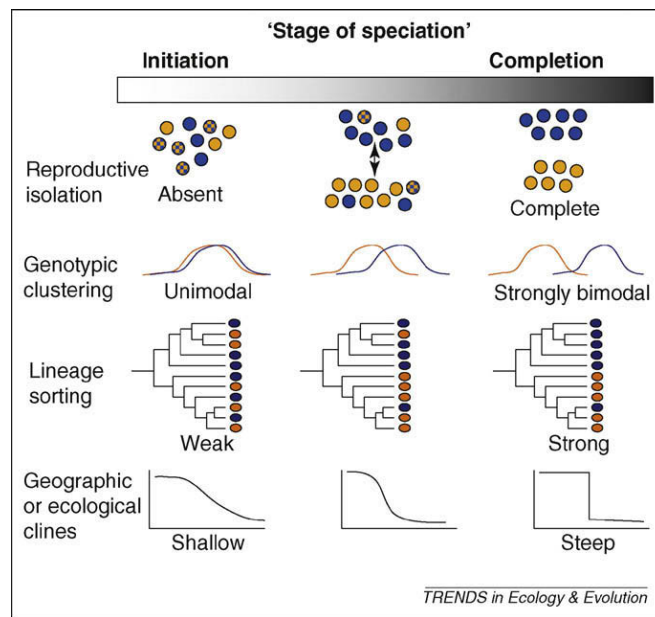


Figure 3 – **The continuous nature of divergence during speciation.** Different means of quantifying divergence can be used to measure arbitrary “stages” of speciation, representing stages from the initiation through to the completion of the speciation process (when two populations are depicted, one is shown in blue and the other in orange). Figure and legend taken from [Nosil et al., 2009].

speciation has recently been more broadly defined as the process by which divergence and reproductive isolation arise ([Nosil et al., 2009]; [Mallet, 2009]). Thus, it is also related to the origin of taxa below the species level, and it may or not represent a step towards the origin of species: ecotypes or subspecies may never attain the species status ([Rueffler et al., 2006]; [Nosil et al., 2009]). Importantly, many contemporary species must have gone through the divergence stage shown by populations or subspecies, and thus, their exploration is crucial to understand the causes of diversification. Studying speciation can involve looking at partially isolated divergent ecotypes (the “magnifying glass” approach) informing on how ecology and genetics interact to cause the evolution of reproductive barriers, or by studying “good”, fully isolated species (the “spyglass” analysis) to infer the causes of speciation ([Via, 2009]). Both approaches are complementary to understand the nature and biogeography of the speciation process but they need to be connected. Studying clades composed by taxa placed along a speciation continuum allow to integrate knowledge on the gradual transition from fully interbreeding populations to totally isolated species (e.g. [Hendry et al., 2009]; [Powell et al., 2013]); Figure 3). Figure 3 indicates the continuity of the speciation process and shows different ways to quantify divergence between pairs of taxa placed on distinct “stages” of speciation. Few clades representing such continuity are as well documented as butterflies in the genus *Heliconius* (see below for an in-depth description of this clade).

### 0.1.1 Distinct forms of isolation contribute to diversification and speciation

Under the genotypic cluster species concept, complete reproductive isolation is not a requirement for diversification and not even for speciation to be achieved. The question arises whether reproductive isolation is needed for adaptive diversification or if it is rather a side effect of divergent evolution ([Mallet, 2005]). The idea that isolating barriers evolve through direct selection against gene flow is unlikely (reinforcement being a possible exception; see below). Instead, reproductive isolation is currently believed to be a by-product of the response to natural selection, or of stochastic processes ([Coyne and Orr, 2004]). Isolating barriers accumulate and/or get stronger along the advancement of the speciation process. Therefore, understanding diversification and speciation require pinpointing ecological traits which actually or potentially prevent gene flow between coexisting taxa. Although the barriers we detect at present are not necessarily the ones that were implicated in the initial reduction of gene flow between populations, and although their current importance may distort their historical importance during the speciation process ([Coyne and Orr, 2004]), they remain highly informative of how reproductive isolation evolves.

Complete lists of possible reproductive barriers between divergent taxa have been compiled by Dobzhansky and Dobzhansky ([Dobzhansky and Dobzhansky, 1937], p. 231–232), and Coyne and Orr ([Coyne and Orr, 2004] p. 28–29). Basically, reproductive mechanisms can be divided into those acting before and after fertilisation (pre and post-zygotic barriers, respectively). Pre-zygotic isolation comprehends ecological divergence (e.g. habitat preference and phenological differences), sexual isolation due to assortative mating and gametic incompatibility ([Seehausen et al., 2014]). By assortative mating we mean that individuals mate preferentially with other individuals using the same resource or habitat, or being similar for characters that also cause ecological specialisation. Some examples of this phenomenon will be given below. Post-zygotic reproductive barriers, which affect hybrid sterility and inviability, can be intrinsic or extrinsic. Although the frontiers between these are not always so precise, intrinsic post-zygotic barriers, which do not depend on the external environment, are the result of genetic incompatibilities such as negative epistatic interactions between alleles at two or more loci (Bateson–Dobzhansky–Muller incompatibilities; e.g. ([Brideau et al., 2006])). In contrast, extrinsic post-zygotic isolation is the result of disruptive selection against hybrids mediated by environmental factors. For instance, hybrids between a young species pair of sticklebacks that are ecologically and morphologically differentiated (limnetic and benthic) grow slower than the parental types in both parental habitats ([Hatfield and Schluter, 1999]). Several other examples of maladaptation of hybrids to the parental habitats have been reported ([Crespi, 1989]; [McBride and Singer, 2010]), including *Heliconius* butterflies (see below; ([Merrill et al., 2012])). Speciation is usually the result of the accumulation of several forms of isolation; the relative importance and the timing at which these forms appear along the speciation process are not easy to assess, but seem to vary depending upon the geographic mode of selection (i.e. the degree of geographic-isolation) and upon the studied taxa ([Coyne and Orr, 1998]; [Schemske, 2000]).



Targeting single components of isolation, like intrinsic post-zygotic barriers ([Orr and Turelli, 2001]; [Brideau et al., 2006]) or assortative mating ([Crespi, 1989]; [Kondrashov and Shpak, 1998]; [Jiggins et al., 2001]; [Emelianov et al., 2001]; [Jiggins et al., 2005a]; [Malausa et al., 2005]) between taxa is highly useful for the understanding of the speciation process. However, exploring the relative importance of distinct reproductive barriers, and of particular barriers in related lineages at varying degrees of evolutionary divergence (i.e. the speciation continuum) gives an even better picture of how diversification proceeds. In some speciation events, divergence along multiple phenotypic axes (due to multifarious selection) is involved. For instance, multiple additive minor barriers can contribute to reproductive isolation ([Ramsey et al., 2003]; [Dopman et al., 2010]). For instance, Dopman et al. 2010 have found seven significant sources of isolation between the E and Z host-races of the European corn borer. Alternatively, in some speciation instances major reproductive barriers play alone an outstanding role in diversification, by resulting in the incidental evolution of distinct components of reproductive isolation (e.g. pleiotropic effects). For instance, in Darwin's finches, beak size conditions both trophic specialisation and mate recognition due to song produced by males ([Podos, 2001]). Similar cases where traits under ecological selection are important for mating include host specificity and cuticular hydrocarbon composition in *Drosophila* ([Etges and Ahrens, 2001]), wing patterns involved in defence from predators and signalling to conspecific types in butterflies ([Vane-Wright and Boppre, 1993]; [Brunton and Majerus, 1995]; [Jiggins et al., 2001]); e.g. mimicry in *Heliconius* butterflies, see below), among others ([Boughman et al., 2005]). In such cases, one single trait is able to enhance or maintain both extrinsic post-zygotic and pre-zygotic isolation. In some particular cases, shifts in habitat preference lead directly to pre-zygotic isolation (i.e. phytophagous insects that mate on or near the host plant; see [Drès and Mallet, 2002] for a review). It remains largely unsolved, however, whether these traits of major effect are able alone to drive speciation or if they are protagonists at the beginning of a process which requires multidimensional divergence to be achieved.

Mate choice has been long highlighted as an important factor in speciation ([Dobzhansky and Dobzhansky, 1937]; [Paterson, 1985]; [Kondrashov and Shpak, 1998]). Already Darwin noted that several rapidly evolving traits of animals served for intraspecific communication, especially during competition for mates, and had nothing to do with ecological adaptation. He proposed that these traits evolved through sexual selection and not through natural selection. A variety of signals are displayed usually during courtship, and act as characteristic features of individuals for mate recognition, not only informing about mate quality (what for Darwin was sexual selection), but more generally about the species or ecotype, the gender, etc. Therefore, changes in these signals, which can be visual, chemical, behavioural, tactile, morphological, acoustic, etc., can constitute, either singly or in combination, isolating factors between diverging taxa ([Pivnick et al., 1992]; [Svensson, 1996]; [Marco et al., 1998]; [Martín and López, 2000]). Assortative mating based on reproductive traits such as coloration (including variation in the ultraviolet ([Silberglied, 1979]) is widespread across animal taxa ([Jiggins et al., 2001]; [Salzburger et al., 2002]; [Elmer et al., 2009]; [Jiang et al., 2013]; [Pérez i de Lanuza G, 2013]). For instance, the implication of wing colour patterns

**Box 1** The role of visual and chemical signals in mating behaviour in Lepidoptera

Very diverse signal modalities for mate recognition exist in the order Lepidoptera. Wing coloration and chemical signalling are often implicated in courtship and serve in inter and intraspecific communication as well as in sexual selection ([Boppré, 1984]; [Silberglied, 1984]). For species and subspecies recognition in particular, butterflies and moths show clear differences in the general trend of use of particular kinds of reproductive traits. In moths, chemical communication is more or less ubiquitous, with females emitting important quantities of pheromones that assemble conspecific males from long distance ([Lofstedt, 1993]; [Millar, 2000]; [Jurenka et al., 2003]). Sex pheromones from around 1500 species of moths have been identified, in part motivated by the interest on detecting the molecules that could be potentially used for the bio-control of populations of moths that affect crops ([Arn et al., 1992]). In contrast, in butterflies, males are usually attracted to females by long-range visual signals, following specific behavioural displays by both sexes ([Stride, 1957]; [Silberglied and Taylor, 1973]; [Rutowski, 1991]; [Jiggins et al., 2001]; [Ellers and Boggs, 2003]). Females of some species use male colour patterns for species recognition ([Robertson and Monteiro, 2005]). In butterflies, chemical communication has received much less attention than in moths, but it seems evident that pheromones come into play in the short range, in a late phase of courtship, together with other signals including tactile communication ([Rutowski, 1991]; [Vane-Wright and Boppre, 1993]). Therefore, not only volatile compounds, but also cuticular compounds acting as contact pheromones ([Hay-Roe et al., 2007]; [Dapporto, 2007]; [Ômura and Honda, 2011]; [Heuskin et al., 2014]) seem to take place as message senders in butterflies. Such olfactory signals are usually emitted by males ([Schulz et al., 1993]; [Costanzo and Monteiro, 2007]; [Nieberding et al., 2008]) and thus drive female choice. For example, research on pierid butterflies has demonstrated that males emit mainly scents that are both species and sex specific; namely, males of species *Pieris napi*, *P. rapae* and *P. brassicae* smell strongly of citral, indole and benzylcyanide, respectively ([Bergström and Lundgren, 1973]; [Andersson et al., 2007]). Very rarely, female pheromones are important for male choice. Such cases have been scarcely reported ([Lundgren and Bergström, 1975]; [Wago, 1978]).

in mate recognition has been reported in several butterfly species ([Rutowski, 1977b]; [Silberglied and Taylor, 1978]; [Jiggins et al., 2001]; [Fordyce et al., 2002]; [Sweeney et al., 2003]; [Robertson and Monteiro, 2005]; [Kemp, 2007]); see Box 1 for more details about mate choice in Lepidoptera). Although the evolutionary role of assortative mating is well understood, punctual questions remain largely unsolved. For instance, when does assortative mating evolve relative to other reproductive barriers? Does it get stronger along the speciation continuum? These are questions that need to be tackled in a comparative way within a clade representing the speciation continuum.

Pre-zygotic reproductive isolation based on coloration seems to be one of the causes of the astonishing diversity of colour patterns in some groups, like birds, fishes, frogs and insects. A good example of such diversity is the ~17.000 existing butterfly species, which are mostly distinguishable on the basis of their wing colour motifs. Visual signals across animals have been extensively investigated because of their high diversity, their conspicuousness and their importance in sexual communication in nature. However, comparatively few studies have looked at the importance of chemical-based signalling in mate choice, although it is often regarded as the most widespread form of communication, reported for taxa ranging from unicellular organisms to mammals ([Wyatt,

---

**Box 2** A few words about pheromones
 

---

Chemical information is transferred through semiochemical substances, among which pheromones are the best known. A pheromone is a chemical (or a mixture of chemicals) secreted by an animal, that “releases a specific reaction” in an individual of the same species ([Karlson and Lüscher, 1959]). Pheromones can reveal the gender, age, social status, belonging to a given group or taxon, etc., or help to find mates, to warn about a danger, to delimitate an area of distribution, to enhance aggregation behaviour, etc. ([Wyatt, 2003]); such messages may or not be transferred during courtship. The information that courtship pheromones (also called sex pheromones) carry needs to be sufficiently conserved to contribute unambiguously to the perception and location of a suitable mate (e.g. of the same species; [Wyatt, 2003]), but also versatile to send a precise message about “mate quality” to prospect mates (e.g. [Andersson, 1994]).

---

2003]; [Ferveur, 2005]; see Box 2). Chemical signals are often divergent across species and even between ecotypes (e.g. [Cardé et al., 1978]) and thus serve for conspecific recognition (e.g. [Baer, 2003]). Interestingly, the chemical structure or composition of pheromones of sister sympatric taxa often show drastic differences, a form of character displacement ([Mérot et al., 2015]; [Coyne and Orr, 1997]; [Shine et al., 2002]). For example, closely-related species of beetles in the genus *Ips* display more dissimilar pheromones acting as attractant to mating sites than far related species ([Symonds and Elgar, 2004]).

When visual signals are constrained, for instance by a protective function (e.g. mimicry), increased dependence on chemical signals for species recognition may be expected ([Vane-Wright and Boppre, 1993]). Little empirical data exist on testing that hypothesis. However, it seems to hold true for *Amauris* butterflies in eastern Africa, where low colour pattern divergence contrast with a high specificity of the complex volatiles bouquets associated with males abdominal hairpencils ([Schulz et al., 1993]). Textbook examples of mimicry complexes like the ones formed by neo-tropical *Heliconius* butterflies have been deeply explored in terms of the involvement of wing colour patterns in male choice ([Jiggins et al., 2001]; [Estrada and Jiggins, 2008]; [Merrill et al., 2011a]); see below) but little is known about the role of pheromones in mate recognition and in the diversification of *Heliconius* butterflies (but see ([Mérot et al., 2015]; [Estrada et al., 2011])). Do closely-related mimetic taxa rely on chemical cues for species recognition? Does differentiation in chemical signals evolve at similar rates than visual signals?

### 0.1.2 Phenotypic divergence as a necessary step towards ecological speciation

When populations diverge phenotypically in response to environmental constraints, ecological speciation may happen ([Schemske, 2000]; [Schluter, 2001]). Although stochastic changes under geographic isolation were long seen as the main factor for speciation to proceed ([Mayr, 1942]), the role of adaptive change in speciation has recently regained importance. Divergent natural selection is expected to pull populations toward distinct adaptive peaks if these inhabit distinct environments composed

of different ecological niches. In the same way, specific processes such as competition for resources between coexisting taxa and predation may favour the adaptation to new environments where the taxa experience contrasting selection regimes ([Mallet and Barton, 1989a]; [Schluter, 1994]; [Rundle et al., 2003]; [Nosil and Crespi, 2006]). Most importantly, the evolution of reproductive isolation may occur as a “by-product” of the genetic changes underlying diverging features among taxa ([Coyne and Orr, 2004], p.385). In a number of widely cited cases of ecotypes or closely-related species, barriers to gene flow seem to have evolved as a side effect of divergent ecological selection. For instance, the beaks of Darwin’s finches, host plant specificity in *Rhagoletis* fruit flies and *Timema* stick insects, body size in stickleback fishes in postglacial lakes, coloration in poison frogs and in cichlid fishes, pollinator targets and flowering time in plants, wing patterns in mimetic *Heliconius* butterflies, are all examples of ecological divergence resulting in reproductive isolation ([Grant and Grant, 1992]; [Feder, 1998]; [Mallet et al., 1998a]; [Via et al., 2000]; [Nosil, 2004]; [Jiggins, 2008]).

The genus *Heliconius*, which we will study here, has undergone a radiation of wing coloration which has received the attention of evolutionary biologists and ecologists since Bates ([Merrill et al., 2015]); see below and Box 3 for more details on this group). Striking patterns of ecological and phenotypic diversity in rapidly multiplying lineages, or adaptive radiations, are also studied in certain clades of fishes, like cichlids in African lakes or Nicaraguan crater lakes, for Caribbean *Anolis* lizards, spiders in Hawaii, Darwin’s Galapagos finches, Hawaiian crickets, columbine flowers, among others([Grant and Grant, 1992]; [Hodges and Arnold, 1994]; [Losos and De Queiroz, 1997]; [Shaw, 2002]; [Gillespie, 2004]; [Barluenga et al., 2006]). These radiations range from around fifteen species (Galapagos finches; [Grant and Grant, 2008]) to over 500 species (Lake Victoria cichlids; ([Johnson et al., 2000]; [Nagl et al., 2000]).

Ecological specialisation as a result of divergent selection has a well-reported potential to favour the formation of new species (e.g. ([Orr and Smith, 1998]; [Schemske, 2000]; [Schluter, 2001]; [Via, 2002]), but “non-ecological” speciation is also thought to be an important factor explaining extant diversity. Speciation under uniform selection, polyploidy speciation and speciation by genetic drift may all occur without the direct involvement of natural selection, although only the latter may be strictly non-ecological (see [Sobel et al., 2010] for a review on this topic). So, a combination of adaptive and non-adaptive changes may concur to observed diversification patterns, but there is a large consensus today about adaptation playing a significant role in most speciation instances ([Templeton, 2008]). According to Coyne and Orr (2004, p.179) “...virtually all barriers can be considered ecological in the sense that they arise from environmentally imposed selection”. Similarly, all instances of ecological speciation result from environmentally-imposed selection and this may be one of the few ways to diverge in presence of gene flow. I will now introduce our study system, where phenotypic diversification is importantly shaped by predation as an ecological agent of selection ([Mallet et al., 1998a]).

---

**Box 3** Butterflies in the genus *Heliconius* are an example of an adaptive radiation
 

---

According to [Schluter, 2000] an adaptive radiation is the rapid diversification of an ancestral population into several ecologically different species, each of which is adapted to a specialised environmental niche. The genus *Heliconius* has been presented as a good example of an adaptive radiation, given that their impressive diversity of wing colour patterns seems to have originated over a relatively short evolutionary time. Namely, the around 40 species and the 10-fold so many geographic races have started diversifying ~10 million years ago (Figure 4; [Kozak et al., 2015]). The striking diversification of the genus strongly contrasts with the low one observed in sister genera within the tribe Heliconiini, where rather conservative wing patterns extend over large geographic areas. This contrast was suggested to rely importantly on the evolution of mimicry as a force possibly leading to the exploration of existing but previously underutilised adaptive zones by members in the genus *Heliconius* (i.e. the mimicry rings formed by several far-related butterfly species). Nevertheless, given that mimicry is based on selection against rare morphs, the rampant geographic variation in *Heliconius* is hard to explain.

There seems to be a change from disruptive to stabilising selection on individual populations as the number of species participating in local mimetic environments increases, but explaining how the initial divergence occurs is elusive. It has been suggested that differentiation arises within essentially continuous populations thanks to disruptive natural and sexual selective pressures on colour loci. The “shifting balance” hypothesis, according to which new forms can evolve in local populations and spread to others ([Wright, 1932], [Wright, 1977]) is now better accepted to explain the emergence and maintenance of new mimetic forms than the “Pleistocene refugia” hypothesis, which states that distinct colour patterns evolved in allopatry in forest refuges formed during cycles of Pleistocene cooling and came into secondary contact when the geographic barriers disappeared ([Brown and Benson, 1974]; [Sheppard et al., 1985]; [Whinnett et al., 2005]). The shifting balance hypothesis considers three distinct phases. First, a new form arises locally via genetic drift or temporary changes in the selection pressures (phase I). During phase II, the new form becomes fixed locally through natural selection, and finally, during phase III, the fitness-enhancing trait spreads to other populations via interdemic selection. Once a new form has become locally, the depression of fitness of hybrids relative to that of parental types is enhanced through the involvement of wing colour patterns in mating. In addition to mimicry, the rapid adaptive radiation undergone by the genus is possibly linked to some traits unique among Lepidoptera (pollen feeding and pupal-mating behaviour; see [Beltran et al., 2007] for a description) and with the co-evolution with Passifloraceae host plants ([Benson et al., 1975]).

---

### 0.1.3 Müllerian mimicry as a driver of wing pattern diversification in *Heliconius* butterflies

Butterflies in the neotropical genus *Heliconius* are spectacularly diverse in their wing colour motifs (see Figure 4 for a recent phylogeny of the clade). They are unpalatable, and their wing colour patterns act as signals warning predators of their toxicity. Furthermore, several toxic species commonly show wing colour pattern convergence locally, thus belonging to the same mimicry ring. This is a positive frequency-dependent process: more common patterns enjoy better protection from predators. This phenomenon is known as Müllerian mimicry.

Mimicry often involves distantly related species within the genus (e.g. *H. melpomene*/*H. erato*, *H. cydno*/*H. sapho* or *H. sara*/*H. doris*) or *Heliconius* species and taxa outside

the genus (e.g. the case of the silvaniforms- or tiger-patterned butterflies- with butterflies in the Danainae and Ithomiinae subfamilies). In contrast, closely-related species often differ in wing colour pattern (e.g. *H. melpomene*/*H. cydno*), which supports the idea that shifts in colour patterns may favour speciation (see paragraph below; [Mallet et al., 1998a], [Mallet, 2010]; [Jiggins et al., 2001]; [Chamberlain et al., 2009]). In some cases, however, closely related species can share mimetic relationships (e.g. sister species *H. melpomene amaryllis*/*H. timareta thelxinoe* and races of silvaniform species like *H. numata elegans*/*H. ethilla aerotome*/*H. pardalinus butleri* or *H. hecale melicerta* and *H. ismenius bouletti*). The adaptation of these butterflies to mimetic environments, called mimicry rings and recognised by the local community of predators, suggests that wing colour motifs are under strong predator-imposed selective forces.

It is not evident how mimicry, a phenomenon implicating phenotypic convergence, may also be associated with widespread phenotypic divergence, and may contribute to speciation. Mimetic clades are often highly diverse, both at a local scale and at a broad geographic scale ([Alexandrou et al., 2011]; [Rosser et al., 2012]; [Twomey et al., 2013]). In *Heliconius*, several mimicry rings, rather than a single one, coexist at a given location, which contradicts the expectations of Müllerian mimicry as a stabilising process. The reasons for such diversity are not clear yet, but it has been suggested that different wing patterns may respond to microhabitat heterogeneity within the forest ([Joron and Mallet, 1998]). At a larger geographic scale, most *Heliconius* species display astonishing racial colour pattern diversity, representing a mosaic distribution of parapatric races. These subspecies share contact zones that are usually narrow, an indication of strong natural selection acting on alternative phenotypes on either sides of the clines, where local predator communities learn to avoid the local warning pattern. Predator-operated disruptive selection this way favours the pattern that is abundant (i.e. the parental pattern) and disfavors rare and non-mimetic patterns (i.e. recombinant patterns), which are removed by frequency-dependent selection. Evidence of selective pressures acting against non mimetic butterflies has been collected in the field ([Benson, 1972]; [Mallet and Barton, 1989b]; [Kapan, 2001]). Briefly, these experiments consisted in releasing modified local mimetic patterns ([Benson, 1972]), transferring colour pattern subspecies across an interracial contact zone ([Mallet and Barton, 1989b]) and reciprocally transferring polymorphic forms of a single species ([Kapan, 2001]). Hence, disruptive selective forces favour and stabilise alternative mimetic patterns locally, but enhance phenotypic diversity if a larger geographical view is taken. As shown so far, shifts in colour patterning determine phenotypic distinctiveness between lineages already at early stages of divergence in the clade. However, divergent selection does not only stabilise distinct colour races locally, but also distinctly-coloured parapatric species (e.g. the *H. erato*/*H. himera* or the *H. cydno*/*H. pacheus* pairs) and sympatric species (e.g. *H. melpomene* and *H. cydno*). Actually, good evidence supports that shifts in mimetic colour patterns might have facilitated ecological divergence and the maintenance of reproductive isolation in *Heliconius* ([Jiggins et al., 2001]; [Jiggins, 2008]). For instance, sister species *H. melpomene* and *H. cydno* coexist in a large area but belong to distinct mimicry rings. In Panama, for example, *H. melpomene* and *H. cydno* mimic the far-related species *H. erato* and *H. sapho*, respectively, suggesting that mimicry adaptation has caused a shift in colour

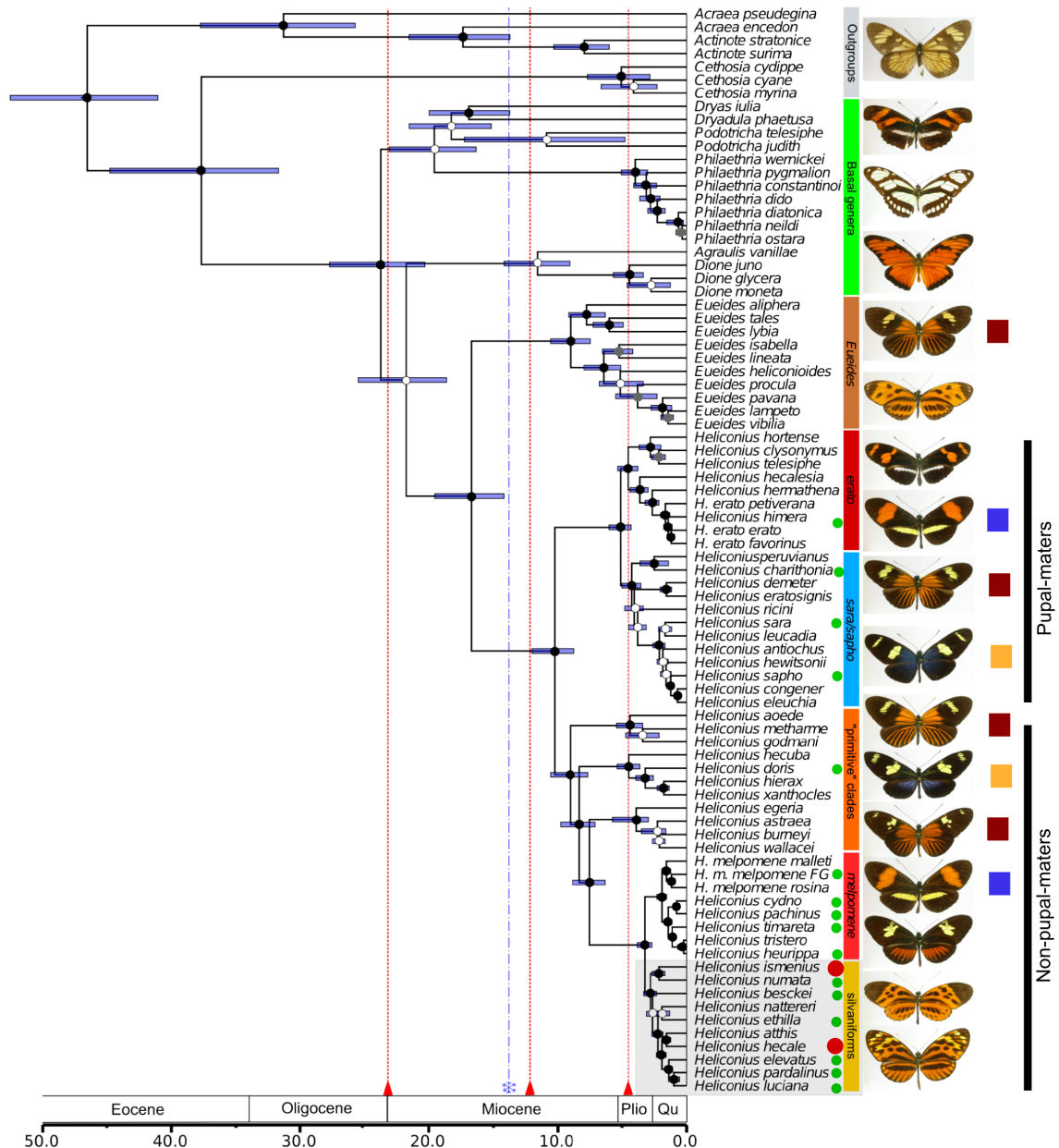


Figure 4 – **Phylogeny of the tribe Heliconiini.** This study focuses on species *Heliconius hecale* and *H. ismenius* (indicated with red circles) which belong to the silvaniform clade (highlighted by a shaded box on the bottom and detailed in Figure 1.4 of Chapter 1). Green dots represent species referred to in the text. The two major clades in the genus *Heliconius*, pupal-maters and non-pupal-maters, are separated by black vertical bars on the right side of the figure. Coloured squares on the right side of the phylogeny highlight mimetic relationships (when the colour is the same) between far-related species. Figure slightly modified from [Kozak et al., 2015].

pattern that might have contributed to species divergence. Divergence of mimetic patterns has led to reproductive isolation, both at the pre-zygotic level (i.e. assortative mating based on colour; see next paragraph; [Jiggins et al., 2001]) and at the post-zygotic level (i.e. disruptive selection due to predation; [Merrill et al., 2012]). In combination with a degree of hybrid sterility and ecological divergence, mimicry might have led to ecological speciation. Martin et al. (2013) found that gene flow has persisted throughout the process of divergence among these species, including the present time, the latter observation being supported by previous studies ([Bull et al., 2006]; [Kronforst et al., 2006b]; [Heliconius Genome Consortium, 2012]). Although the findings by Martin et al. (2013) cannot rule out brief periods of geographic isolation, these support the hypothesis of speciation-with-gene-flow driven by mimicry. Together, all these observations show how Müllerian mimicry, a phenomenon of convergence, is believed to be involved in speciation, which implies divergence. However, the mechanisms underlying the initial divergence remain elusive (see Box 3 for a brief description of the explaining hypotheses).

Colour patterns in *Heliconius* butterflies are also implicated in mate preferences. Males typically prefer to court females (or female wing models) carrying their own pattern ([Jiggins et al., 2001], [Jiggins et al., 2004]; [Mavárez et al., 2006]; [Melo et al., 2009]; [Merrill et al., 2011a]). Wing patterns in *Heliconius* have been referred to as “magic traits” for speciation (c.f. [Servedio et al., 2011]) because of their simultaneous contribution to the pre-zygotic isolation (assortative mating) and post-zygotic isolation (disruptive selection for alternative mimicry associations). However, how genes for mimetic adaptation and mate preferences become associated remains largely unresolved.

As outlined above, *Heliconius* butterflies represent a diversified clade where distinct stages of the speciation continuum can be found. The stage of lowest reproductive isolation along this continuum is composed by coexisting colour morphs in polymorphic populations. A striking example of this local polymorphism is shown by *H. numata* (see below), and other species such as *H. cydno*, *H. doris*, *H. hecale* and *H. ismenius* also have some polymorphic populations ([Kapan, 2001]; [Chamberlain et al., 2009]). Further along the continuum are parapatric races of the same species that differ in wing pattern separated by transition zones with strong gene flow. Next, in a few cases, incipient parapatric species overlap in contact zones with occasional hybridization (e.g. *H. erato* and *H. himera*, *H. cydno* and *H. pacheus*). Some pairs of coexisting closely-related species show pervasive gene flow despite the evolution of strong isolating barriers (*H. melpomene* and *H. cydno*). Finally, distantly-related species coexist in sympatry and are believed to never hybridise (e.g. *H. melpomene* and *H. erato*). [Merrill et al., 2011b] studied this continuum and evaluated the strength of assortative mating between four pairs of taxa that coexist in a portion or in the totality of their distribution (Figure 5A). Figure 5A shows how premating isolation (in the form of population-based male mate preference) increases throughout the continuum from polymorphic populations to “good” sympatric ecologically divergent species. Recently, comparative genomic surveys of genetic differentiation among different *Heliconius* species and subspecies have also shown the progression in the extent of reproductive isolation (Figure 5B; [Nadeau et al., 2012]; [Hill et al., 2013]; [Martin et al., 2013]). [Martin et al.,

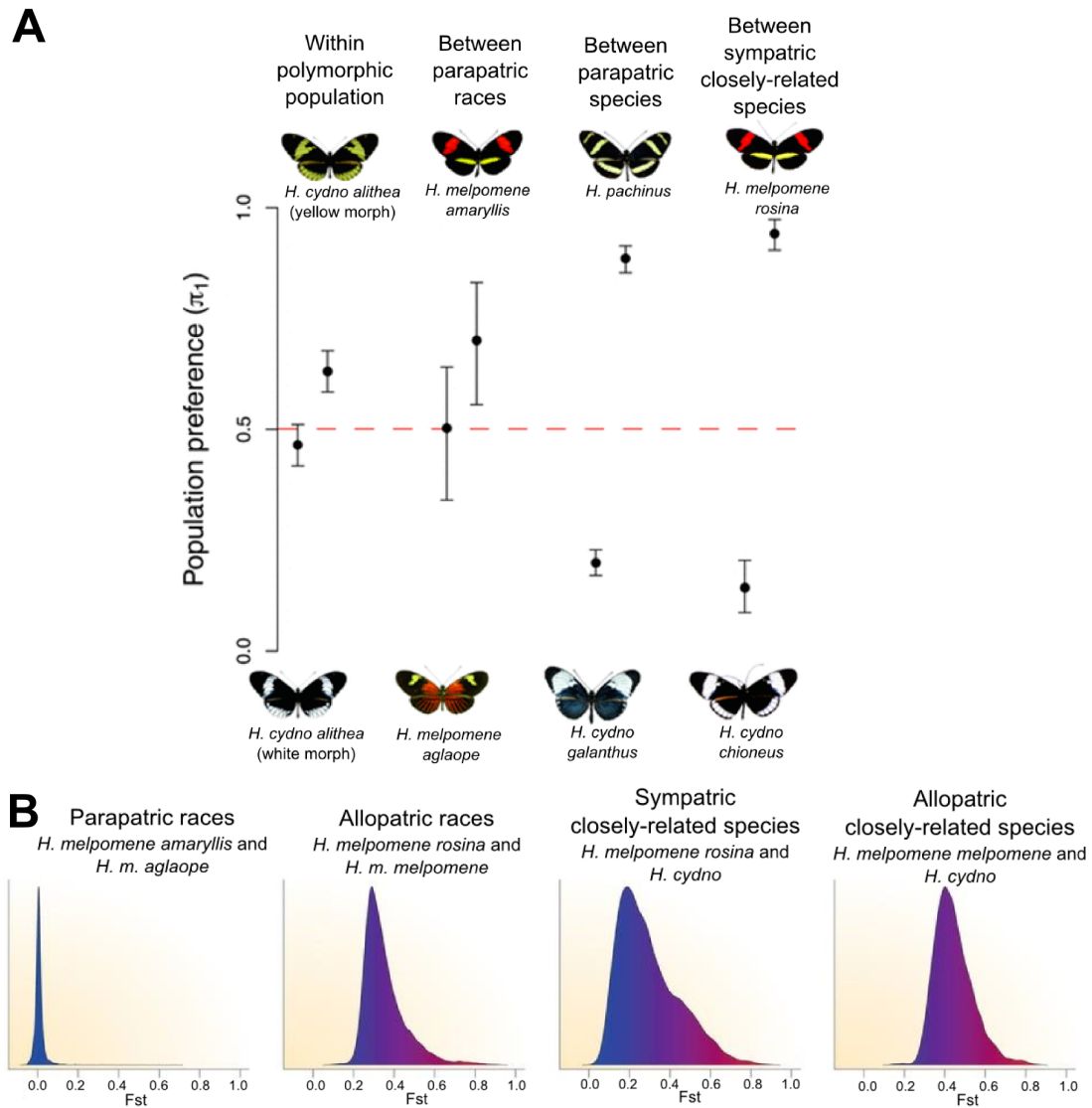


2013] showed that parapatric races of *H. melpomene* display virtually no genomic differentiation apart from the high peaks of divergence associated with wing colour loci (Figure 5B). This contrasts with the comparison of the geographically-isolated subspecies *H. melpomene rosina* and *H. m. melpomene*, displaying similar levels of genome-wide genetic differentiation to the closely-related species *H. melpomene rosina* and *H. cydno* in sympatry. Moreover, allopatric subspecies of *H. melpomene* show partial intrinsic hybrid infertility ([Jiggins et al., 2001]), contributing to the general observation that closely-related species show more divergence in allopatry than in sympatry. Isolation by geography clearly contributes to divergence in the genus while pervasive gene flow occurs between coexisting closely-related species (see below). Together, these observations suggest that distinct geographic modes of divergence and speciation have taken place in the diversification of this clade of butterflies.

#### 0.1.4 Speciation as a response to geography

In order to better understand the ultimate causes of speciation, classifying the cases of divergence according to the degree of spatial isolation existing between the diverging taxa has proven useful (Figure 6). This separates divergence into scenarios without gene flow (allopatric speciation) and with varying levels of gene flow (sympatric and parapatric speciation). Gene flow can this way be investigated as one of the mechanisms of evolutionary change, acting usually in combination with other evolutionary forces such as natural selection, genetic drift and mutation.

When physical barriers to gene exchange exist (allopatry) gradual neutral divergence (the fixation of mutations stochastically in either of the isolated populations) may result in weak or strong disruptive selection and cause reproductive isolation to evolve ([Mayr, 1942]). By contrast, when there is gene exchange between populations, deterministic mechanisms (i.e. natural and sexual selection) can enhance divergence and contribute to maintain species boundaries, and often play a more important role in divergence than genetic drift. In fact, strong disruptive selection (exceeding some threshold proportional to the effective migration rate; [Barton, 2013]) on traits implicated in habitat or resource use is required ([Rice and Hostert, 1993]; [Via, 2001]) to prevent the formation of hybrids both when the diverging populations coexist in a broad portion of their spatial distribution (sympatry) and when they meet in adjoining regions of their range (parapatry). The difference between sympatric and parapatric scenarios of speciation relies on migration (or dispersal in plants). In parapatric models of speciation, only a small fraction of individuals in each region encounters the other, but such a limited coexistence in so called “hybrid zones” (see next section) can still allow considerable gene flow between populations. Parapatric speciation has been considered anecdotal ([Coyne and Orr, 2004]). For instance, hybrid zones are often assumed to derive from secondary contact between taxa diverged in allopatry rather than by *in situ* divergence. However, in some hybrid zones parapatric divergence seems highly likely (e.g. the contact zone between morphs of the *Littorina saxatilis* snails in northern Spain; [Rolán-Alvarez et al., 1997]). Hybrid zones between subspecies differentiated by a few ecological traits are excellent situations to understand the mechanisms for the maintenance of differences in the face of gene exchange.



**Figure 5 – Speciation continuum in *Heliconius* butterflies** (A) mirrored by the strength of assortative mating between taxa at distinct stages of divergence and (B) shown by the extent of genomic divergence between distinct taxa. Part A shows how male mate preference of syntopic females increases throughout the continuum from polymorphic populations to sympatric ecologically different species. In each case 1 would indicate a complete preference for the red or yellow female type of the pair and 0 would indicate a complete preference for the white or rayed female type of the pair. Part B shows the shapes of the frequency distributions of  $F_{st}$  values between races and subspecies with distinct degrees of geographical isolation. Part A is slightly modified and legend is partially taken from [Merrill et al., 2011a]. Part B is slightly modified from Seehausen, but the original data comes from [Martin et al., 2013].

It has been suggested that in presence of gene flow, speciation is greatly facilitated between populations diverging for resources or habitat use (i.e. ecological speciation) when there is assortative mating ([Bush, 1975]; [Felsenstein, 1981]; [Diehl and Bush, 1989]; [Kirkpatrick and Ravigné, 2002]; [Malausa et al., 2005]). The pleiotropic effects of ecological adaptation alleviate some of the problems of the selection/recombination antagonism, according to which evolving assortative mating is highly difficult in the face of recombination, since the latter can break the association between the alleles for performance and the choice allele ([Felsenstein, 1981]; [Coyne and Orr, 2004]). Stringent requirements such as linkage between genes affecting performance and mating favour divergence in presence of genetic exchange ([Kondrashov and Mina, 1986]; [Rice and Hostert, 1993]; [Schluter, 2000]; [Via, 2001]). In the genus *Heliconius*, more than one mechanism may be operating to drive the association between mimetic adaptation and mate choice. Namely, in the *melpomene* sub-clade, a genetic association between each of two loci responsible for a wing colour preference cue and mate preference on each cue separately was observed ([Kronforst et al., 2006c]; [Merrill et al., 2011b]). Additionally, [Kronforst et al., 2007] have suggested that reinforcement, which is the evolution of premating isolation between taxa owing to selection against unfit hybrids, has been critical driving assortative mating preference within the *melpomene* sub-clade. These authors have found that sexual selection was enhanced in areas of contact between hybridising species.

Renewed interest in sympatric speciation came from findings on the monophyly of radiations occurring all in sympatry ([Meyer et al., 1990]; [Schliwen et al., 1994]). Cichlid fishes in African lakes constitute one of the best documented cases where sympatric speciation seems the most plausible explanation of an adaptive radiation ([Wilson et al., 2000]). Other examples include the host-plant specialist races of herbivores such as *Rhagoletis pomonella* flies, *Timema cristinae* walking-stick insects, the larch budmoth *Zeiraphera diniana* and *Acyrtosiphon pisum* pea aphids ([Feder and Filchak, 1999]; [Via, 1999]; [Emelianov et al., 2001]; [Nosil and Crespi, 2006]). They are an example of “mosaic sympatric speciation” ([Mallet et al., 2009]), because broadly sympatric host plants often represent a locally patchy resource (see Figure 6). These ecotypes often show high levels of reproductive isolation ([Wood and Keese, 1990]; [Emelianov et al., 2001]). Nevertheless, ecological divergence can also be accompanied by only weak reproductive isolation and little genetic differentiation, as observed in *Ostrinia nubilalis* corn borers, *Timema cristinae* stick insects, among others ([Dopman et al., 2005]; [Nosil, 2007]).

Most examples of divergence with gene exchange have been criticised for not being robust enough to reject the null hypothesis of having diverged in allopatry at some point of their evolutionary history ([Coyne and Orr, 2004]). Indeed, some outstanding examples of sympatric speciation appear to have involved a scenario of geographic isolation ([Feder et al., 2003]). For instance, *Rhagoletis* fly races were found to carry chromosomal inversions that might have evolved in allopatry and provided the genetic variation that facilitated sympatric divergence ([Feder et al., 2003]). Another example comes from sympatric distinct forms of the lake whitefish (*Coregonus* sp.) which are repeatedly found in different lakes in Canada. The origin of these forms is believed to have followed genetic divergence evolved in allopatry

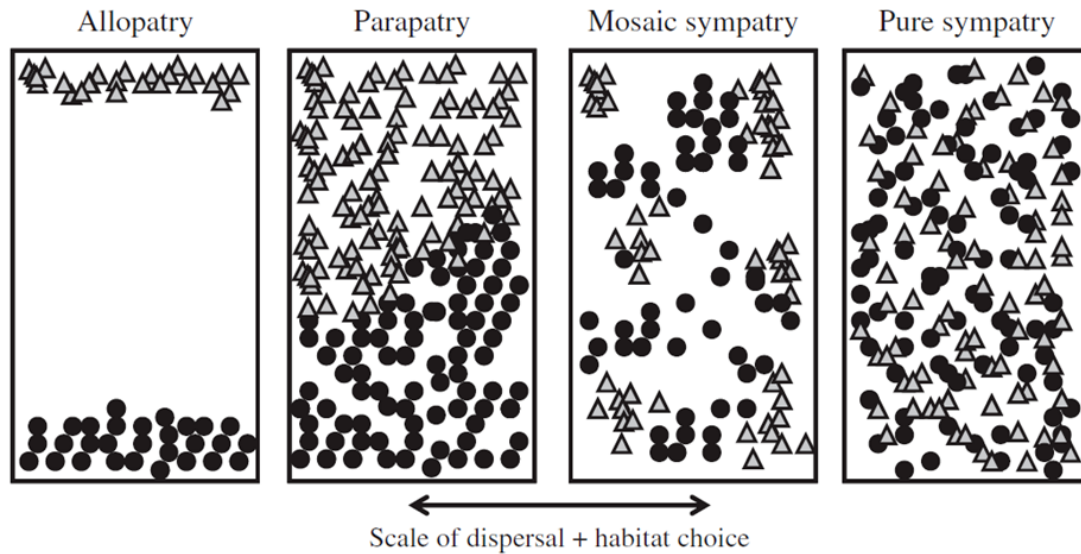


Figure 6 – **The allopatry/sympatry spectrum, using spatial definitions.** Figure and legend taken from [Mallet et al., 2009].

without phenotypic differentiation ([Bernatchez and Dodson, 1990]) and the evolution of a limnetic dwarf form from a benthic normal form after secondary contact ([Landry et al., 2007]). The mentioned cases show, however, the crucial role of character displacement for divergence in sympatry, responding to ecological opportunity and competitive interactions.

Currently, there has been a shift in speciation research away from an emphasis on geography and increasingly towards how selection can favour divergence and what kinds of traits are involved in early stages of speciation ([Via, 2002]; [Jiggins, 2008]).

### 0.1.5 The genomic landscape of divergence with gene flow

A central tenet of population genetics is that the portions of the genome that confer reproductive isolation between diverging taxa are expected to resist gene flow, which allows phenotypic differentiation to be preserved in response to natural selection. In contrast, most of the genome is neutrally affected by migration and genetic drift, and is exchanged freely between hybridising populations. Under a more detailed view of a verbal model of diversification and speciation with gene flow ([Feder et al., 2012]), divergence is expected to start at specific “barrier loci” (c.f. [Abbott et al., 2013]) or “speciation genes” (c.f. [Wu, 2001]), which anticipate and cause the evolution of reproductive incompatibility. At this initial step of divergence, punctual so-called “genomic islands of divergence” (c.f. [Turner et al., 2005]), which divergence exceeds neutral background expectations contrast with the majority of the genome, which is supposed to introgress freely between hybridising populations since it does not confer some degree of isolation. It has been suggested that the next stage implies divergence across greater portions of the genome through hitchhiking of other physically linked adaptive and neutral loci (i.e. divergence hitchhiking). Finally, towards the completion of the speciation process, “genome hitchhiking” further reduces the global genomic

average effective migration rate. The following paragraphs include some evidence supporting and challenging this model.

The view idea of “semi-permeable” species boundaries has been supported by genome-wide scans of divergence between pairs of hybridising taxa, empowered by the innovation on next-generation sequencing (NGS) technologies ([Seehausen et al., 2014]). This has benefited from a population genomics approach relying on the genotyping of a large set of loci to accurately determine the loci associated with adaptation and reproductive isolation, in relation to the level of genomic differentiation under neutrality.

The discovery of genomic islands of divergence in empirical analyses is one of the main contributions of the emerging field of speciation genomics to the understanding of the molecular underpinnings of reproductive isolation barriers ([Seehausen et al., 2014]). However, the enthusiasm about these islands faces some caveats. In particular, relation between divergence and divergent selection remains controversial. Low recombination genomic regions (i.e. near centromeres and within chromosomal inversions) and low nucleotide diversity regions can also show elevated differentiation ([Carneiro et al., 2009]; [Joron et al., 2011]; [Ellegren et al., 2012]; [Cruickshank and Hahn, 2014]). The results of Cruickshank and Hahn (2014) are particularly intriguing. They reanalysed data from five different outstanding studies that had reported genomic islands of divergence, and did not recover such islands when using a measure of divergence that is independent of intra-group diversity ( $D_{xy}$ ) instead of one that depends upon such diversity, and which is usually estimated ( $F_{st}$ ; [Beaumont, 2005]). These results diminish reliability on divergence scans alone for detecting regions functionally involved in diversification and speciation. To distinguish unambiguously divergent selection from other causes in shaping the genomic landscape, it is advantageous to combine top-down and bottom-up approaches to identify genomic regions involved in speciation ([Michel et al., 2010]), that is to say, to combine genome scans with a knowledge on the association of adaptive traits and reproductive isolation. If the genetic architecture of divergent traits is not known at all, genomic scans are “blind” ([Ellegren et al., 2012]), and misleading conclusions are susceptible to be made ([Turner et al., 2005]). For instance, genome scans of divergence among sibling species of the malaria-transmitting mosquito *Anopheles gambiae* found three genomic islands of divergence ([Turner et al., 2005]) and suggested them to be due to selection against hybrid genotypes during backcrossing. However, a recent analysis of reproductive isolation between these species coupled to genomic scans of divergence found that the introgression of the insecticide resistance mutation *gsc-1014F* homogenised one of the tallest genomic islands, but did not hinder reproductive isolation between the species ([Clarkson et al., 2014]). Hence, these authors concluded that islands of divergence may not necessarily have an impact on speciation.

Finally, speciation with gene flow might not always start at islands of divergence. Instead, “continents” of multiple differentiated loci (widespread divergence) were found already at early stages of speciation between *Rhagoletis pomonella* host races ([Michel et al., 2010]). Widespread genomic divergence was also observed between sibling *Anopheles* species ([Lawniczak et al., 2010]). These results indicate that the steps proposed by this model can be intricate along a speciation continuum. Concerning more

advanced stages of speciation, there is empirical data that seem to support divergence and genomic hitchhiking ([Rogers and Bernatchez, 2007]; [Nosil et al., 2008]; [Jones et al., 2012]). However, by comparing three pairs of *Heliconius* species showing distinct degrees of reproductive isolation [Kronforst et al., 2013] found that in advanced stages on the speciation continuum, divergence evolves due primarily to the origin of new divergent regions, and not to the enlargement of the islands of divergence related to wing coloration loci. Other studies have also shown small individual regions of genomic divergence between well isolated species ([Turner et al., 2005]).

The understanding of the genomic landscape of divergence with gene flow is enabled by comparative studies between populations at distinct stages along the speciation continuum, even if the study systems may not represent a direct evolutionary progression. *Heliconius* butterflies are among the few cases where the sequential accumulation of presumed barrier loci has been explored at different points of the speciation continuum within one single clade. These studies have, however, only performed comparisons from the subspecies to the sister-species levels. What are the outcomes of a longer evolutionary time in shaping the genomic landscape of divergence between co-existing lineages? Also, does the genomic profile of differentiation reflect the existence of a few barrier loci at early stages of divergence when other sub-clades in the genus are explored? Can differentiation outliers be really considered barrier loci having a role in diversification? These are questions that need to be further investigated to get a better understanding of the speciation process.

## 0.2 How does hybridisation affect diversification?

*We used to make fun of Edgar Anderson by saying that he was finding hybrids under every bush. Then we realized that even the bushes were hybrids.*

Warren H. Wagner; taken from Abbott et al. 2013

Following [Harrison, 1993], hybridisation is “the interbreeding of individuals from two populations, or groups of populations, which are distinguishable on the basis of one or more heritable characters”. This definition overcomes the difficulty of categorising taxonomic units (e.g. species) implicated in the process, and extends over less differentiated taxa (e.g. subspecies). The role of hybridisation on biodiversity genesis and dynamics depends on how far this process advances into producing hybrid generations. In particular, the generation of (at least partially) fertile or viable offspring and the transfer of genetic material between the hybridising taxa through backcrossing leads to genetic admixture; this process is called introgressive hybridisation or introgression ([Anderson, 1949]).

Confusion exists in relation to the evolutionary outcomes of hybridisation. This controversy is fuelled by evidence of both the restraining and creative effects of hybridisation. An overview of this evidence will be given below. Then, a glimpse at the incidence of natural hybridisation will be presented, including the approaches applied to detect it, given that the importance of hybridisation in shaping biodiversity may be directly proportional to its frequency in nature. Next, I will present a general

description of areas where populations are believed to be undergoing important levels of hybridisation (i.e. hybrid zones), because such systems provide natural laboratories to examine the speciation process. The organisation of the subsections that will be outlined here responds somehow to the history of research on hybridisation reviewed by ([Schwenk et al., 2008]).

### 0.2.1 The evolutionary outcomes of hybridisation

For adaptation, divergence and speciation to proceed, hybridisation represents a problem because of its homogenising effect on genetic variation, and hence, its counteracting effect on differentiation. Hybridisation was long considered unbeneficial for diversification, since the cessation of gene flow and recombination was thought to be absolutely necessary for population differentiation and the generation of new species ([Dobzhansky and Dobzhansky, 1937]; [Mayr, 1942]). Hybridisation can act as a constraint on divergence by continuously breaking down linkage disequilibria between adaptive loci ([Felsenstein, 1981]; [Kirkpatrick and Ravigné, 2002]), and this is one of the causes of the production of unfit hybrids.

However, rather than always counteracting the diversification process, the effects of hybridisation can sometimes be constructive and enhance diversification by generating new species, adaptive novelty and adaptive variation ([Anderson, 1949]; [Stebbins, 1959]). One of the first outstanding studies that showed experimentally the potential significance of introgressive hybridisation in animal evolution was the work by Lewontin and Birch ([Lewontin and Birch, 1966]) on the Australian fruit flies *Dacus tryoni* and *D. humeralis*. Nowadays, it is accepted that the effects of hybridisation on the speciation process are highly dependent upon the involved hybridising taxa, the stage of divergence and the ecological context where it occurs ([Abbott et al., 2013]). Hybridisation is, however, an inevitable component of diversification, and needs to be investigated to get a better picture of the evolutionary process.

Findings about the recurrence of natural hybridisation have unveiled the importance of reticulate evolution, which is based on natural hybridisation between different evolutionary lineages ([Arnold, 1992]). Hybridisation occurs inevitably in almost all proposed speciation processes, excepting cases of instantaneous or strictly allopatric speciation ([Barton, 2013]; [Abbott et al., 2013]), for instance in species complexes that have evolved through adaptive radiations ([Seehausen, 2004]). Seehausen (2004) proposed that the high incidence of hybridisation between populations invading new environments, its potential to increase genetic diversity and to elevate the response to selection, might all represent reasons for rapid diversification under divergent selection in instances of adaptive radiation. Box 3 presents the butterflies in the genus *Heliconius* as a good example of an adaptive radiation, where pervasive gene flow has been detected in spite of the rarity of hybrids found in nature ([Gilbert, 2003]; [Bull et al., 2006]; [Kronforst et al., 2006b]; [Mallet et al., 2007]; [Mallet, 2009]; [Pardo-Díaz et al., 2012]; [Heliconius Genome Consortium, 2012]; [Martin et al., 2013]). Nevertheless, hybridising species (and subspecies) of *Heliconius* remain distinct in sympatry and parapatry ([Mallet et al., 2007]; [Mallet, 2009]). Are high hybridisation rates universal to entire clades or are they instead restricted mainly to groups of species within such

clades? When does hybridisation between diverging species stop? These are questions that remain unanswered.

The deleterious evolutionary consequences of hybridisation include environmentally mediated or intrinsic incompatibilities that lead to unviable offspring or affect their fertility, and/or the loss of genetic integrity ([Rhymer and Simberloff, 1996]). Hybrid unfitness has extensively been reported both in experimental and natural settings ([Hatfield and Schluter, 1999]; [McBride and Singer, 2010]). Moreover, there are special cases where hybridisation collapses the species boundaries. For instance, speciation reversal can happen between species maintained by a selection/migration balance, due to environmental changes that can have anthropogenic origin, like pollution or predators dynamics changes ([Seehausen et al., 1997]; [Taylor et al., 2006]) or species translocations that bring together otherwise allopatric species (e.g. biological invasions).

The rupture of the selection/migration equilibrium can also take the opposite direction. Namely, if disruptive selection (and sexual selection) overcomes the effect of gene exchange, if there is a progression towards larger genomic regions being protected from introgression ([Via et al., 2000]; [Wu, 2001]) or if reinforcement evolves ([Servedio and Noor, 2003]), reproductive isolation can be strengthened further and further between the hybridising taxa. In particular, reinforcement of reproductive barriers through selection of assortative (contypic) mating has attracted big attention as one of the favourable indirect effects of hybridisation on divergence, but has often been controversial ([Coyne and Orr, 2004]). Although empirical work shows the potential of reinforcement to generate increased prezygotic isolation ([Servedio and Noor, 2003]; [Coyne and Orr, 2004]), it is not clear if it can result in complete isolation (e.g. ([Bímová et al., 2011])).

Alternatively, if circumstances remain unchanged and the selection/migration balance lasts over time, hybridisation may not have an obvious impact on diversification among the parental taxa ([Barton and Hewitt, 1985]); existing differentiation may be maintained, but there may be no progress towards further divergence. If selection pressures are high enough, local adaptations may persist in spite of high levels of gene exchange ([Endler, 1977]). For instance, species-related morphological integrity can be maintained in the face of persistent genetic admixture ([Noor et al., 2000]; [Saint-Laurent et al., 2003]; [Kraus et al., 2012]). A good example of this scenario are sympatric species of ducks (family Anatidae), which hybridise very often in the wild producing viable and fertile offspring and thus exchange genetic material ([Kraus et al., 2012]), and still remain morphologically distinct (males species-specific plumage, ornamentation and courtship behaviour; see Figure 7). Another nice example is provided by hybridising *Heliconius* species and even parapatric races (e.g. ([Jiggins et al., 1996]; [Martin et al., 2013])). Analysis of putatively neutral markers have shown that species *Heliconius melpomene* and *H. cydno* show important levels of gene exchange ([Bull et al., 2006]; [Kronforst et al., 2006b]; [Martin et al., 2013]). [Martin et al., 2013] have uncovered the persistence of gene flow at several moments of the process of divergence. Impressively, these authors provided evidence that up to 40% of the genome in *H. melpomene* shows admixture with *H. cydno* or *H. timareta* in sympatry. This percentage of the genome clustered by geography rather than by species. However, species *H.*



*melpomene* and *H. cydno* remain “good” species with distinct ecologies and strong barriers to gene flow, including both pre and post-zygotic isolation ([Jiggins et al., 2001]; [Naisbit et al., 2002]; [Merrill et al., 2012]). Similar cases of distinctiveness despite a high incidence of introgression are found in several other groups of hybridising species or subspecies ([Noor et al., 2000]; [Saint-Laurent et al., 2003]), and among polymorphic populations of a number of species, including the Galapagos lava lizard *Microlophus albemarlensis* ([Jordan et al., 2005]) and the Dominican anole *Anolis oculatus* ([Stenson et al., 2002]).

As suggested so far, hybridisation can be a beneficial process that potentially allows populations to gain genetic diversity and to promote adaptation to new environments while either driving speciation or retaining their integrity. In that sense, [Mallet et al., 2007] suggested that hybridisation can act as a multilocus “macro-mutation” that results in big phenotypic shifts to colonise new adaptive peaks. In an extreme case of such a shift, hybrid speciation can happen virtually instantaneously if hybrids obtain the entire chromosomal set of both parental species (i.e. allopolyploid speciation; ([Spolsky et al., 1992]; [Soltis and Soltis, 2000])). Hybrid speciation can also occur with hybrids retaining the parental chromosome number (i.e. homoploid speciation). In such scenarios, the hybrid species is a separate cluster of genotypes that becomes stabilised and remains distinct when in contact with either parent ([Mallet et al., 2007]). Whereas polyploid speciation has widely been reported in plants, homoploid speciation has been much less documented, presumably due to the difficulty of detecting diploid hybrid species with certainty. There are, however, some species of animals and plants, which appear to have a homoploid hybrid origin. For instance, the species *Senecio squalidus* in the British Isles has resulted from hybridisation between *S. aethnensis* and *S. chrysanthemifolius* ([James and Abbott, 2005]). A putative example of a hybrid species is *Heliconius heurippa*, which is suggested to have formed from the interbreeding between species *H. melpomene* and *H. cydno* ([Mavárez et al., 2006]). The forewing of *H. heurippa* has an intermediary morphology between these parental species, which was accurately reconstructed by backcrossing F1 hybrids to *H. cydno*. Additionally, the resulting hybrids prefer approaching and courting models having their own colour pattern than the parental species’ patterns ([Melo et al., 2009]).

Whenever there are no fitness costs related to the introgressive events (e.g. maladaptive combinations of alleles), hybridisation can increase the genetic diversity of hybrids. For instance, populations of Darwin’s finches *Geospiza* show little evidence of poor genetic diversity, in spite of regular bottlenecks often caused by rainfall-related environmental extremes. This has been partially attributed to ongoing hybridisation among coexisting species ([Freeland and Boag, 1999]). Extremely high levels of genetic diversity presumably due to hybridisation have also been reported in plants (e.g. species in the genus *Mimulus*; [Sweigart and Willis, 2003]). Genetic diversity of hybrids can sometimes exceed the combined variation of the parental types (i.e. transgressive segregation; see ([Rieseberg et al., 1999]) for a review on this topic). In such circumstances, hybrids can be fitter than the parental types (i.e. heterosis or hybrid vigour), although this phenomenon is mostly observed in the first hybrid generation and decreases in later generations ([Grant and Grant, 1992]). For example, hybrids between Darwin’s finches *Geospiza* in Daphne Major Island display environment-dependent hy-

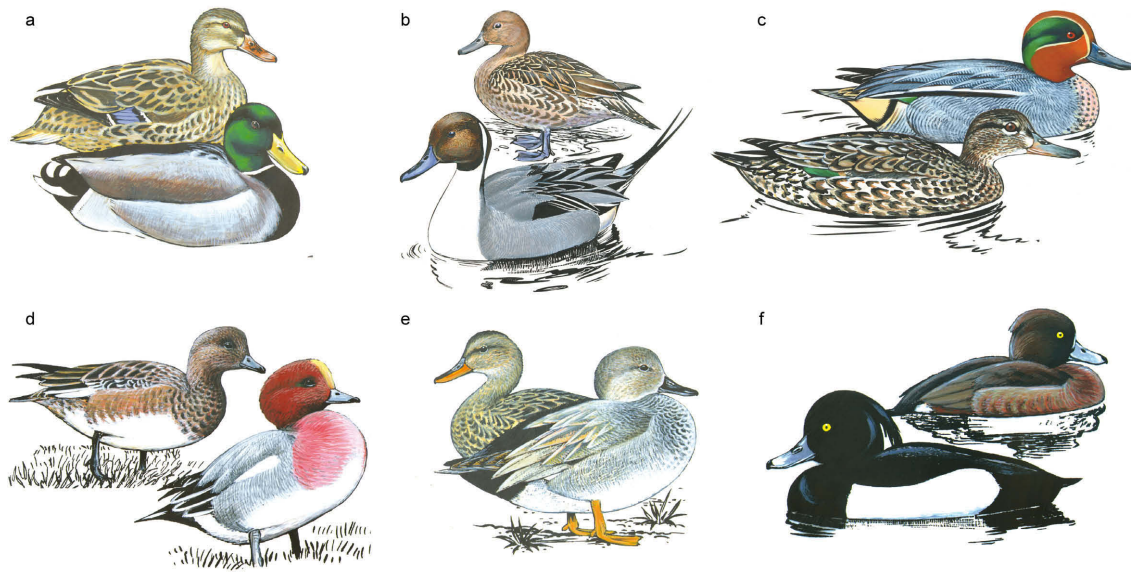


Figure 7 – **Distinctive plumage ornamentation among duck species despite high levels of interspecific hybridisation.** Male and female of each of the studied duck species: (a) *Anas platyrhynchos* (mallard), (b) *Anas acuta* (northern pintail), (c) *Anas crecca* (teal), (d) *Anas penelope* (Eurasian wigeon), (e) *Anas strepera* (gadwall), (f) *Aythya fuligula* (tufted duck). Figure taken and legend modified from [Kraus et al., 2012].

brid superiority ([Grant and Grant, 1992]).

Finally, interspecific hybridisation can permit the transfer of adaptively advantageous genes from one species to the other (i.e. adaptive introgression), potentially accelerating speciation. Adaptive introgression has been demonstrated in only a few convincing cases in plants and animals, sometimes in settings with significant levels of human intervention. Well known examples in mammals include the transfer of an allele that provides resistance to rodent poisons from the Algerian mouse *Mus spretus* to the partially sympatric house mouse *Mus musculus* ([Song et al., 2011]) and the spread of melanism in wolves and coyotes following admixture with domestic dogs ([Anderson et al., 2009]). In *Heliconius* butterflies, alleles of a wing colour pattern locus called *optix* also seem to have been transferred through hybridisation between closely-related species, triggering phenotypic convergence for mimicry ([Pardo-Diaz et al., 2012]; [Heliconius Genome Consortium, 2012]). More precisely, unidirectional introgression of genomic segments containing the gene *optix* from *H. melpomene* to *H. timareta* explains at least in part their resemblance for mimicry ([Pardo-Diaz et al., 2012]). In the tree topology of the genomic region containing the gene *optix*, the taxa are clustered by colour pattern rather than by species ([Pardo-Diaz et al., 2012]; [Heliconius Genome Consortium, 2012]).

### 0.2.2 The occurrence of hybrid zones

According to [Barton and Hewitt, 1985], hybrid zones are “narrow regions in which genetically distinct populations meet, mate, and produce hybrids” (see Box 4 for a classification of hybrid zones). The level of current hybridisation observed in a hybrid

---

**Box 4** Classification of hybrid zones
 

---

Hybrid zones are classified either by the extent of reproductive isolation existing between the hybridising populations or by their origin ([Barton and Hewitt, 1985]; [Harrison and Bogdanowicz, 1997]; [Jiggins and Mallet, 2000]). The first classification relies on the distribution of genotypic classes ([Harrison and Bogdanowicz, 1997]) and on the view that hybrid zones may constitute a continuum of distinctive levels of differentiation, which mirror speciation as a gradual process ([Jiggins and Mallet, 2000]). On the one extreme of this continuum, “unimodal” hybrid zones consist mostly of intermediate hybrid genotypes. On the other extreme, “bimodal” hybrid zones are predominantly composed by individuals genetically similar to one or the other parental genotype, with few intermediates (see [Jiggins and Mallet, 2000], for examples). In other words, unimodality suggests random mating and complete hybrid viability, whereas bimodality indicates that a degree of reproductive isolation has developed and speciation is partially complete ([Harrison and Bogdanowicz, 1997]; [Jiggins and Mallet, 2000]), representing early and advanced stages of differentiation, respectively. The second classification of hybrid zones separates “primary” from “secondary” contact hybrid zones. The former ones result from *in situ* divergence in the face of gene flow, thanks to selection towards opposite ends of an environmental gradient. In contrast, the latter ones result from secondary contact between populations having diverged during a period of spatial isolation. Given that in the former case, differentiation has occurred in the face of gene flow (i.e. sympatric or parapatric speciation models) while in the latter case, it occurred in allopatry, distinguishing between these alternative scenarios informs about how the speciation process takes place ([Barton and Hewitt, 1985]). It has been suggested that most reported hybrid zones are secondary ([Barton and Hewitt, 1985]). For instance, notable examples of hybrid zones like the ones between the toads *Bombina bombina* and *B. variegata* ([Szymura and Barton, 1991]; [MacCallum et al., 1998]), or between the grasshoppers *Chorthippus parallelus parallelus* and *C. p. erythropus* ([Butlin and Ritchie, 1991]), represent suture zones between Pleistocene refugia in Central and Northern Europe.

---

zone between lineages somehow reflects the level of differentiation and hence, of the stage of speciation attained between these lineages. Therefore, hybrid zones have been referred to as “natural laboratories” ([Harrison, 1993]) where to explore how taxa diverge and speciate, or how their identity is maintained while exchanging genes to some extent. In particular, hybrid zones allow investigating how the distinct kinds of reproductive isolating barriers evolve and participate in diversification.

The structure of a hybrid zone can be predicted by using a theoretical framework called cline theory (see Box 5 for a more detailed description of the concepts and uses of cline theory). This framework also allows modelling hybrid zone dynamics and inferring the relative contribution of natural selection and dispersal to their structuring, based on the measurement of the width, slope and concordance of (phenotypic or genotypic) clinal variation in hybrid populations ([Mallet and Barton, 1989a]; [Barton and Hewitt, 1989]; [Barton and Gale, 1993]). In particular, assessing comparatively the width of multi-locus clines allows investigating differential introgression, pinpointing adaptive loci and characters, and comparing the strength of selection among distinct traits in the face of introgressive hybridisation. Finally, the inference of the factors contributing to the maintenance of hybrid zones also benefits from applying the rationale behind cline theory.

---

**Box 5** Using cline theory for characterising a hybrid zone and the selective pressures on adaptive traits diverging across it
 

---

Cline-based analyses have been used to evaluate the strength of selection operating on a particular trait or locus, and can be a first step to uncover local adaptations and their genetic basis. For such analyses, allele frequencies are measured at distinct localities and compared to the geographic position across a hybrid zone. The width of the cline (which is the inverse of its maximum slope) is inversely proportional to the strength of selection, following the simplified equation  $\sigma/\sqrt{s}$ , where  $\sigma$  is equivalent to dispersal and  $s$  is selection. Another useful variable is the cline centre, which is the geographic position where the maximum allele frequency gradient is observed. The width and the centre of a cline can be statistically compared to these variables for other clines to investigate the hybrid zone dynamics. Different clines are said to be “consistent” (or “congruent”) if they have the same centre and “concordant” if they have virtually the same width.

Multi-locus cline consistency informs about the history of the hybrid zone. Cases of consistent clines like, for instance, at the hybrid zone between fire-bellied toads *Bombina bombina* and *B. variegata*, for allozymes, mtDNA, mating call and colour patterns ([Szymura and Barton, 1986]) have been interpreted as the result of secondary contact after allopatry; such an interpretation is stronger if some of the characters or loci are known to be neutral ([Barton and Hewitt, 1981]). In contrast, coincident clines are not expected in primary hybrid zones, unless alleles at the distinct loci respond to the environment in the same way. Also, clinal variation for neutral loci is not expected under parapatric divergence, unless neutral loci are linked to selected loci. Multi-locus cline concordance, on the other hand, does not inform about the history of divergence, but about the current ecology and in particular, about patterns of introgression expected to respond to selection. As explained above, remarkable variance in introgression across the genome is often observed; regions under selection are less permeable to gene flow than neutral regions. Thus, adaptive traits show clinal variation whereas neutral regions do not. Finally, the strength of selection acting on a given trait is inversely proportional to the width of the cline for that trait.

---

### 0.3 What is the genetic architecture of adaptation?

*It is almost impossible with any brevity to exemplify the notion of adaptation. Just because adaptation consists, even in the simplest cases, in a multiplicity of correspondences between one sufficiently complicated system, the organism itself, and another equally complicated, the environment in which it finds itself. It is, indeed, just this multiplicity that makes the thing recognizably adaptive.*

Fisher 1934. Taken from [Orr, 1998]

Understanding the link between diversification and genetic structure allows getting a better understanding of the speciation process. In contrast to the availability of ecological data showing the importance of selection as a main force for population divergence and speciation, genetic data underlying these processes is still scarce ([Linnen and Hoekstra, 2009]; [Butlin et al., 2012]). How does ecological and phenotypic divergence occur at the genetic level is a question that remains largely unsolved.

Ecologically-important traits are usually complex and the sets of loci and their relationships underlying this complexity constitute the genetic architecture. The term “genetic architecture” encloses the number and genomic position of the loci contributing to a trait, their effect size, the way their alleles interact (i.e. dominance), the way the loci interact to modulate each others’ contributions (i.e. epistasis and additivity), the way they interact with the environment (i.e. epigenetics) and the contribution of a single locus to distinct traits (i.e. pleiotropy).

Genetic architecture influences the potential of a trait to vary, and thus to evolve (see [Hansen, 2006]). This is why the study of genetic architecture has been central to many biological questions such as the understanding of biological diversification, of diseases, and of plant and animal breeding (e.g. [Rogers et al., 2013]). Although unravelling individual genetic variants underlying complex traits has shed important light on the control of phenotypic variation, the understanding of gene action and trait evolvability can only be fully assessed in its architectural context. Moreover, unveiling the genetic architecture of complex traits is often the first step towards the identification of the trait genes and trait nucleotide variants, which ultimately determine trait variation.

As stated above, the genetic architecture defines how the trait value is determined by the genotype (i.e. the genotype-phenotype map). Most population genetic models rely on the assumption that the genotype-to-phenotype map remains unchanged over short time scales, and thus, that the genetic architecture of a trait is constant, and selection only acts on allele frequencies. However, several simulation studies have shown that genetic architecture is expected to evolve when any of its components changes (e.g. [Wagner and Altenberg, 1996]). Such findings have, for example, suggested that gene effects, and not only gene frequencies, are evolutionary variables ([Hansen, 2006]). The evolution of genetic architecture has been proposed to have an important influence on the evolution of recombination ([Azevedo et al., 2006]) and on reproductive isolation and speciation ([Fierst and Hansen, 2010]).

The evolution of components such as dominance and epistasis has been largely explored theoretically but to a lesser extent empirically ([Lande, 1980]; [Barton, 1995]; [Coyne and Orr, 1998]; [Carlborg and Haley, 2004]; [Kroymann and Mitchell-Olds, 2005]; [Carter et al., 2005]; [Hansen, 2006]; [Phillips, 2008]). Here, our work focuses mainly on the number of loci contributing to a complex trait and the distribution of their effect size. In the following lines I will outline the most influential theoretical work regarding these components of genetic architecture and will not give many details regarding dominance and epistasis. A central question to this theoretical framework is how many loci are involved in adaptation: few or many? Here, I will focus on the model by [Fisher, 1930], which has been extensively used to describe how the genetic architecture of a trait can determine the access to a fitness optimum. This model postulates that a quantitative trait is encoded by many loci with small, additive effects. Fisher’s pioneer geometric model of evolution settled the first principles of character evolution, but made some assumptions (like treating complex gene interactions like statistical noise) that limited the subsequent research on the evolution of genetic architecture ([Hansen, 2006]). Fisher concluded that evolution is fuelled by very small mutations (consistent with the micromutational or infinitesimal views), and that factors

of large phenotypic effect have little or no importance in adaptation. More recently, however, Robertson ([Robertson, 1967]) proposed a nearly exponential distribution of allelic effects, where most phenotypic change is driven by a few loci with moderate to large effect, and an increasingly larger number of loci with increasingly smaller effects making up the remainder. Over a single bout of adaptation towards a local optimum, Orr ([Orr, 1998]) also found an exponential distribution of gene effect sizes. The adaptive walk proposed by Orr ([Orr, 1998]) seems to match the theoretical expectations about the distribution of gene effect sizes fixed during mimicry evolution ([Baxter et al., 2008a]). More precisely, it fits the widely accepted “two-step” model proposed to explain the adaptive peak shift during the evolutionary transition to a new mimicry ring ([Turner, 1985]). Under this model, a large mutational step (step 1) contributes to get to the “slopes” of a new adaptive peak (the approximate resemblance to a locally protected mimicry ring). Subsequently, several small mutational steps allow climbing up to the phenotype optima (the perfection of resemblance between the co-mimics). The effect of migration was not taken into account in the model by Orr ([Orr, 1998]) but was proposed by Griswold ([Griswold, 2006]) to skew the distribution of gene effects towards larger effect sizes during an adaptive walk.

A largely unsolved question is how the genetic architecture of an adaptive trait is shaped by factors such as selection and migration. What determines how many loci are involved in adaptation? Here again, models trying to predict the evolutionary trajectory have been developed. [Rajon and Plotkin, 2013] have developed a population genetics model for the evolution of the number of loci controlling a trait. Their model predicts that relatively few loci are expected to contribute to a trait under weak or strong selection, but many loci with variable effects encode a character under moderate selection. Under a migration-selection balance, highly concentrated architectures of adaptation were found, with few and tightly linked functionally variants of large effect ([Yeaman and Whitlock, 2011]). [Kopp and Hermisson, 2006] on the other hand, found that frequency-dependent disruptive selection also favours the evolution of an asymmetric genetic architecture, with most of the genetic variation concentrated on a small number of loci. The message coming out of these results, as pointed by [Kopp and Hermisson, 2006] is that genetic architecture is under strong selection, and that neither the details of the genetic architecture nor its potential to evolve should be ignored. However, scant empirical data exists allowing to test these models and more generally, allowing to understand the evolution of the genetic architecture of adaptive traits.

I will next present how distinct approaches, and in particular Quantitative Trait Locus (QTL) mapping, have come into play to help testing the mentioned predictions. I will mention outstanding cases where the genetic architecture of adaptive traits has been unraveled, making emphasis on wing colour patterns for mimicry in *Heliconius*. Finally, I will present some data aiming at questioning the repeatability of evolutionary paths at the genetic and molecular levels.

### 0.3.1 Unravelling the genetic architecture of adaptive traits

In a review about the genetic theory of adaptation, Orr ([Orr, 2005]) has emphasised that there is a big gap between the rich body of mathematical theory on phenotypic

evolution and the large and growing body of data on the genetic basis of adaptation. In concordance to this statement, it is still not possible to make any general conclusions about the genetic architecture of adaptive traits, about its response to factors such as selection and migration, and about the relative importance of additive and non-additive variation.

Quantitative Trait Locus (QTL) mapping ([Lander and Botstein, 1989]) put in evidence a broad empirical diversity of observed genetic architectures of complex traits, which range from Fisherian to Mendelian. Pictures of Fisherian genetic architectures have been mainly detected in agriculturally and medically important complex characters, as well as for traits in model species ([Laurie et al., 2004]; [Schön et al., 2004]; [Ehrenreich et al., 2009]). For instance, in *Arabidopsis*, the gene network controlling flowering time consists of more than sixty genes ([Ehrenreich et al., 2009]). Cases have been reported where traits believed to rely on few large effect loci have been shown to have a substantially more complex genetic architecture than previously thought. These inconsistencies, believed to be due to experimental constraints such as sampling size and other factors affecting statistical power, have been observed for traits such as seed oil content ([Laurie et al., 2004]) and grain yield ([Schön et al., 2004]) in maize, and for numbers of *Drosophila* sensory bristles ([Dilda and Mackay, 2002]). Hence, it is important to keep in mind that the gene effect might sometimes be so small that researches might lack power to detect them.

Notably, a myriad of instances rejecting the long accepted Fisherian expectations have been found, in particular for traits that are fitness-related or function as barriers to gene flow. More precisely, genetic analysis of phenotypic variation between populations or species have frequently revealed several traits controlled by a small number of genes of major effect and their interactions (e.g. [Tanksley, 1993]; [Kearsey and Farquhar, 1998]; [Reed et al., 2011]). Adaptations having simple modes of inheritance include coat colour variation in mouse species, colour polymorphism in the moth *Biston betularia*, trichome variation and resistance to salinity in *Arabidopsis lyrata*, armour-plate reduction and pelvic reduction in the threespine stickleback *Gasterosteus aculeatus*, floral characters in *Mimulus* flowers, the elements of wing colour patterns in *Heliconius* butterflies, among others ([Bradshaw et al., 1995]; [Cresko et al., 2004]; [Shapiro et al., 2004]; [Colosimo et al., 2004]; [Joron et al., 2006]). Despite these findings, many QTL of minor effect, whose detection is possibly limited by statistical limitations, may also contribute to adaptation.

Among the mentioned adaptive traits, *Heliconius* butterfly wing patterns stand out because of their complexity, which contrasts with their apparently simple genetic basis. Big efforts have been made on identifying the genes accounting for mimicry in this clade (see Box 6), since this can reveal what is the raw material for adaptation and how selection acts on this material. In distinct species in the genus *Heliconius*, a few major loci of simple Mendelian inheritance have been recurrently found to control separate wing colour pattern elements; variants of these elements appearing together on the wing generate a given mimetic pattern. Since Chapter 1 provides details about the colour loci (see in particular Table 1.1 and Figure 1.4) and Box 6 gives details about the approaches used to discover them, I will not give any details here. Briefly, the *Heliconius* colour-pattern *toolbox*, how it has been called, is distributed along 4 of

the 21 chromosomes characterising the genus. Although mimetic patterns are usually controlled by a multilocus architecture dominated by a low number of genes of major effects, loci of minor and continuous effect have also been detected, thanks to a few attempts made to map quantitative colour pattern variation ([Baxter et al., 2008a]; [Jones et al., 2011]; [Papa et al., 2013]). This distribution of colour gene effect sizes matches the expectations of the two-step model predicting the colonisation of a new mimicry ring (see above), especially because minor effect loci influence the major colour elements controlled by the larger effect loci, presumably improving the resemblance for mimicry ([Papa et al., 2008]; [Jones et al., 2011]). Interestingly, this multilocus architecture, although being found in most *Heliconius* species explored so far, is not generalised.

Wing coloration in *H. numata*, a tiger-patterned species which belongs to the silvaniform or *numata* sub-clade (see Figure 4) is almost entirely controlled by a *supergene*, which is a cluster of tightly linked genes ([Joron et al., 2006], [Joron et al., 2011]). The *supergene P*, how it is named, is homologous to one of the genomic regions present in the multilocus *toolbox* ([Joron et al., 2006]). This completely different genetic picture of adaptation underlies an impressive local polymorphism restricted to *H. numata*, which is absent or very limited in other *Heliconius* species. More precisely, *H. numata* has multiple coexisting colour morphs, each of them mimicking a different species of *Melinaea*, a genus in the Ithomiinae subfamily (see Figure 14A; [Brown and Benson, 1974]). Each morph is determined by a specific allele of the supergene with precise allelic dominance (Figure 14A), and maintains its phenotypic integrity thanks to the co-action of strongly linked genetic elements. This strong linkage is due to the suppression of recombination owing to specific chromosomal rearrangements within a ~400 Kb region (Figure 14B). In consequence, distinct haplotypes at the supergene and inversion breakpoints are in complete association with colour morphs ([Joron et al., 2011]; see Figure 14B). The variation of the genetic architecture of wing patterning in the genus *Heliconius*, ranging from a nearly symmetric architecture (with loci having similar gene effects) in *H. melpomene*, *H. erato* and *H. cydno*, to a highly asymmetric architecture (with largely unequal gene effects) in *H. numata*, makes it possible to explore which factors play a role in the evolution of the genetic architecture of complex traits.

Our knowledge on the molecular basis of adaptive variation remains limited, because the identification of the trait genes and ultimately the nucleotide variants is clearly more difficult than the identification of QTLs. The genetics of speciation have often been addressed by combining QTL mapping with other approaches such as candidate genes, multilocus genome scans, functional assays and molecular scans ([Nielsen, 2005]; [Vasemägi and Primmer, 2005]; [Noor and Feder, 2006]; [Stinchcombe and Hoekstra, 2007]). This has allowed identifying some specific genes (and in rarer cases, specific nucleotides) underlying characters under selection ([Colosimo et al., 2005]; [Hoekstra et al., 2006]), including two of the genes controlling wing patterns in *Heliconius* (called *optix* and *WntA*; [Reed et al., 2011]; [Martin et al., 2012]; see Box 6 for more details).



---

**Box 6** A brief history of the mapping of wing colour pattern genes in *Heliconius* butterflies

The exploration of the genetics of wing colour pattern in *Heliconius* began formally in the 1950s and became increasingly feasible with the arrival of the technologies necessary to access and analyse genetic information. A classical forward genetics approach (which starts from the phenotype and goes towards the genetics causing it) has been used to disentangle the genetic basis of these adaptive traits. Central to this was the development of methods for rearing *Heliconius* in captivity, thus opening the way for diverse genetic experiments. Intra-specific crosses between geographic races differing in wing pattern revealed that colour pattern was under the control of a few loci with Mendelian inheritance in several species ([Sheppard et al., 1985]). The development of molecular markers such as Amplified Fragment Length Polymorphisms (AFLPs), Restriction Fragment Length Polymorphisms (RFLPs), microsatellites and single-copy nuclear loci (SCNL), facilitated linkage mapping to localise the regions controlling colour in *H. melpomene*, *H. erato* and *H. cydno* ([Tobler et al., 2004]; [Jiggins et al., 2005b]; [Kapan et al., 2006]), which are species belonging to two major clades within *Heliconius*. These studies targeted a handful of genes of major effect which were found to be located on four of the twenty ones *Heliconius* chromosomes and which are all properly described in the text and in Table 1.1 of Chapter 1. With the ultimate aim of uncovering the genetic changes underlying these complex phenotypes, major efforts have been made to get closer to the identification of the sites that are causally involved in the variation of the pattern elements in *Heliconius*. A crucial step towards the colour genes in *Heliconius* was the creation of Bacterial Artificial Chromosome (BAC) libraries, and the identification of BAC clones which could be assembled into chromosomal walks across key regions, helping to define candidate genes within them ([Ferguson et al., 2010]; [Baxter et al., 2010]; [Counterman et al., 2010]). Additionally, population genetics studies have provided molecular signals of selection on the candidate genes. When estimating gene flow and population differentiation between different races of the same species, the candidate colour regions show higher levels of structuring than the markers in the rest of the genome ([Baxter et al., 2010]; [Joron et al., 2011]). Also, genotype-by-phenotype association tests on wild specimens confirmed the strong association between some genes within the regions of interest and the patterns they control for ([Counterman et al., 2010]; [Reed et al., 2011]). These results, together with functional analyses of gene spatial expression, have led to the identification of two of the genes responsible for important elements in colour variation across the genus. The first is a morphogen homologous to *WntA* in *Drosophila*, that determines the size and position of melanic patterns around the discal cell of the forewing, which was attributed to the loci called *Ac* and *Sd* in a wide range of geographic races of *H. melpomene*, *H. cydno* and *H. erato* and also has important effects in *H. himera*, *H. sara* and *H. atthis* ([Martin et al., 2012]). The second is a transcription factor, homologous to the *Drosophila* gene *optix*, which prefigures the variety of red wing elements controlled by loci named *B-D/D/G-Br* in *Heliconius melpomene*, *H. erato* and *H. cydno* ([Reed et al., 2011]). The genes *WntA* and *optix* both govern conspicuous elements in wing patterning and drive the convergence for mimicry between *H. melpomene* and *H. erato*, as well as the divergence of these species into many distinct geographic races.

---

### 0.3.2 What is the genetic basis for phenotypic convergence?

The unification of the terms *convergence* and *parallelism* to the term convergent evolution ([Arendt and Reznick, 2008]; [Manceau et al., 2010]) has proved useful in several cases ([Manceau et al., 2010]; [Nadeau and Jiggins, 2010]; [Losos, 2011]; [Martin and Orgogozo, 2013]), particularly when the adjectives *genetic* and *phenotypic* are also

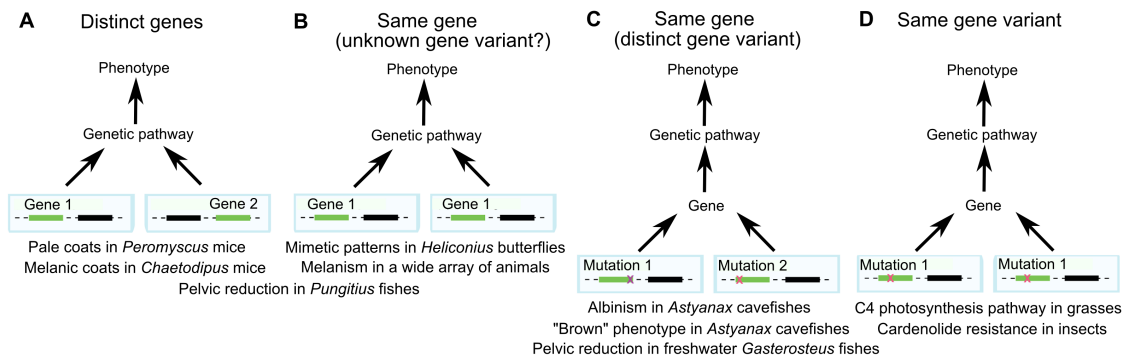


Figure 8 – **Molecular patterns at multiple levels underlying phenotypic convergence.** Similar phenotypes can evolve by (A) distinct genes, (B) the same gene but unknown causal mutations, (C) different mutations in the same gene or (D) the same mutation in the same gene. Some cases included and explained in more detail in the text are mentioned above the molecular patterns they exemplify. Figure modified from [Losos, 2011]; [Elmer and Meyer, 2011]).

adopted ([Elmer and Meyer, 2011]). The term phenotypic convergent evolution refers to the independent evolution of the same features in different evolutionary lineages (independently of the phylogenetic distance between them), typically in response to similar environmental challenges. Such phenotypic convergence has been observed in a wide array of organisms ([Brown, 1981]; [Losos, 1992], p. 199; [Colosimo et al., 2005]; [Rogers and Bernatchez, 2007]; [Arendt and Reznick, 2008]) and has been suggested to be shaped by natural selection ([Endler, 1986]; [Schluter, 2000]; but see [Losos, 2011] for a nice discussion about the variable role of selection on convergence). As described above, mimicry is one of the most outstanding examples of phenotypic convergence shaped by selective forces. Natural systems displaying phenotypic convergence provide ideal scenarios to investigate whether this resemblance involves or not the same genetic mechanisms (i.e. the developmental genetic pathways, loci, genes and genetic variants), thus leading to a better understanding of the molecular underpinnings of adaptive evolution. Convergent phenotypes can occur by the evolution of distinct molecular patterns across taxa: distinct genes, the same gene but unknown causal mutations, different mutations in the same gene and the same mutation in the same gene (see Figure 8).

To what extent the evolutionary process is repeatable, and therefore predictable, is an important current question. Is there a common genetic basis for convergence between distinct taxa? At which functional level does genetic convergence occur? Are these genetic similarities the result of parallel evolution or collateral evolution? Does convergent evolution more often involve standing genetic variation or new mutations?

Based on an explosion of research into these issues, the answer to the first question, on which I will mainly focus here, is not a clear *yes* or *no*, but rather *sometimes*. In cases where convergence results from similar molecular mechanisms, such mechanisms can occur through distinct historical patterns: by parallel evolution or by collateral evolution, the latter being the product of shared ancestry or of hybridisation ([Stern, 2013]). Nice reviews questioning the evolutionary repeatability or versatility have been written ([Arendt and Reznick, 2008]; [Manceau et al., 2010]; [Elmer and Meyer, 2011]; [Kronforst et al., 2012]; [Martin and Orgogozo, 2013]; [Stern, 2013]).

Before describing a little tip of the increasingly accumulating empirical data aiming at answering the questions presented in the last paragraph, I will highlight some issues that should be kept in mind. A first thing to focus the attention on is that finding a convergent genetic basis can sometimes be the result of focusing on candidate genes, thus enhancing ascertainment biases and precluding the discovery of other possibly evolutionary important genes. One solution to this problem is to complement gene “hunting” with unbiased experimental approaches such as genetic mapping. A second thing to take into consideration is that determining whether or not the same phenotype is defined by the same genetic basis clearly depends upon the mechanistic level that is analysed: phenotypic convergence mirrors genotypic convergence more often at higher functional levels (e.g. at the gene level) than at finer levels (e.g. at the nucleotide level). For example, it could be stated that repeated complete pigment loss (i.e. albinism) in cave populations of *Astyanax mexicanus* fishes has a homologous genetic basis, because it involves the same gene (see below for more details). This situation is presumably very common when the causal mutations underlying a phenotype are unknown. Alternatively, we could affirm that convergent albinism in different populations has not a common molecular basis, given that actually, it involves distinct mutations of the same gene ([Protas et al., 2006]).

The observation that some closely-related taxa can use distinct genetic mechanisms for the control of similar phenotypes highlights the ability of natural selection to build specific structures from distinct starting points ([Losos, 2011]). The use of different genetic solutions to solve similar ecological problems might be favoured for phenotypes potentially attained via many different pathways, such as pigmentation-related traits. Examples include adaptive melanisation in different populations of the pocket mice *Chaetodipus intermedius*. The increased pigmentation in mice populations living in geographically distant lava flows in the United States, which provides cryptic protection, is attributed to the gene encoding the melanocortin-1 receptor (*Mc1r*) in one population in Arizona but to an unknown gene in New Mexico populations ([Hoekstra and Nachman, 2003]). A similar study case involves the evolution of light coats in oldfield mice (*Peromyscus polionotus*). Pale mice inhabit sand dunes on Florida’s Gulf Coast and benefit from reduced predators attack by crypsis on the substrate colour ([Vignieri et al., 2010]). Such lightly-coloured phenotype was found to be mainly controlled by a single amino acid mutation in the gene *Mc1r* ([Hoekstra et al., 2006]). Surprisingly, in a different population of pale mice occupying the dunes of Florida’s Atlantic Coast, variants of *Mc1r* were not involved in the control of pale coats.

Phenotypic convergence implicating disparate taxa and having a similar genetic basis have also been reported, and have been suggested to support the idea that genetic evolution displays some predictability ([Stern, 2013]), and that constraints may exist on adaptive evolution ([Losos, 2011]). A remarkable example of such a scenario is the evolution of dark or pale coloration in a wide variety of animals including birds, felids, lizards, the black bear, woolly mammoths and mice ([Theron et al., 2001]; [Eizirik et al., 2003]; [Rosenblum et al., 2004]; [Mundy, 2005]; [Römpler et al., 2006]; [Hoekstra et al., 2006]). Interestingly, variation of these melanin-based traits in all these far-related taxa is controlled by the same gene *Mc1r*. Another nice example of coupled phenotypic and genetic convergence at a high taxonomic level comes from multiple

species belonging to four orders of insects, which have evolved cardenolide resistance to feed on poisonous milkweed plants through precisely the same substitution at one amino acid position in the  $(Na^+ + K^+)ATPase$  gene ([Dobler et al., 2012]; [Zhen et al., 2012]). Molecular convergence in the latter example was driven by parallel evolution (c.f. [Stern, 2013]). Other cases of parallel evolution at the gene function level have been reported between far-related taxa: loss of larval trichomes in *Drosophila sechellia* and *D. ezoana*, resulting from regulatory changes in the gene *shavenbaby* ([Sucena et al., 2003]; [Frankel et al., 2012]; although the adaptive basis of this phenotype remains to be established). The kind of findings I have mentioned in this paragraph pushed researchers to suggest that “all genes are not equal in the eyes of evolution” ([Stern and Orgogozo, 2009]). More precisely, the evolution of similar traits in different taxa often involves changes in the same genes (“gene reuse”); hence, *hotspot* genes (c.f. [Stern and Orgogozo, 2009]) and *hotspots of evolution* (c.f. ([Martin and Orgogozo, 2013])) seem to exist that “lead the way” of an important portion of the evolutionary events.

Several cases of convergent phenotypic evolution show the combination of evolution at homologous genes and also non-homologous evolution involving distinct genes. In the Mexican cave tetra, *Astyanax mexicanus*, multiple populations have pairs of surface-dwelling morphs and cave-dwelling morphs ([Strecker et al., 2012]) the latter associated with pigment and eye loss. Complete loss of pigment in multiple populations has resulted from different deletions in the same gene, the ocular albinism 2 (*Oca2*) coding region ([Protas et al., 2006]). Also, in two of these albino cave populations, partially reduced pigmentation (another phenotype, the “brown” phenotype) has been attributed to different mutations in the gene *Mcl1r* and to an undetermined mutation at a different gene in a third population ([Gross et al., 2009]). These observations suggest that although the same gene is implicated in both cases, complete or partial loss of pigment in cavefishes has occurred independently in the distinct populations. Another impressive example of repeated evolution is the repeated reduction of armour-plate and pelvic structures in lake populations of the threespine stickleback, *Gasterosteus aculeatus*, contrasting with the heavily armoured marine forms of this species. The gene *Pituitary homeobox 1* (*Pitx1*) has been identified as the major effect locus on pelvic reduction in freshwater populations from across the globe ([Shapiro et al., 2004]; [Coyle et al., 2007]). It was confirmed that deletions in the *cis*-regulatory region upstream of *Pitx1* were implicated in pelvic reduction ([Chan et al., 2010]). This deletion has variable sizes (suggesting that it arose from independent events) and was found to be important in several independent freshwater populations. Interestingly, *Pitx* also controls the pelvic reduced phenotype in some ninespine stickleback populations, a clade diverged around 13 My ago from a common ancestor to ninespine and threespine sticklebacks (Figure 9), although variation at a different unknown gene seems to be responsible for this phenotype in ninespine sticklebacks from Alaska ([Shapiro et al., 2009]). Also, a derived allele at the gene *Ectodysplasin* (*EDA*) has caused reduced armour plating in different threespine stickleback populations, also in response to the repeated adaptation of fish to freshwater ([Colosimo et al., 2005]; [Jones et al., 2012]). In most populations, this derived phenotype resulted from fixation of the same ancestral allele of *EDA* (i.e. collateral evolution through shared ancestry) and in one case from de novo evolution of an *EDA* allele ([Colosimo et al., 2005]).

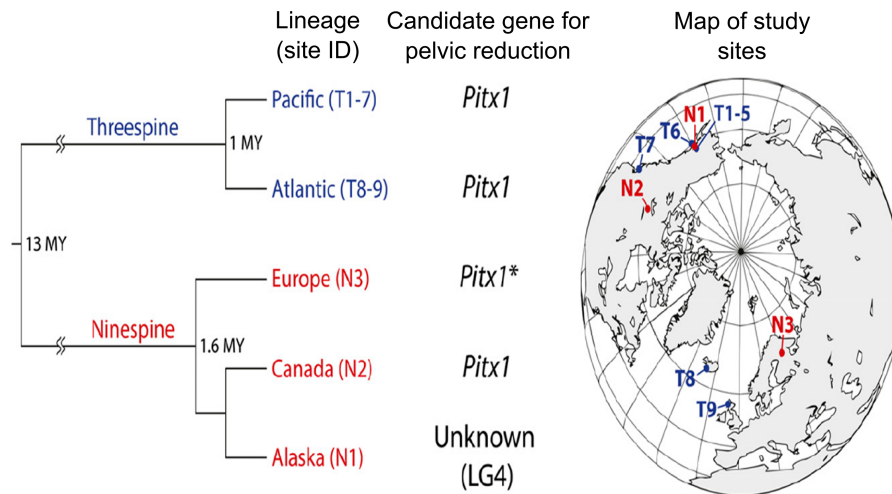


Figure 9 – **Phylogenetic relationships and candidate genes for pelvic reduction in threespine and ninespine sticklebacks.** *Pitx1* controls this convergent phenotype in all but one studied populations. Figure taken and legend slightly modified from [Shikano et al., 2013].

The latter paragraph encompasses examples of phenotypic convergence among relatively closely-related taxa, some of which resulted through changes in the same genes but did not involve the same mutations. The genetic basis of convergence for mimicry within the genus *Heliconius* has not been yet solved to the functional mutation level. However, some interesting findings have been gathered so far. Crosses between closely related species have shown that the same wing colour pattern loci in the “toolkit” not only operate within species but also underlie phenotypic divergence between species ([Jiggins and McMillan, 1997]; [Gilbert, 2003]; [Naisbit et al., 2003]; [Kronforst et al., 2006a]). Comparative mapping work across distinct species, including *H. erato* and *H. melpomene*, which belong to sub-clades split from a common ancestor ~10 My ago (see Figure 4) showed that loci affecting similar wing pattern elements routinely mapped to the same position on homologous chromosomes ([Joron et al., 2006]; [Kronforst et al., 2006a]; [Chamberlain et al., 2011]). Convergence in red pattern elements making up the mimetic patterns is the best understood at the genetic level, with recent expression and association data pointing to the gene *optix* as the causal gene of red patterning in several *Heliconius* species ([Reed et al., 2011]; [Martin et al., 2014]). Compelling evidence for collateral evolution through species hybridisation (*c.f.* [Stern, 2013]) has been provided for the sharing of *optix* alleles between *H. melpomene* and *H. timareta*, and with the further related species *H. elevatus* ([Pardo-Diaz et al., 2012]; [Heliconius Genome Consortium, 2012]). The second mimicry gene identified so far, *WntA*, is also responsible for a convergent pattern element throughout the genus ([Martin et al., 2012]). Thus, the current picture we have is one of a conserved and shared “toolkit” of genes across the genus, underlying both convergence for mimicry and an astonishing diversity of wing patterns.

## 0.4 How does our work contribute to answer largely unresolved questions?

More than one and a half centuries of research on the genus *Heliconius* has brought together several domains of knowledge, including ecology, evolutionary biology, genetics, population genetics and genomics, among several others. Chronologically speaking, there seems to be a point of inflection on the type of questions addressed concerning this group, which coincides with the development of genetic tools. A first bulk of data is related to the biology and ecology of the clade and a second period of research has mainly focused on evolutionary genetics. Brown ([Brown, 1981]) summarizes the early stages of research on this clade. I will mention some of the topics included in that review to briefly highlight some general features of our study system (see also [Beltran et al., 2007] for a good description of some of these features). The study of the geographic variants in the genus and proposals for their taxonomic relationship started in the nineteenth century (and continues until today). Simultaneously, ecological approaches have been used to study the biology and life history of different *Heliconius* species and to examine the co-evolution of these butterflies and their host plants in the genus *Passiflora*. *Heliconius* larvae feed on these plants and get cyanogenic glycosides that make them distasteful to birds. The particular diet of adult butterflies, which consists of amino-acids taken from pollen complementing nectar as a glucose source, was reported as one of the reasons for the long reproductive period of adult *Heliconius*. Efforts have been made to investigate clade-specific behaviours like this pollen-feeding behaviour and nocturnal communal roosting. In addition, odour production has also been explored and proposed to be important for chemical communication and as a defence against predators.

Although fruitful research has continued on these distinct topics and on several other domains (e.g. [Mérot et al., 2015]; [Beltran et al., 2007]; [Estrada et al., 2011]; [Llaurens et al., 2014]; [Rosser et al., 2014]; [Le Poul et al., 2014]; [Kozak et al., 2015]), the exploration of the genetics of wing colour pattern in *Heliconius* has taken comparatively more strength. As it could be seen above, in Box 6, a long list of studies has aimed at determining the genetic basis of these adaptive traits. In the last years, we entered a time where an important part of the evolutionary questions is answered in the light of genomics. The research linked to *Heliconius* is not an exception. The genome of *Heliconius melpomene* was published three years ago ([Heliconius Genome Consortium, 2012]) and has opened a door into the symbiosis between genomics and traditional approaches used for mapping colour genes, for exploring their mechanisms, the selective forces acting on them and their role in speciation with gene flow, for determining the contribution of introgression of adaptive alleles for mimicry, etc. Our project takes advantage of this combination of approaches to contribute to the understanding of the genetic architecture of colour patterns and the implication of these and other traits at distinct stages of diversification, with special emphasis on species in the silvaniform clade (Figure 5).

Research on the genus *Heliconius* has mainly focused on species within the far-related sub-clades *melpomene* and *erato*, which belong to the non-pupal-maters and pupal-maters clades, respectively, two main subdivisions of the genus (Figure 4; see

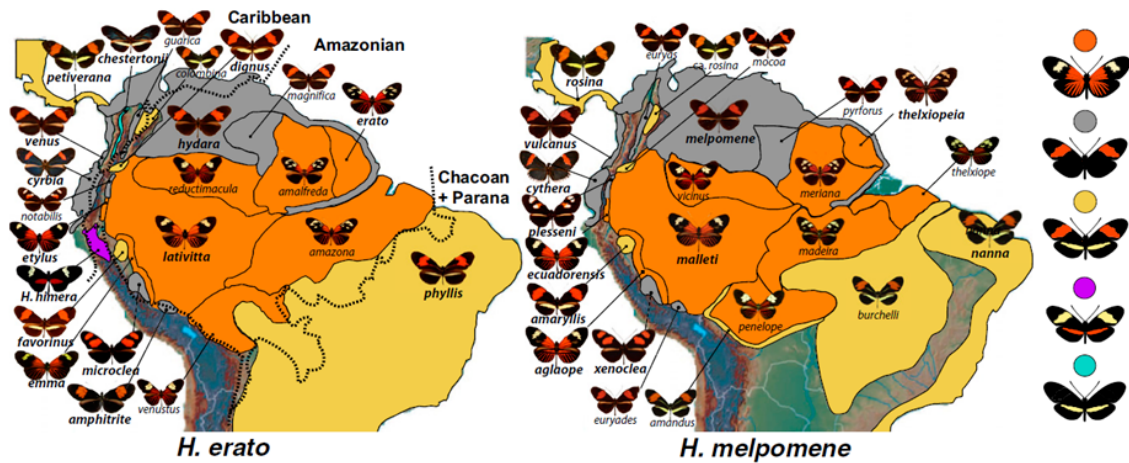


Figure 10 – The parallel radiation of mimetic colour-pattern complexes in *Heliconius erato* and *H. melpomene*. Each region demarcates the boundaries of a different colour pattern and is coloured based on broad categorization of the colour patterns. The highly divergent races/incipient species are indicated with separate colours. Figure and legend taken from [Hines et al., 2011].

([Beltran et al., 2007]). The best studied mimicry ring is the one formed by species *H. melpomene* and *H. erato*. These two species show a remarkable parallel radiation, each having several sharply distinct geographic races that mimic one another perfectly in any particular place (Figure 10). Another species that has attracted great interest is *H. numata*. This species belongs to the so called “silvaniform” or *numata* clade. As we stated in the section above, *H. numata* is remarkable on account of an extraordinary local polymorphism that is absent or very limited in other *Heliconius* species ([Brown and Benson, 1974]). Most of the silvaniform species share mimicry rings with other groups of butterflies, mainly in the Danainae and Ithomiinae subfamilies, and such mimicry complexes are an important component of the butterfly biota in the Neo-tropics. A typical “tiger” colour pattern composed by a mosaic of black, orange and yellow/white elements is characteristic of most silvaniforms (*H. besckei* and *H. elevatus* are exceptions). Given the complexity of these patterns, the taxonomic classification of the silvaniforms based on colour patterns has been considered as one of the most difficult in Lepidoptera ([Brown, 1981]). This task has taken advantage from biosystematic work on the field and insectary, showing that very similar adults can be distinguished as eggs or larvae ([Brown, 1976] and references therein). Nowadays, ten species are described in the clade; the most recent phylogeny of *Heliconius* shows their relationship (Figures 4 and 1.4; see [Kozak et al., 2015]).

Here, we investigate two scarcely explored silvaniform species, *H. ismenius*, which is the sister species of *H. numata*, and *H. hecale*, a quite far related species of *H. ismenius* within the clade. Species *H. hecale* has a broad distribution across South and Central America, but *H. ismenius* is limited to the northwest of South America and Central America, where both species overlap (Figures 11 and 12). Species *H. hecale* displays particularly strong phenotypic variation across their range, having around thirty distinctly coloured geographic races (Figure 12), whereas *H. ismenius* has less than ten subspecies (Figure 11). In spite of this geographic diversity, both species show very limited local polymorphism, with at most two coexisting morphs in one single



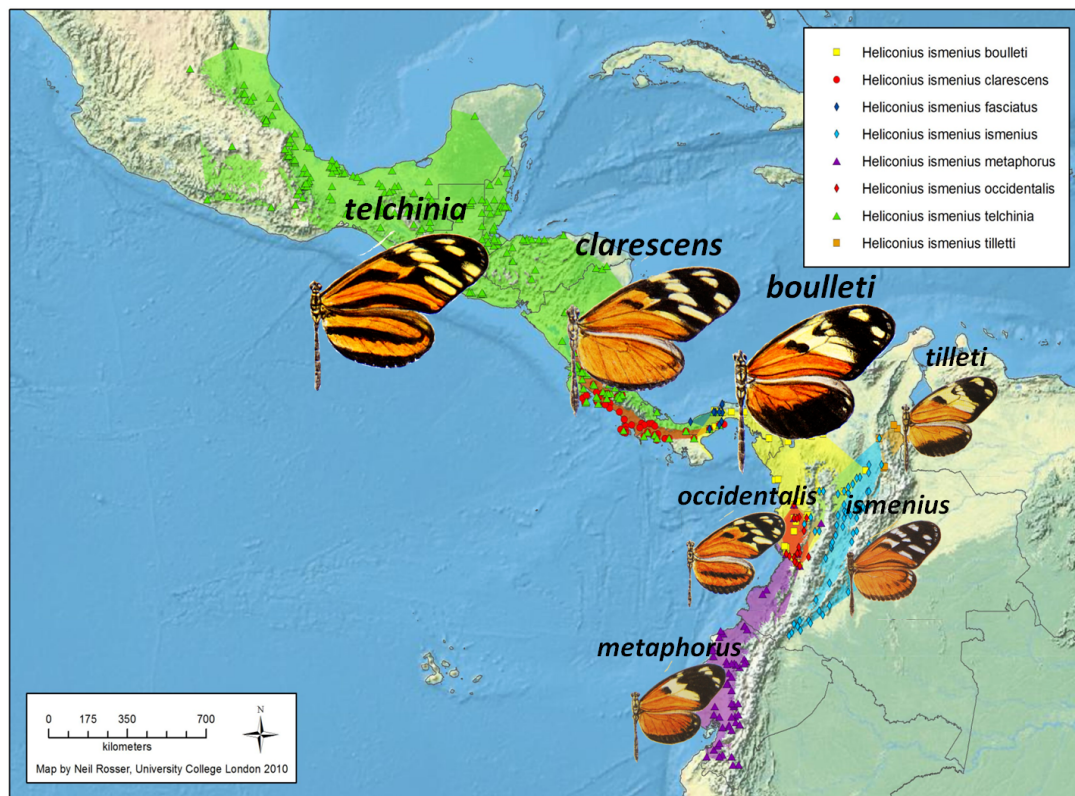


Figure 11 – **Distribution of *Heliconius ismenius* geographic races.** Coloured surfaces represent the range of each subspecies, named on top or bottom of each wing image. Races used in this study are shown in increased size in relation to the rest. Map taken from [Rosser et al., 2012] and wing images taken from [Holzinger and Holzinger, 1994].

locality (i.e. yellow and white morphs of *H. hecale zuleika* in North West Panama or of *H. ismenius telchinia* in Central Panama). Although some races of *H. hecale* and *H. ismenius* belong to the same mimicry ring in some areas, these two species can be easily distinguished on the basis of the resource use and morphology of immature stages, in concordance to what I said above. For instance, these species are found on distinct *Passiflora* host plants in Panama; *H. hecale* usually feeds on *P. vitifolia*, and *H. ismenius* on *P. quadrangularis* or *P. ambigua*. We observed in the insectaries that females of *H. hecale* usually lay separate eggs on the tips of the host plants, whereas *H. ismenius* females lay diffuse clutches of eggs on mature and immature leaves, and thus have semi-gregarious larvae (see Figure 13). The morphology of the caterpillars of both species differs in the number and size of longitudinal rows of black dots on a white body (see Figure 13).

Here, we exploit the increasingly robust synergy between genomic, phenotypic and ecological data to tackle open questions in evolutionary biology. By making the link between the gene and the community levels of biological organisation we investigate how does biological diversification take place in the face of gene flow, and how do reproductive barriers evolve and function. This thesis is composed of three chapters, which are summarised on the schema in Figure 15. In this study we investigate the factors and mechanisms that have favoured diversification of the genus *Heliconius*, by



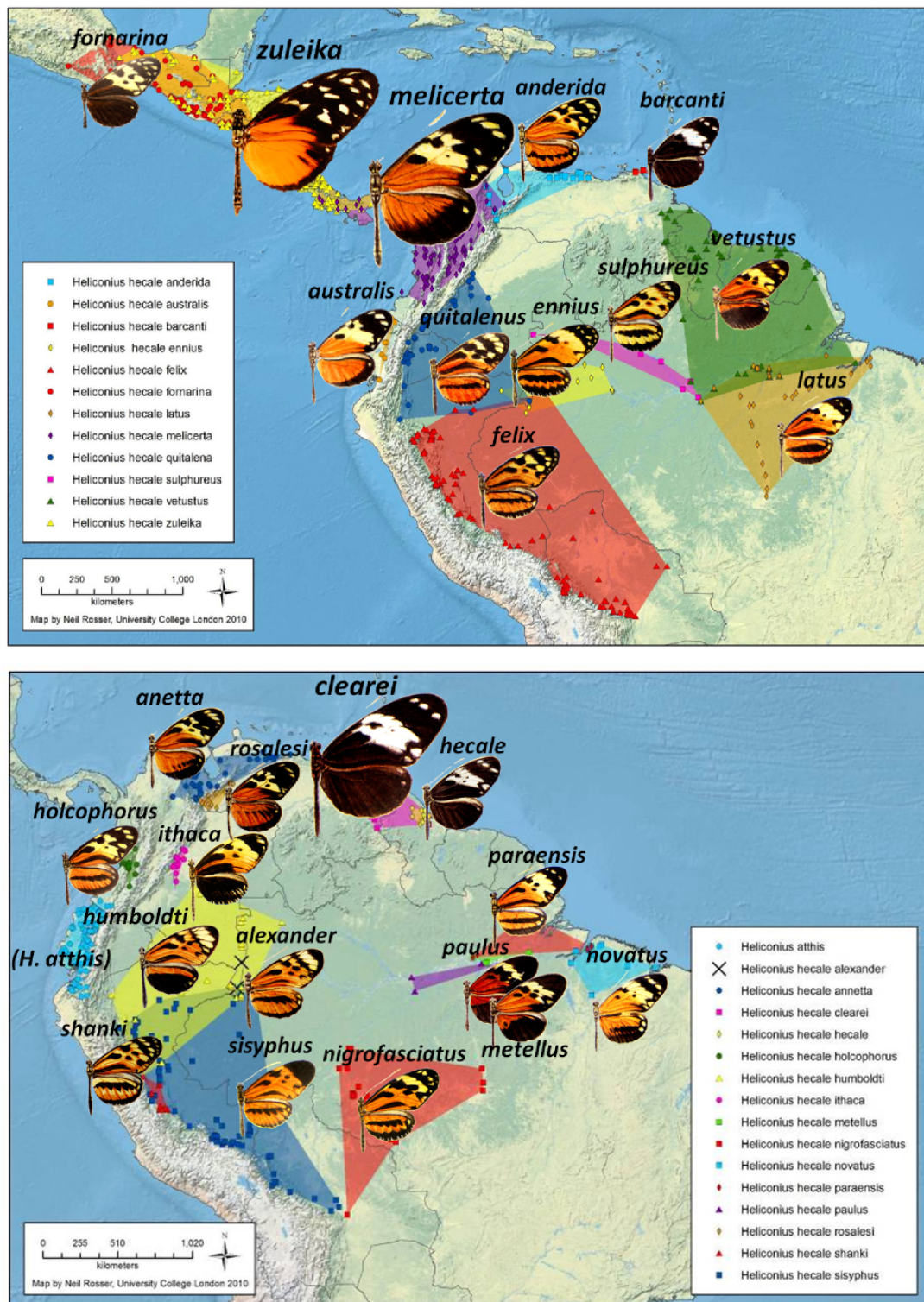


Figure 12 – Distribution of *Heliconius hecale* geographic races. Coloured surfaces represent the range of each subspecies, named on top or bottom of each wing image. Races used in this study are shown in increased size in relation to the rest. Maps taken from [Rosser et al., 2012] and wing images taken from [Holzinger and Holzinger, 1994].

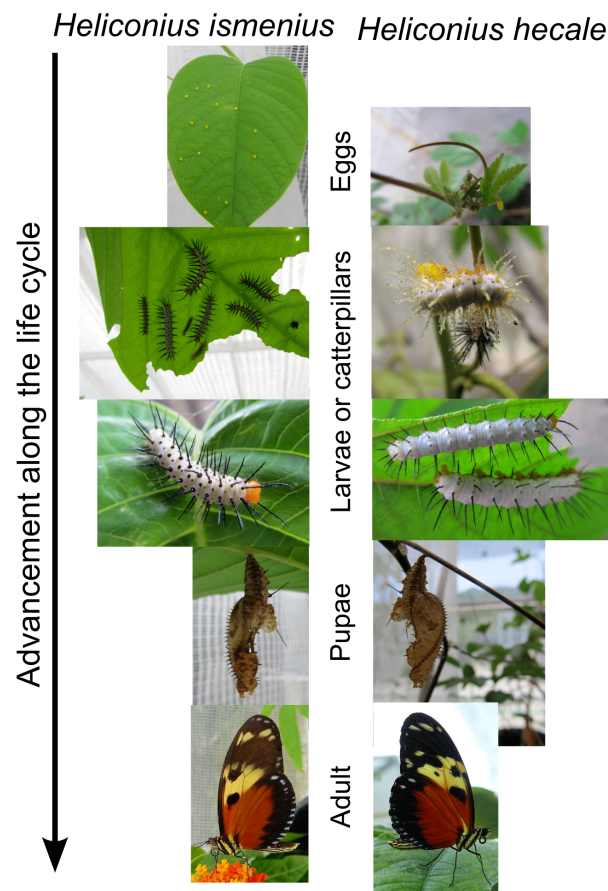


Figure 13 – Pictures showing (sometimes slight) differences at distinct stages in the life cycle between *Heliconius ismenius* and *H. hecale*.

focusing particularly on *H. hecale* and *H. ismenius*, two species having complex wing colour motifs which are very different from those mainly studied so far.

In the first part of this work (Chapter 1, green boxes in Figure 15) we ask what the genetic basis of biological diversification is. To what extent is adaptive phenotypic diversity determined by similar genetic architectures? Are convergent phenotypes necessarily defined by similar genetic architectures? What factors influence the evolution of the genetic architecture of adaptive traits? We will explore the extent of conservatism in the genetic architecture of wing colour control for mimicry within the silvaniform clade, to get a better picture of the evolution of distinct genetic architectures accounting both for convergence and diversification in the genus *Heliconius*, and to provide insights into the origin of the *supergene* architecture in *H. numata*.

This helps defining more precisely the question of the role of single major effect adaptive traits on divergence. Can such ecologically very important traits drive speciation alone to completion or is ecological speciation rather a multidimensional process relying on several factors? How can fully interbreeding lineages remain phenotypically distinct? What is the relative importance of different reproductive isolation barriers along the speciation continuum? How is genomic divergence built through time?

In chapter 2 (orange boxes in Figure 15), we study the hybrid zone between *H. hecale*

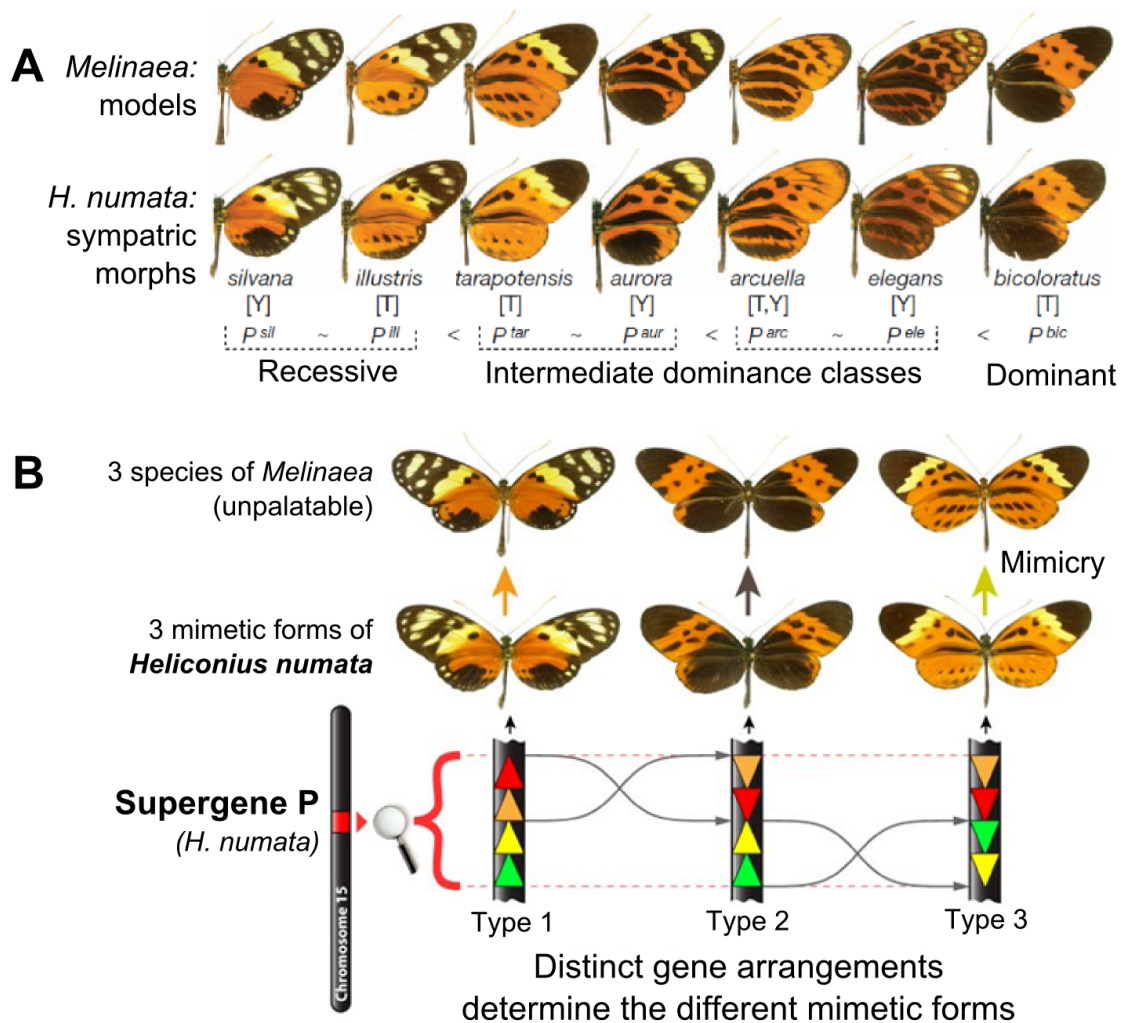


Figure 14 – **Local polymorphism in *H. numata* is controlled by a supergene.** (A) *H. numata* has multiple coexisting colour morphs at a single locality, and each of them mimics a different species of *Melinaea*, a genus in the Ithomiinae subfamily. Morphs of *H. numata* display a hierarchical dominance relationship, with dominance increasing from left to right. (B) Particular chromosomal rearrangements are associated with distinct coexisting morphs. Part A slightly modified from [Joron et al., 2011]; part B provided by Mathieu Joron.



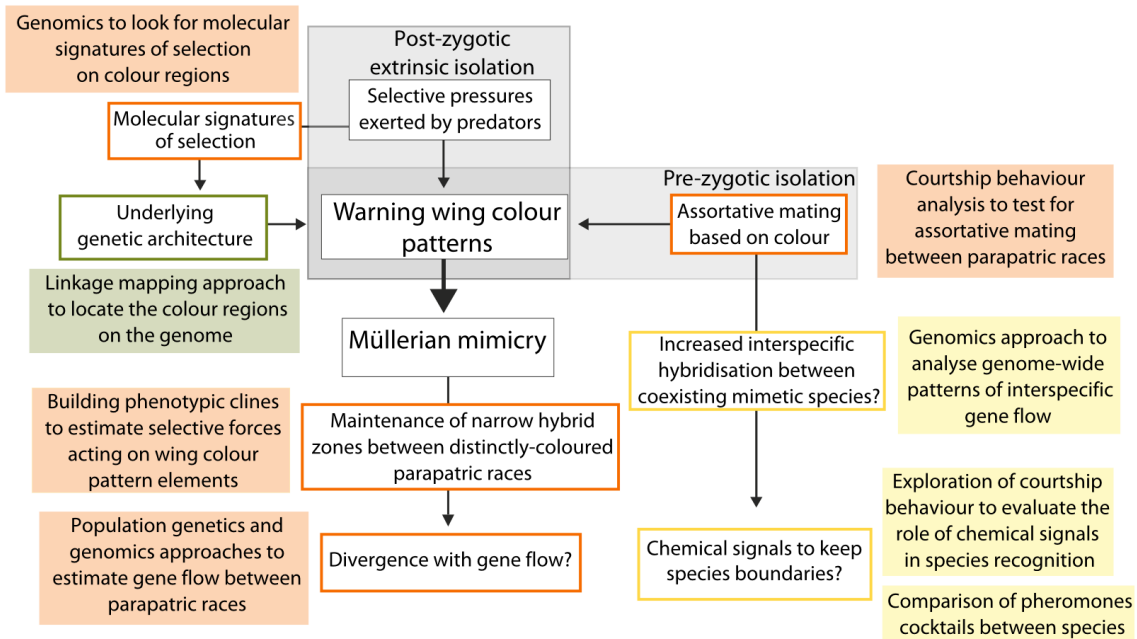


Figure 15 – **Schema summarising this study.** Some important features of *Heliconius* butterflies and unsolved questions are presented within the empty boxes. The approaches used to understand particular issues are mentioned within the fully coloured boxes. Green, orange and yellow boxes correspond to Chapters 1, 2 and 3 of this thesis, respectively.

*melicerta* and *H. hecale zuleika* in Panama and explore how selective pressures acting on wing coloration contribute to maintain phenotypic distinctiveness in presence of gene flow at early stages of divergence. By assembling population genetics and genomics, cline theory and male-based mate choice behavioural assays, we assess the extent of genome-wide gene flow between these races, we explore the strength of extrinsic post-zygotic isolation resulting from selective pressures imposed by predators on the mapped wing colour elements and we evaluate the strength of pre-zygotic isolation resulting from mate recognition based on wing patterns.

Finally, in Chapter 3 (yellow boxes in Figure 15), we investigate the relative role of colour and chemical communication as barriers to gene flow at an advanced stage of speciation. We use genomic data linked to chemical ecology approaches and behavioural data to perform genome-wide analysis of gene flow between natural populations of *H. hecale* and *H. ismenius*, to investigate whether pervasive interspecific hybridisation is universal to the whole genus *Heliconius*, to explore the genomic landscape of divergence between these “good” species and to investigate whether chemical cues may be involved in keeping the species boundaries between them. We take benefit from contrasting populations where both species share the same wing-patterns (Eastern Panama) to populations where the two species have distinct wing patterns (Western Panama).

## Chapter 1

# Conservatism and novelty in the genetic architecture of adaptation in *Heliconius* butterflies

Bárbara Huber (1,2,3), Annabel Whibley (1), Yann Le Poul (1), Nicolas Navarro (4,5), Arnaud Martin (6), Simon Baxter (7,8), Abhijeet Shah (1,9), Benoît Gilles (1,3), Thierry Wirth (1,2), W. Owen McMillan (3), Mathieu Joron (1,3)

1. Institut de Systématique, Evolution, et Biodiversité, UMR 7205 CNRS, Muséum National d'Histoire Naturelle, Paris, France
2. Laboratoire Biologie Intégrative des Populations, Ecole Pratique des Hautes Etudes (EPHE), Paris, France
3. The Smithsonian Tropical Research Institute, Balboa, República de Panamá
4. Laboratoire PALEVO, Ecole Pratique des Hautes Etudes, Dijon, France
5. UMR uB/CNRS 6282-Biogéosciences, Université de Bourgogne, Dijon, France
6. Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720-3200, USA
7. School of Molecular and Biomedical Science, The University of Adelaide, Australia
8. Department of Zoology, University of Cambridge, United Kingdom
9. Department of Animal Behaviour, Universität Bielefeld, Germany

## Abstract

Understanding the genetic architecture of adaptive traits has been at the centre of modern evolutionary biology since Fisher, but evaluating how the genetic architecture of ecologically important traits influences their diversification has been hampered by the scarcity of empirical data. Now, high-throughput genomics facilitates the detailed exploration of variation in the genome-to-phenotype map among closely-related taxa. Here, we investigate the evolution of wing pattern diversity in *Heliconius*, a clade of neo-tropical butterflies which have undergone an adaptive radiation for wing-pattern mimicry and are influenced by distinct selection regimes. Using crosses between natural wing-pattern variants, we used genome-wide RAD-genotyping, traditional linkage mapping and multivariate image analysis to study the evolution of the architecture of adaptive variation in two closely-related species: *Heliconius hecale* and *H. ismenius*. We implemented a new morphometric procedure for the analysis of whole-wing pattern variation, which allows visualising spatial heatmaps of genotype-to-phenotype association for each QTL separately. We used the *H. melpomene* reference genome to fine-map variation for each major wing-patterning region uncovered, evaluated the role of candidate genes and compared genetic architectures across the genus. Our results show that although the loci responding to mimicry selection are highly conserved between species, their effect size and phenotypic action vary throughout the clade. Multilocus architecture is ancestral and maintained across species under directional selection, whilst the single-locus (supergene) inheritance controlling polymorphism in *H. numata* appears to have evolved only once. Nevertheless, the conservatism in the wing-patterning toolkit found throughout the genus does not appear to constrain phenotypic evolution towards local adaptive optima.

**Keywords** *Heliconius*, mimicry, genetic architecture, adaptive traits, Next Generation Sequencing, morphometric analysis.

## 1.1 Introduction

Over the past decade, next generation sequencing technologies have provided increased power to identify the genomic targets of selection: the loci, genes and genetic variants that control adaptive phenotypes ([Stinchcombe and Hoekstra, 2007]). These tools expand the frontiers beyond classical model species and, in particular, have provided powerful insights into convergent evolution (c.f. [Arendt and Reznick, 2008]), whereby the same phenotype evolves in two or more lineages independently, typically in response to similar environmental challenges. Natural systems displaying phenotypic convergence provide a robust framework to investigate whether this convergence derives from the recruitment of the same or different genes and genetic mechanisms, thus allowing a better understanding of the molecular basis of adaptive evolution ([Stern, 2013]).

The evolution of an adaptive trait is influenced by its genetic architecture. This term encapsulates the often complex genotype-to-phenotype relationship and includes the number and nature of genetic elements (genes and alleles), their absolute and relative genomic locations, their effect sizes and their interactions. These interactions can occur with the environment (e.g. via epigenetic effects), between distinct genes (i.e. epistasis, additivity), between variants at the same locus (i.e. dominance) and in additional effects on other phenotypic traits (i.e. pleiotropy). The genetic architecture of phenotypic variation can influence both convergence and diversification processes, and selective pressures may operate on any of its components, either singly or in combination ([Hansen, 2006]). A large number of theoretical studies

described the evolution of these different features ([Lande, 1980]; [Barton, 1995]; [Orr, 1998]; [Carter et al., 2005]). Nevertheless, scant empirical data exists on the factors associated with the evolution of genetic architectures and on how heterogeneity in the genetic architecture of complex traits can influence their diversification ([de Visser et al., 1997]; [Lair et al., 1997]).

Butterflies in the genus *Heliconius* represent an excellent system to investigate the evolution of the genetic architecture of complex adaptive traits. This clade contains distinct lineages which display different wing colour patterns, show heterogeneity in the genetic architecture of these traits, and permit a comparative approach across lineages thanks to the occurrence of both convergent and divergent evolution. Additionally, the ecological roles of wing patterns and the selection regimes shaping their variation have been relatively well studied in this genus ([Brown, 1981]; [Kapan, 2001]; [Jiggins et al., 2001]). *Heliconius* butterflies are unpalatable to predators and the spectacular wing colour patterns advertise their toxicity. Several species within and outside this genus converge in wing patterns, enjoying survival benefits in the face of predation by using similar signals of toxicity. This convergence is known as Müllerian mimicry. This adaptation to the local prey environment recognised by educated predators suggests that the genes controlling wing colour are subject to strong selective pressures.

Previous studies have defined a palette of genomic regions of large phenotypic effect shared by distinct *Heliconius* species and underlying the diversification of colour patterns ([Joron et al., 2006]; [Kronforst et al., 2006a]; [Papa et al., 2008]; [Kronforst et al., 2006a]). This conserved “toolkit” of genes is mainly distributed across four of the 21 *Heliconius* chromosomes, but several minor effect loci have also been detected (see summary Table 1.1). Two of the causal genes that drive adaptive pattern variation have been identified. One is the *WntA* signalling ligand, a putative morphogen that determines the size and position of melanic patterns in the forewing median region (corresponding to the effects of loci *Ac/Ac/Sd* in *Heliconius melpomene*, *H. cydno* and *H. erato*, respectively; [Martin et al., 2012]). Another gene is a transcription factor, homologous to the *Drosophila* gene *optix*, which prefigures the variety of red wing elements controlled by the cluster of loci *B-D/D/G-Br* in *Heliconius melpomene*, *H. erato* and *H. cydno* ([Reed et al., 2011]; [Martin et al., 2014]). Causal genes at two other major loci have yet to be formally characterised at the gene level: *K* which controls the white/yellow switch in *H. melpomene* and *H. cydno* ([Kronforst et al., 2006c]), and a tight cluster of loci that controls most of the variation in yellow and white pattern elements. The latter is a complex of at least three linked loci (*Yb*, *Sb* and *N*) in *H. melpomene*, two of which have also been described in its sister species *H. cydno* (*Yb* and *Sb*). In the more distantly related *H. erato*, this region harbours the *Cr* locus which controls similar pattern variation ([Jiggins and McMillan, 1997]). Recombination occurs between loci *Yb*, *Sb* and *N* in *H. melpomene*, but *Cr* in *H. erato* segregates as a single genetic unit ([Sheppard et al., 1985]; [Mallet, 1989]; [Ferguson et al., 2010]).

This variation in the level of linkage reveals slight modifications in the genetic architecture, nested within an otherwise highly conserved multilocus architecture throughout the *Heliconius* genus for the control of pattern variation ([Kronforst et al., 2006a]; [Papa et al., 2008]). There are other subtle architectural differences. For instance, the red/yellow forewing band switch is caused by variation in a single locus, *D*, in *H. erato*, but by the interaction of two unlinked loci (*B* and *N*) in *H. melpomene* ([Sheppard et al., 1985]).

To date, almost all our knowledge about the architecture of colour pattern variation in *Heliconius* comes from studies of species displaying variable shapes of red, white and yellow elements within a mainly black wing ([Jiggins and McMillan, 1997]; [Jiggins et al., 2005b]; [Kronforst et al., 2006a]; [Reed et al., 2011]; [Nadeau et al., 2014]). In contrast, the genetic basis of variation for “tiger” patterns, which are composed of a mosaic of black, orange and yellow/white elements, is less known. These patterns are widely used by species of the so-called

“silvaniform” sub-clade of *Heliconius*, which contains ten described species sharing mimicry relationships with other groups of butterflies, mainly in the Danainae subfamily. Within this clade, the genetic basis of colour pattern has only been characterized in species *H. numata* ([Joron et al., 2006]). Compared to what is known in other *Heliconius* species, *H. numata* shows a strikingly different genetic architecture of wing pattern variation. Indeed, a single-locus (supergene *P*) virtually monopolizes the control of wing pattern variation in this species. According to [Thompson and Jiggins, 2014], a supergene is “a genetic architecture involving multiple linked functional genetic elements that allows switching between discrete, complex phenotypes maintained in a stable local polymorphism”. The supergene *P* is positionally homologous to the *Yb-Sb-N* cluster of *H. melpomene* ([Joron et al., 2006]).

Mimetic selection regimes are largely determined by the distribution and abundance of distinct signals used by local prey communities. Most *Heliconius* species, including tiger-patterned species, display geographic races differentiated in wing patterning in response to directional selection imposed by positive frequency dependence favouring one single well-defended pattern in each locality ([Brown, 1981]). By contrast, *H. numata* displays a rich local polymorphism, and all populations harbour distinct forms mimicking multiple distinct tiger-patterned species ([Brown and Benson, 1974]). This polymorphism is believed to be driven by fine-scale variations in the abundance of alternative tiger-patterned mimicry rings, causing balancing selection at the regional level ([Joron et al., 1999]). The heterogeneity in selection regimes shaping *Heliconius* wing patterns, i.e. local monomorphism under directional selection vs. polymorphism under balancing selection, allows investigating the relationship between selection regimes and the evolution of distinct genetic architectures underlying complex adaptive traits.

Here, we focus on the silvaniform clade within *Heliconius* and ask whether the genetic architecture of colour pattern variation is associated with the phenotypic variation itself or with the selection regime shaping it. To this end we carefully analyse wing pattern inheritance in two unexplored tiger-patterned species in this sub-clade, *H. hecale* and *H. ismenius*, which show geographic variation under local directional selection for mimicry. We combine traditional linkage mapping powered by next generation sequencing, multivariate quantitative genetics and fine-mapping of candidate genes, to identify the genomic regions controlling wing pattern variation in these two species and to explore the evolution of genetic architectures in a broader comparative framework.

## 1.2 Materials and Methods

### 1.2.1 Crossing experiments

Intraspecific crosses were performed between geographic races of *H. hecale* and *H. ismenius* (Figure 1.1 and Table 1.2). For *H. hecale*, we crossed subspecies *melicerta* (eastern Panama) with *zuleika* (western Panama), and *melicerta* with *clearei* (Venezuela) to obtain F1 males which were backcrossed to *melicerta* females (Figures 1.1, 1.3AI and 1.3BI). For *H. ismenius*, we crossed *boulleti* (eastern Panama) with *telchinia* (western Panama), and then backcrossed F1 males to *boulleti* females (Figures 1.1 and 1.3CI). Breeding was performed at the Smithsonian Tropical Research Institute (STRI) in Gamboa, Panama. Butterflies were kept in  $\sim 2 \times 2 \times 2m$  cages and provided with ample sugared water and pollen. *Passiflora vitifolia* and *P. edulis* were used for *H. hecale* oviposition and as larval food-plants, while *P. quadrangularis* was used for *H. ismenius*. The bodies of parents and progeny were preserved in NaCl-saturated DMSO solution at  $-20^\circ\text{C}$  and wings were stored separately in glassine envelopes.



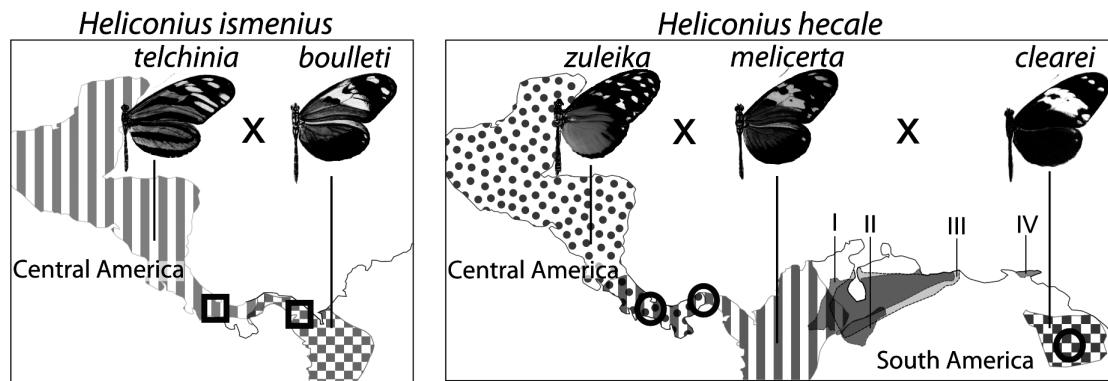


Figure 1.1 – Summary of crosses performed in *Heliconius hecale* and *H. ismenius*. Geographic distribution of the subspecies used for the crosses are indicated by filling patterns, and sampling localities by circles and squares. The distribution of other *H. hecale* races found in Northern South America is also shown: *H. h. annetta* (I), *H. h. rosalesi* (II), *H. h. anderida* (III) and *H. h. barcanti* (IV).

### 1.2.2 Phenotypic analysis of the broods

Wing pattern variation was quantified in three distinct ways. First, variation segregating with largely discrete alternative phenotypic states (e.g. presence/absence) was scored in all progeny. This included the number of marginal yellow spots in the dorsal and ventral views of the hindwing. However, much continuous variation was observed and hard to score by eye. Therefore, in order to get a comprehensive measure of colour pattern variation in the mapping families, we used morphometric quantification of pattern with the Colour Pattern Modelling tool (CPM; [Le Poul et al., 2014]). This method uses recursive alignment of wing outlines and image segmentation to identify conserved and homologous pattern elements. Briefly, CPM consists of a first colour clustering step, where colours are treated as classes of pigments. Second, wing images are aligned based on pattern and outline, onto a modal wing “model” built recursively from the image stack. Finally, a Principal Component Analysis (PCA) of colour variation of homologous pixels across wings is performed to reduce the dimensionality of pattern variation. Hindwings and forewings were treated separately for the first two steps, but combined for the PCA. For each cross, all offspring of the largest broods with intact wings were included.

### 1.2.3 Restriction Site-Associated DNA (RAD) library construction and sequencing

RNA-free genomic DNA was extracted from thoracic tissue using the Qiagen DNeasy Blood and Tissue kit following manufacturer’s instructions. Three RAD libraries were prepared from the backcross parents and 62 offspring of the largest broods of each type of cross, using the protocol adapted by [Heliconius Genome Consortium, 2012]. Briefly, 300-350ng of genomic DNA were digested with the 8bp-cutter restriction enzyme, *SbfI*. We expected 1,053 cutting sites based on CCTGCAGG occurrences in the reference *H. melpomene* genome. For brood parents, reactions were scaled up to 1,000-1,400ng inputs to increase representation. One of 64 Illumina P1 adapters, each with a unique 5-base barcode, was used to tag each specimen within a library. During the final PCR amplification step, 18 cycles of PCR were used, with eight independent amplifications pooled to minimize the contribution of PCR errors. Each library was paired-end sequenced in one lane of an Illumina HiSeq2000 with 100-base read length.

### 1.2.4 Bioinformatics analysis

The function *process\_radtags* implemented in *Stacks v0.9991* ([Catchen et al., 2013]) was used to demultiplex the separate libraries and apply basic quality filters. The processed reads of each individual were mapped to the reference genome of *H. melpomene* version 1.1 ([Heliconius Genome Consortium, 2012]) using *Stampy v1.0.17* ([Lunter and Goodson, 2011]) with default parameters except for setting the substitution rate to 0.01. SAM/BAM file conversion, analysis, and filtering were performed using *SAMtools v0.1.18* ([Li et al., 2009]) and *Picard Tools v1.67* (<http://picard.sourceforge.net>). To limit genotype miscalling due to PCR bias, PCR duplicates were removed using *Picard Tools v1.67*. At this stage, samples for each type of cross were combined in the same alignment file and processed together. Local realignment around indels was performed using the *Genome Analysis Tool Kit (GATK) v1.6-2* ([DePristo et al., 2011]).

SNP genotypes were called using the *GATK v1.6-2 UnifiedGenotyper* with default parameters with the exception of setting expected heterozygosity to 0.015. We applied positional filters to exclude repetitive regions of the genome ([Heliconius Genome Consortium, 2012]). Filters for coverage ( $> 10\times$  and  $< 200\times/249\times$  for offspring/parents, respectively), genotype ( $GQ \geq 30$ ) and mapping quality ( $MQ \geq 40$ ) were applied using a custom *Perl* script ([Kanchon Dasmahapatra, pers. comm.]). After filtering, markers with genotype calls at fewer than 80% of individuals were excluded. Sites showing Mendelian inconsistencies were removed and missing genotype calls were imputed, scaffold-by-scaffold, using *Beagle v3.3.2* ([Browning and Browning, 2009]). Different subsets of individuals per brood were tested to define the dataset that generated the best quality downstream linkage maps. Final linkage maps were constructed from populations of 41, 42 and 29 offspring for the larger *melicerta* $\times$ *zuleika*, *melicerta* $\times$ *clearei* and *boulleti* $\times$ *telchinia* broods, respectively.

### 1.2.5 Linkage map construction

Crossing-over does not occur during oogenesis in Lepidoptera ([Turner and Sheppard, 1975]), so an intact haplotype of each chromosome is passed from mother to offspring. Consequently, any female informative marker on a given autosome can inform on the segregation of linked maternal variation (“chromosome print”, c.f. [Jiggins et al., 2005b]). In contrast, male-informative SNPs (heterozygous in father but homozygous in mother) and intercross sites (heterozygous in both parents) do recombine and inform on genetic distances within chromosomes. Genetic maps were computed independently for each cross using *Joinmap v3.0* ([Van Oijen and Voorrips, 2001]). We filtered out SNPs that deviated from the expected 1:1 segregation ratio (female and male-informative backcross markers) and 1:2:1 ratio (intercross markers) to generate a genotype matrix. Linkage groups corresponding to the 20 *Heliconius* autosomes were reconstructed using female informative markers and a LOD threshold  $\geq 6$  for all three datasets. The sex chromosome was not reconstructed. Male informative and intercross markers were collapsed to unique segregation patterns using a custom *Perl* script ([John Davey, pers. comm.]). Collapsed markers were combined with female-informative chromosome prints and clustered by linkage group ( $LOD \geq 5$ ). Individual linkage maps were built using the Kosambi mapping function and a  $LOD \geq 1$ .

### 1.2.6 Mapping wing pattern loci

Phenotypes segregating with discrete alternative states were incorporated directly into map construction, alongside the collapsed marker sets, and were thus co-localised with the markers with which the phenotype was most strongly associated. A generalized linear model (GLM) was

used to test for an association between the number of spots on the margin of the hindwing and each of the 20 “chromosome prints”, using *R*. This analysis was possible since the mother of the brood (and not the father) was heterozygous for this trait. The overall colour pattern variation quantified with CPM was mapped as QTLs by using a genome-wide Haley-Knott regression implemented in the *R/qtl* package ([Broman et al., 2003]). This analysis was extended by performing multivariate analysis on all principal components with an eigenvalue  $\geq 2\%$  using the *R/shapeQTL* package (available on request; see 1.5.1 for more details). Statistical thresholds for significant linkage were based on 1000 permutations. To further evaluate the identified QTLs, we ran a stepwise multiple QTL model search using the algorithm developed in ([Broman and Sen, 2009]) and implemented in the *R/shapeQTL* package for multivariate traits. The search was restricted to additive QTLs but epistatic interactions between discovered QTLs were evaluated in the final model. Since colour pattern analysis is affected by sex ([Jones et al., 2011]), we used gender as an additive covariate. We employed only male informative markers on the 20 dense autosomal linkage maps for the different broods. The subsets of offspring in the final linkage maps having intact wings were used for this analysis. A conventional threshold of  $\text{LOD} \geq 3$  and other relaxed requirements were used to detect suggestive QTLs (SI Methods).

### 1.2.7 Refining candidate intervals

To fine-map candidate intervals associated with discrete phenotypic variation, we genotyped additional markers within each region of interest in an extended panel of progeny. A combination of newly designed and previously published markers was used, generally targeting single-copy nuclear loci (SCNL), but on occasion anchored in non-coding regions (Table 1.3). Markers were first amplified in brood parents, then in the progeny when allelic variation was found (see 1.5.1 for more details about genotyping methods).

### 1.2.8 In situ hybridisation

Larval wing disc *in situ* hybridisations were performed following a previously described procedure ([Martin et al., 2012]). Wing imaginal discs of three *H. hecale melicerta* individuals and two *H. hecale zuleika* individuals originated from phenotypically pure stocks maintained in insectaries at STRI, in Gamboa, Panama. The *WntA* riboprobe was synthesized from a 885bp cDNA amplicon previously cloned from the closely related species *Heliconius cydno* ([Martin et al., 2012]).

## 1.3 Results

Mapping families show variable progeny. In each cross type, we obtained two families sired by the same F1 male crossed to unrelated mothers. In *H. hecale*, we reared a total of 120 (98 + 22) and 290 (183 + 107) butterflies for the *melicerta* × *zuleika* and *melicerta* × *clearei* crosses, respectively (Figure 1.3). *H. ismenius* was more difficult to rear, and we obtained 54 (36 + 18) offspring for the *boulleti* × *telchinia* crosses. The offspring of the broods showed segregation of discrete colour pattern characters affecting large portions of the wings (Figure 1.3), as well as some minor quantitative variation.

Four major loci segregated independently in Mendelian ratios (Table 1.5) and were named according to the inferred homology with mapped loci of similar phenotypic effect in other *Heliconius* species. First, **HhK** governed the white (*HhK<sub>c</sub>*) / yellow (*HhK<sub>m</sub>*) switch for the forewing band in the *H. h. melicerta* × *clearei* families (Figure 1.3BI), white being dominant to

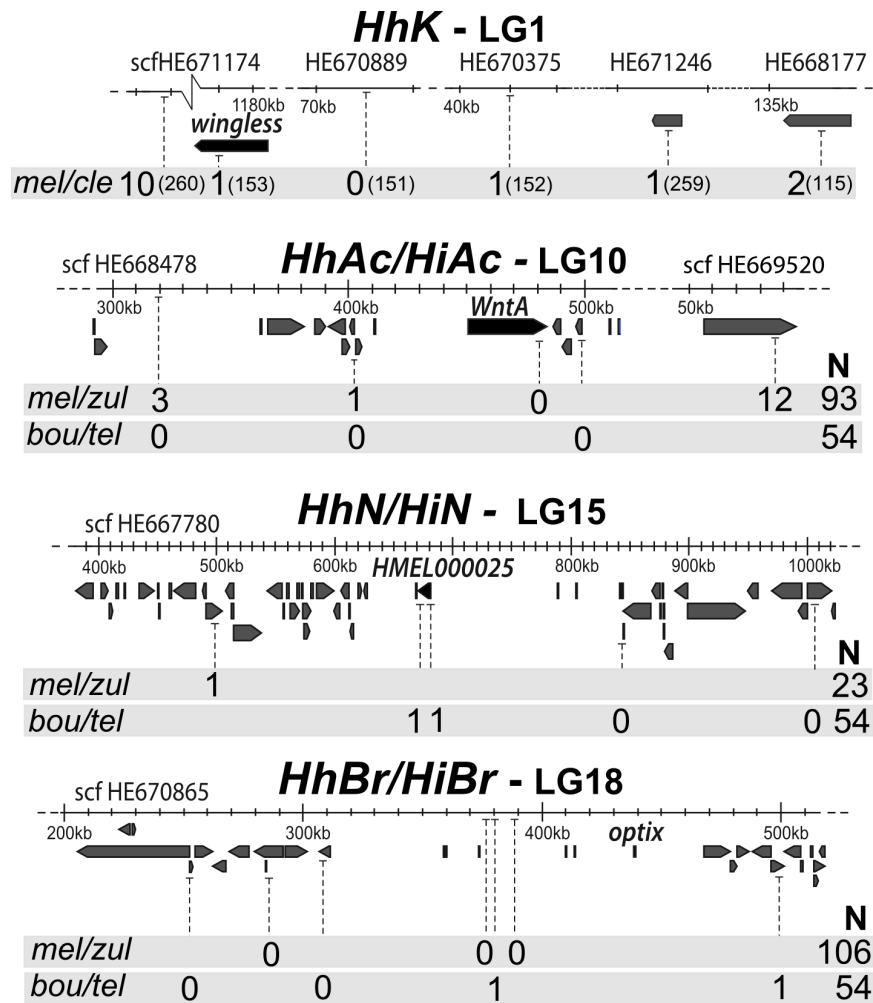


Figure 1.2 – Fine mapping of wing patterning loci in *H. hecale* and *H. ismenius*. Gray shaded boxes show recombinant individuals found in a total of *N* offspring in *H. hecale melicerta* × *H. h. clearei* (*mel/cle*), *H. hecale melicerta* × *H. h. zuleika* (*mel/zul*), and *H. ismenius bouletti* × *H. i. telchinia* (*bou/tel*) crosses. Annotated genes on each scaffold and candidate colour genes are represented by gray and black block arrows, respectively. Scaffolds on LG1 (top panel) are ordered according to the *H. melpomene* reference genome, but the order is unknown for the three scaffolds indicated on the right (HE670375, HE671246 and HE668177).

yellow. Second, **HiAc** and **HhAc** shaped the size and position of black patterns close to the forewing discal cell in *H. ismenius* (Figure 1.3CI) and in the *H. h. melicerta* × *zuleika* families (Figure 1.3AI), respectively. The *H. i. telchinia* (*HiAc<sub>t</sub>*) and *H. h. zuleika* (*HhAc<sub>z</sub>*) alleles were fully dominant over the *H. i. bouletti* (*HiAc<sub>b</sub>*) and *H. h. melicerta* (*HhAc<sub>m</sub>*) alleles, respectively. Dominant alleles break the continuity of the forewing yellow band by increasing the size of a black spot next to the discal cell. Third, **HhN** and **HiN** controlled presence vs. absence of small yellow dots in the largely black forewing submarginal area in *H. hecale* and *H. ismenius*, respectively. Although those loci are similar in the wing position and type of variation controlled, the two species showed differences in phenotypic action and dominance, which is detailed in Figures 1.3AI and 1.3CI. Because of its restricted phenotypic effect on the wing, *HiN* was considered a locus of minor effect. Fourth, in *H. hecale* and *H. ismenius*, respectively, **HhBr** and **HiBr** controlled the shape of the black marginal band of the hindwing, defined by the orange elements around this band. In our crosses, *H. h. zuleika* (*HhBr<sub>z</sub>*) and *H. i. telchinia* (*HiBr<sub>t</sub>*) alleles were strongly dominant over *H. h. melicerta* (*HhBr<sub>m</sub>*) and *H. i. bouletti* (*HiBr<sub>b</sub>*) alleles, respectively (Figures 1.3AI and 1.3CI). Dominant alleles produce a broken boundary of the black marginal band, whereas recessive homozygotes show a smooth, wide black band.

Additionally, we recognised two presumably polygenic traits with continuous variation in our *H. hecale* families. Throughout this paper, trait names will be written in non-italics (in contrast to Mendelian loci) and will refer simultaneously to the quantitative trait and the QTL of major effect associated with it. First, *melicerta* × *zuleika* families showed segregation for the number of yellow spots (2 to 7) along the distal hindwing margin, a trait we called **Hspot** (Figure 1.3AI). The alleles were not fixed in specific parental races: the F1 father seemed to be homozygous and the mother of the backcross brood heterozygous for this trait. Second, *melicerta* × *clearei* crosses showed continuous variation in wing melanisation (**Cm**; Figure 1.3BI). Loci controlling *Cm* variation essentially determined the position of the boundary between black and orange areas on hindwings and in the proximal region of forewings.

### 1.3.1 Construction of RAD-sequence linkage maps

We obtained around 162, 247 and 343 million reads for each of the three libraries (Table 1.6). The number of reads per individual ranged from 132, 182 to 38, 739, 993, excluding 4 individuals in each library which were virtually absent due to presumed barcode failure. The intended over-representation of parental samples was observed. Tables 1.6-1.7 provide a detailed breakdown of RADseq library statistics. After applying basic quality filters, on average 82% of the raw read dataset was retained (Table 1.6), of which approximately 94% was mapped to the *H. melpomene* reference genome. PCR duplicates can potentially lead to biases towards a single allele and thus introduce genotyping errors, and so were excluded from our dataset. Despite efforts to reduce library clonality during the preparation stages, we observed a drastic decrease in data quantity when excluding PCR duplicates from the mapped reads: only around 9% of the mapped reads were retained. To maximise genotype accuracy, the duplicate-removed dataset was used for the map construction, with the corollary that some individuals were excluded from the analysis due to a high proportion of missing data. Subsets of individuals with the highest total number of high quality calls were retained in the analysis: a subset of 41, 42 and 29 offspring plus the 2 backcross parents for the bigger *melicerta* × *zuleika*, *melicerta* × *clearei* and *bouletti* × *telchinia* broods (Figure 1.5). At this stage, any remaining missing genotypes were imputed and markers with improbable segregation patterns, such as Mendelian inconsistencies, were excluded. On average, 2,187 total polymorphic sites were informative for map construction: 963 female informative, 857 male informative and 367 intercross markers (Table 1.10). These SNPs were

**Brood 96 (36) - Brood 101 (18)**

recovered from 460, 510 and 566 RAD-tags among the 1,053 expected cutting sites by *SbfI* enzyme, in the three broods respectively. Thus these sites are well distributed across the genome. The depth at each of these variable sites averaged  $114\times$  for the parents and  $50\times$  for the offspring (Figure 1.6). The male informative and intercross markers together were collapsed to  $\sim 660$  unique segregation patterns on average, 39.2% of which were supported by more than two markers (Table 1.10). Around 450 markers on average were mapped to the 20 autosomes for each type of cross, but they were heterogeneously distributed across the linkage groups (Figures 1.7-1.9).

### 1.3.2 Colour loci in *H. hecale* and *H. ismenius* map to previously identified regions

The loci segregating in our broods map to the same genomic regions where colour genes have previously been localised in other *Heliconius* species.

*HhK* maps to chromosome 1, and clusters with markers on the genomic scaffold that contains the developmental gene *wingless* (HE671174; Figure 1.10). We found 1 recombinant in 153 genotyped offspring between *HhK* and *wingless* (Figure 1.2). The mapping interval is described on the basis of the order of the scaffolds defined for *H. melpomene* and may indicate the position of *HhK* between scaffolds HE670375 or HE671246 (relative position unresolved) and the scaffold containing *wingless* (Figure 1.2).

Both *HhAc* and *HiAc* map to chromosome 10. In both *H. h. melicerta* $\times$ *zuleika* and *H. i. bouletti* $\times$ *telchinia* crosses, these loci co-segregate with markers on scaffold HE668478 (Figure 1.10), which contains the gene *WntA*. By genotyping additional markers around this locus for a larger number of offspring, we revealed a perfect association of *HhAc/HiAc* with *WntA* in our crosses (no recombinants in the vicinity of *WntA*; Figure 1.2).

---

Figure 1.3 (*preceding page*) – **Phenotypic effect of Mendelian wing-patterning loci and major QTLs identified in *H. hecale* and *H. ismenius* crosses.** For each type of cross (A-C), panel I (left) shows the crosses performed, the phenotypes associated with inferred genotypes at the major Mendelian loci (colour *HhK*; forewing melanisation *HhAc/HiAc*; forewing distal band layer/spot *HhN/HiN*; hindwing band *HhBr/HiBr*) and variation of the quantitative traits (dashed boxes: Hindwing spots Hspot, continuous melanization Cm). Parental races (top left) are represented by their dorsal views, the F1 male siring the mapping families (top right) by its dorsal and ventral views, and typical backcross specimens (bottom) have arrows pointing to the variable character. The name of the mapping families is written on the bottom of the panels of each cross type, with total number of offspring shown in brackets. Families labelled in bold were used to build the RAD libraries. Panel II (right) shows the genomic position and phenotypic effect of major QTLs. Coloured wing diagrams show the spatial distribution of individual QTL effects on pattern variation extracted from multi-variate wing pattern analysis. Phenotypic variation is broken down into heatmaps for each of the three main colours (black, orange and yellow), representing, for every wing position, the strength of association between colour presence and allelic transition at the QTL (from blue to red). For analytical simplicity, both white and yellow elements in the *H. hecale melicerta* $\times$ *H. h. clearei* cross were considered as yellow elements. Genomic plots show genome-wide association (LOD) between wing pattern variation and markers along the 20 autosomes, with 5% (solid line) and 10% (dashed line) association thresholds. Panel AIII shows the detection of *WntA* transcripts by *in situ* hybridisation on wing imaginal discs of the last larval instar of *H. h. melicerta* and *H. h. zuleika*. *WntA* expression shows marked differences along the discal crossvein (arrows), in the M3-Cu2 inter-vein region (brackets), and in the Cu2-Cu1 inter-vein region (arrowheads). Colour dots indicate vein intersection landmarks. Phenotypic variation controlled by the *HhAc* locus is represented on the right.

The *HhN* and *HiN* loci map to chromosome 15 (Figure 1.10). More specifically, they cluster with RAD markers placed close to the superscaffold containing *Yb-Sb-N/Cr/P* in other species (HE667780; Figure 1.10). In *H. ismenius* we found one recombinant between *HiN* and gene *HMELO00025* (Figure 1.2), one of the candidate genes for *Yb* in *H. melpomene* and putatively part of the *P* supergene in *H. numata* ([Wu et al., 2010]; [Nadeau et al., 2012]). In *H. hecale*, we found 1 recombinant between *HhN* and the only informative marker available for genotyping in 23 offspring (Figure 1.2). For this particular analysis, we only used individuals of genotype *HhN<sub>z</sub>HhN<sub>z</sub>*, due to the difficulty of distinguishing *HhN<sub>z</sub>HhN<sub>m</sub>* and *HhN<sub>m</sub>HhN<sub>m</sub>* genotypes with certainty.

Finally, both *HhBr* and *HiBr* map to chromosome 18 (Figure 1.10). *HhBr/HiBr* co-segregate with RAD markers on the scaffold containing *optix* (HE670865; Figure 1.10). In the *H. hecale* families, we found no recombinants between *HhBr* and markers near *optix* in 106 offspring genotyped (Figure 1.2). In *H. ismenius*, a region that excludes the coding region of *optix* was delimited (Figure 1.2).

We found a significant association between the number of yellow spots on the margin of the hindwing (Hspot) and the maternal variation (chromosome prints) of linkage groups 15 ( $P = 1.37 \times 10^{-4}$ ), 6 ( $P = 8.69 \times 10^{-3}$ ) and 12 ( $P = 1.21 \times 10^{-2}$ ). The variation associated with this QTL is highly correlated with the discrete effect of *HhN* ( $\tau = 0, 55$ ;  $p = 1, 09 \times 10^{-9}$ ), which suggests that the number of yellow spots is largely controlled by a locus located close to the *Yb-Sb-N/Cr/P* region on chromosome 15.

The genomic position and phenotypic effect of three of the mapped major Mendelian loci was confirmed through quantitative, multivariate analysis of whole wing variation (Figures 1.3AII, 1.3BII, 1.3CII and 1.11). We found peaks of significant association between the variation of specific wing areas and the genomic regions described above. Our morphometric analysis allowed visualising in the form of heatmaps the genotype-to-phenotype association, which enabled a fine description of the effects associated with each identified QTL (Figures 1.3AII, 1.3BII and 1.3CII). The effect of these QTLs corresponded to that controlled by *HhAc/HiAc* (on LG10; Figure 1.3AII and 1.3CII), *HhN* (on LG15; Figure 1.3AII), and *HhBr/HiBr* (on LG18; Figure 1.3AII and 1.3CII). For all but one of those QTLs, confidence intervals included markers placed on the scaffolds containing known candidate colour genes (Figure 1.11). These intervals are relatively precise, extending over  $8.43 \pm 7.39$  cM on each chromosome.

In both species, loci *HhAc* and *HiAc* essentially control variation in black patterns situated around the forewing discal crossvein and extending into the M3-Cu1 region (Figures 1.3AII and 1.3CII). In situ hybridisation assays showed that *WntA* mRNA is expressed in the median forewing region in *H. hecale* larval wing disks, overlapping with the presumptive position of the *HhAc/HiAc*-dependent pattern variations (Figure 1.3AIII).

Additionally, our quantitative approach highlighted that the Cm trait is mainly explained by a major QTL mapped to a narrow region around *optix* on LG18 (Figures 1.3BII and 1.11B). This variation is also affected, albeit to a lower extent, by a second QTL mapping near *HMELO00025* on LG15. The effect of this second QTL is mainly restricted to the medial region of the hindwing, similar to the region affected by *Yb* in *H. melpomene* and close allies. We did not detect significant epistasis between these two QTLs ( $F(13, 25) = 1.93$ ,  $P = 0.08$ ). Furthermore, the QTL on LG15 in the *melicerta* × *clearei* family explains minor yellow/black variations in the distal area of the forewing band (*HhN* in Figure 1.3BII) in the same position as locus *HhN*. Also, this QTL is associated with variation in yellow apical forewing spots (Fspot) in the same *melicerta* × *clearei* family (Figure 1.3BII). Finally, the Hspot trait was also highlighted in the *melicerta* × *zuleika* family in association with the QTL on LG15. We found a significant epistatic interaction between this QTL on LG15 and the QTL on LG10 in the *melicerta* × *zuleika* brood ( $F(7, 23) = 3.189$ ,



$P = 0.02$ ). Our morphometric analysis allowed detecting some cases where a genomic position is associated with multiple pattern elements on the wing (Figures 1.3AII and 1.3BII). However, this might be caused by different genetic elements, as suggested by the “apparent” pleiotropy of the *HhN*-Hspot cluster in Figure 1.3AII, which may be homologous to *N-Sb* of *H. melpomene* and therefore may represent the action of distinct, tightly linked loci (see Discussion below). Therefore, we do not claim pleiotropic effects of any mapped QTL, even though such effects have previously been reported in *Heliconius* (Table 1.1).

Several suggestive QTLs ( $\text{LOD} \geq 3$ ) located across eight distinct chromosomes (including LG1, LG10 and LG15) were found to be modulating pattern element variations controlled by major QTLs (Figure 1.12 and Table 1.11). Remarkably, we find a suggestive QTL on LG15 in the *boulleti* × *telchinia* brood, which explains the variation controlled by the minor effect locus *HiN* in this brood (Figure 1.12C).

## 1.4 Discussion

### 1.4.1 *Heliconius hecale* and *H. ismenius* bear a multilocus architecture for the control of wing patterning

Using image analysis of wing patterns and linkage mapping based on dense genome-wide genotyping, we have characterised the genetic architecture of mimicry variation in two species, *Heliconius hecale* and *H. ismenius*, belonging to the underexplored “silvaniform” clade of *Heliconius*. These approaches revealed multiple, unlinked colour loci in those species, and showed that the combination of high density genotyping, use of a reference genome, and multivariate phenotypic analysis can yield detailed information on the genetic underpinnings of the major components of adaptive traits, as well as a sensitive description of the effect of individual QTL loci on the variation of such complex traits. Mapping was based both on Mendelian characters traditionally scored by eye and on a multivariate morphometric analysis of whole-wing pattern complexity. The latter does not rely on the subjective detection of variable elements, and proved powerful to extract major components of variation from the complexity of the entire wing pattern variation.

The power and precision of a QTL analysis relies on an accurate phenotypic description, a dense array of molecular markers, and a sufficient number of offspring. The limiting factor here was the number of offspring genotyped (between 29 and 42 individuals), leading to an easier detection of QTLs of large effect. We retrieved each of the individual characters scored manually (mostly of major effect), which validates the relevance of our quantitative analysis and gives credit to the additional QTLs revealed. The credible intervals were relatively narrow around each mapped QTL (Figure 1.11) and (with one exception) encompassed candidate genes known from other studies. This reflects the good resolution introduced by our phenotyping, despite the low number of offspring analysed. Additionally, novel candidate minor effect genomic regions were identified as suggestive QTLs (loci detected with limiting statistical power; Table 1.11). One of the strengths of the method used is that it permitted whole-wing visualisation of all phenotypic changes associated with each QTL separately.

Mapped loci include both Mendelian loci and QTLs affecting relatively large wing regions strongly differentiated between the variants used in our crosses. Minor effect loci (i.e. suggestive QTLs) modify pattern elements also controlled by major loci, and modulate their phenotypic effects and the resemblance to local co-mimics. These findings are consistent with theoretical expectations concerning the distribution of gene effect sizes fixed during mimicry evolution ([Baxter et al., 2008a]). Notably, our results confirm that a multilocus architecture of wing

pattern variation spread out on multiple chromosomes is a feature shared by most species in the genus ([Sheppard et al., 1985]; [Mallet, 1989]; [Naisbit et al., 2003]; [Jones et al., 2011]; [Papa et al., 2013]).

#### 1.4.2 The wing colour architecture in *Heliconius hecale* and *H. ismenius* is largely homologous to the architecture found in other *Heliconius* species

Major wing patterning loci discovered here finely map in homologous positions to major colour loci previously identified in *Heliconius*, and the pattern elements and wing positions affected are generally conserved (Figure 1.4B). In some cases, their effects in *H. hecale* and *H. ismenius* are very similar to what is observed in other species. For instance, *HhK* causes a similar white/yellow switch in *H. hecale* as *K* in *H. cydno* and *H. melpomene* ([Naisbit et al., 2003]; [Kronforst et al., 2006c]; Figure 1.4A). Both loci map near to *wingless* on linkage group (LG) 1, and the combination of positional and phenotypic effect strongly argues for *HhK* to be the ortholog of *K* ([Kronforst et al., 2006c]; Figure 1.4A). *K* is not formally identified yet, but our results for *HhK* exclude the coding region of *wingless*. Similarly, loci *HhAc* and *HiAc* on LG10 control melanisation of the discal region of the forewing, reminiscent of the variation controlled by *Sd* and *Ac* in other species (Figure 1.4, Table 1.1) and identified to the *WntA* gene ([Martin et al., 2012]). Here, *WntA* is in perfect linkage with *HhAc* and *HiAc* (Figure 1.2), and its expression is markedly reduced in *H. h. melicerta* compared to *H. h. zuleika* around the discal crossvein and the adjacent M3-Cu1 and Cu2-Cu1 domains (Figure 1.3AIII). This strongly suggests that *cis*-regulatory variation of *WntA* expression causes the allelic effects of the *HhAc* and *HiAc* loci, revealing the molecular identity of *HhAc/HiAc* and confirming its homology at the gene level to one of the known “toolkit” colour loci in *Heliconius*.

In other cases, however, the phenotypic effects of toolkit loci were quite different in *H. hecale* and *H. ismenius* to their known effects in other species (see Figure 1.4A). The versatility in the effect of these loci across taxa is consistent with their developmental position as switch genes presumed to act relatively early in scale fate determination. Furthermore, this highlights the importance of the interaction of some components of genetic architecture in generating radically different phenotypes, despite an overall conserved multilocus architecture.

LG15 contains three linked loci related to distinct parts of the forewings and hindwings. *HhN/HiN* control the presence/absence of yellow elements in the forewing of the two species, and we hypothesize their homology to the *H. melpomene N*, also situated on LG15 and affecting a similar wing region. Other loci in other *Heliconius* species have been reported to affect the melanisation on the post-discal and subapical regions of the forewing (*Fs* and *L* in *H. cydno*, *Ro* in *H. erato*), but they map to different chromosomes, or their location is unknown ([Sheppard et al., 1985]; [Nijhout et al., 1990]; [Linares, 1996]; [Nadeau et al., 2014]). Interestingly, locus *Ro* maps to LG13 in *H. erato* ([Nadeau et al., 2014]) which shows that similar wing-pattern elements can have a distinct underlying genetic basis in different species. In *H. hecale*, two QTLs (*Hspot* and *Fspot*) are also situated on linkage group 15. The phenotypic effect of *Hspot* (yellow or white spots along the hindwing margin) and its linkage to *HhN* suggest a homology to *Sb*, a locus tightly linked to *N* in *H. melpomene*. In *H. numata silvana*, the supergene *P*, situated on LG15 and presumed to contain the ortholog of *Sb*, also controls a very similar variation along the hindwing margin as in *H. hecale*. Regarding *Fspot*, no locus has been previously described to affect the forewing apical region, presumably because *Heliconius* species studied to date rarely show pattern variation in this wing region. *Fspot* may represent a new wing patterning locus in *Heliconius* with a role for mimicry variation mainly in silvaniform species.

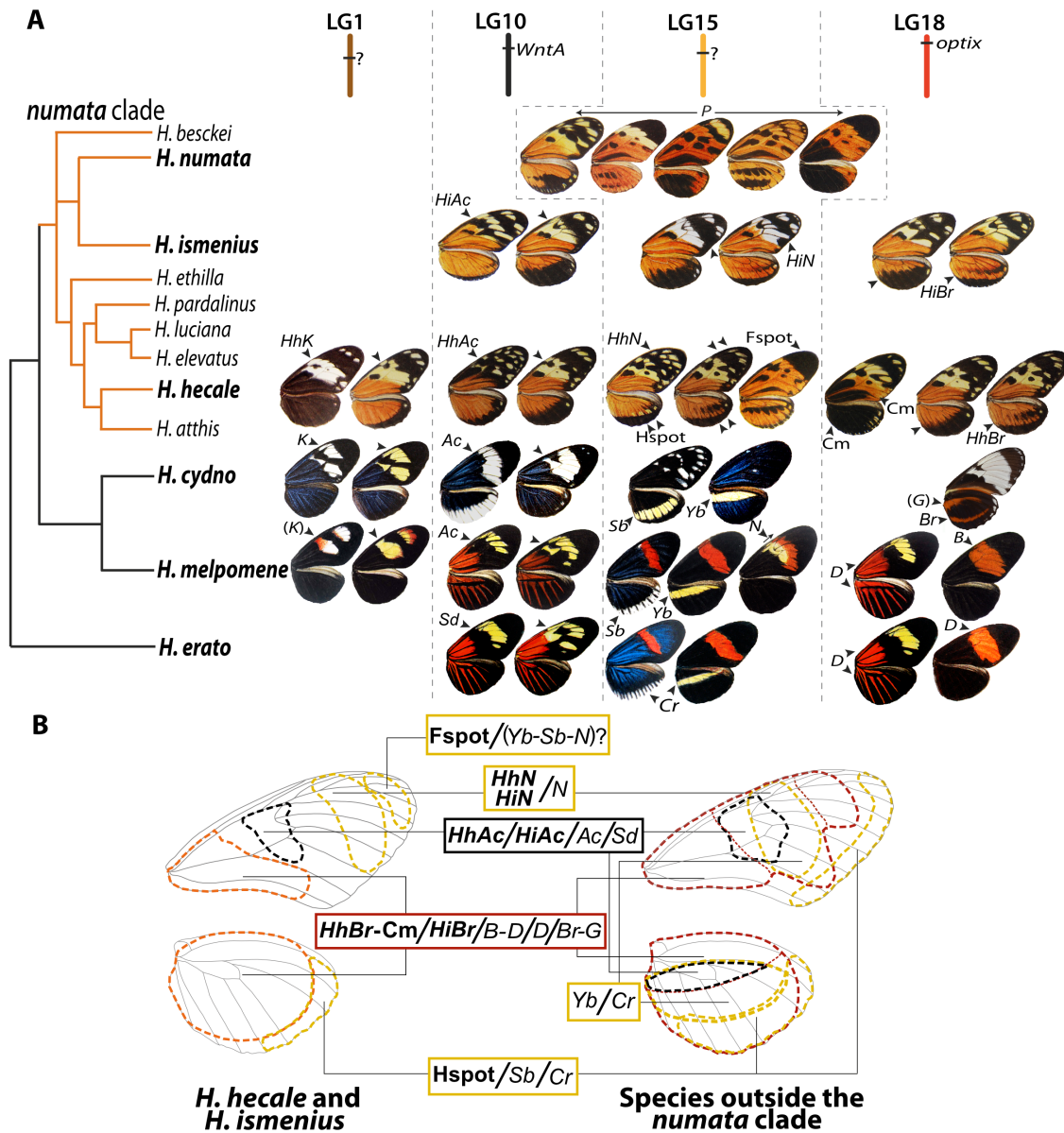


Figure 1.4 – Conservatism and novelty in the genetic architecture underlying the diversity of *Heliconius* wing patterns. (A) Known genetic architectures underlying pattern diversity throughout the clade mapped onto an unscaled phylogeny. Orange tree branches represent nine of the ten species in the silvaniform clade. Major colour variation loci are located on four chromosomes (top) and control variation in similar wing regions (arrows) throughout the genus. Wing phenotypes are represented based on Holzinger and Holzinger (1994). Note that the effect of the *Br* locus in *H. cydno* is shown on the ventral side. Loci with names in brackets were described based exclusively on interspecific crosses. (B) Comparative diagram of the distribution of the gene effects across the wing for toolkit loci in the silvaniform clade (excepting *H. numata*; left) and in the *H. melpomene* and *H. erato* clades (right), showing the general conservatism of the regions affected by homologous elements of the multilocus architecture despite some flexibility.

Despite the lack of functional analyses to pinpoint the causal gene(s) within the *Yb-Sb-N/Cr/P* cluster, molecular signals of selection were reported for the locus *HMEL000025* ([Wu et al., 2010]; [Nadeau et al., 2012]). In *H. ismenius*, fine-mapping excludes the coding region of this gene from the interval for *HiN* (Figure 1.2), but does include its regulatory region, as well as other neighbouring genes. Taken together, fine-mapping and gene effects suggest that *Fspot*, *Hspot* and *HhN* form a cluster in *H. hecale*, partly homologous to the *Yb-Sb-N* cluster in *H. melpomene*, and possibly forming part of the elements participating in supergene *P* in *H. numata* ([Joron et al., 2006], [Joron et al., 2011]).

Finally, LG18 contains *HhBr/HiBr*, mapping close to *optix* both in *H. hecale* and *H. ismenius*. *Optix* underlies the variation controlled by loci *D*, *B-D* and *Br-G* in *H. erato*, *H. melpomene* and *H. cydno*, respectively ([Reed et al., 2011]; [Supple et al., 2013]; [Martin et al., 2014]). Here, the conspicuous variation associated with *HhBr* and *HiBr* affect a similar wing position as *Br* in *H. cydno* ([Gilbert, 2003]; Figure 1.4A) albeit with slightly different phenotypic effects. In *H. ismenius*, mapping excludes the coding region of *optix* but includes a large intergenic region that has been proposed to contain 3' enhancers of *optix* involved in its pattern-related cis-regulatory evolution ([Pardo-Diaz et al., 2012]; [Supple et al., 2013]). While a previous report failed to detect the expression of *optix* in the developing hindwings of *H. hecale fornarina* ([Martin et al., 2014]), this does not rule out a colour patterning role for this gene in silvaniform species. Notably, species-specific delayed expression of *optix* could generate such a negative result in *H. hecale*. The region around *optix* also emerges as the one of largest effect in the melanisation of hindwings and forewings (Cm) in *H. hecale*, specifically at wing positions affected by the *D* locus in *H. melpomene* (Figure 1.4). Interestingly, hindwing melanisation is also associated with markers near *HMEL000025* on LG15, especially at the position of the hindwing bar controlled by *Yb* in *H. melpomene* and *H. cydno*. The lack of significant epistasis between those markers may indicate that these genes largely act additively and relatively independently of each other, but could also be the result of a limited power to detect the epistatic interactions.

Epistasis and dominance are commonly used as criteria to infer gene homology between taxa ([Naisbit et al., 2003]). Here, we did not use these to infer homology since they show wide variations across the genus. For instance, no evident epistasis was detected between loci on LG15 and LG18, in contrast to other species (e.g. *N* and *B* for the forewing submarginal band; [Sheppard et al., 1985]; [Naisbit et al., 2003]). Conversely, loci on LG10 and LG15 show epistasis in *H. hecale*. A similar interaction was also reported between loci *Sd* (LG10) and *Cr* (LG15) in *H. erato* ([Mallet, 1989]). These results suggest that gene interactions can differ between species and their detection depends on individual allele effects. Regarding dominance, the putative orthologs *HhN* (*H. hecale*) and *HiN* (*H. ismenius*) express codominance in different ways. Variation in dominance could be related to variation in mimicry selection pressures. Variable dominance relationships have been reported for some loci across *Heliconius* species ([Joron et al., 2006]) as well as within species ([Le Poul et al., 2014]).

### 1.4.3 The supergene *P* is restricted to *Heliconius numata* and has evolved from a multilocus architecture.

The multilocus architectures of mimicry found in the explored silvaniform species contrast with the single-locus architecture controlling mimicry polymorphism in *H. numata*. This indicates that the supergene evolved uniquely in the *H. numata* lineage from a multilocus architecture shared between its sister species *H. ismenius* and other more distant relatives.

Many *Heliconius* species show geographic variation in their mimicry associations and local populations are usually fixed for a given warning pattern, except in narrow hybrid zones

where unlinked genes segregate freely. A few species such as *H. cydno*, *H. hecale* and *H. ismenius* maintain single-character polymorphisms in some part of their range (e.g. colour switch or presence/absence of a pattern element paralleling similar variation in the local mimicry community ([Brown and Benson, 1974]; [Kapan, 2001])). In contrast, *H. numata* shows rampant polymorphism across its entire range, involving highly differentiated wing patterns and the concerted co-variation of multiple colour elements across the wing. This polymorphism is controlled by a cluster of loci locked into a supergene, allowing the segregation of discrete mimicry types ([Brown and Benson, 1974]; [Joron et al., 1999]). Our data therefore argue for this peculiar architecture to be evolutionarily associated with the maintenance of polymorphism in *H. numata* and further confirm that balancing selection might be shaping the genetic architecture of wing patterns in this species.

Interestingly, single-locus architecture is not associated with a specific type of wing pattern, as both *H. hecale* and *H. ismenius* have “tiger-patterns” very similar to some of the *H. numata* forms (see Figure 1.4A). For instance, *H. numata silvana*, associated with the ancestral allelic class of the *H. numata* supergene ([Joron et al., 2011]), is phenotypically similar to *H. ismenius bouletti* and *H. hecale melicerta*. *H. numata* also participates in mimicry with *H. hecale* or closely related species in many parts of its range ([Brown, 1981]). Mimicry evolution can therefore involve distinct genetic architectures, even though some of the loci may be homologous ([Jones et al., 2011]). Recent research has revealed that similar phenotypes may not always show a parallel genetic basis at the nucleotide or gene level ([Arendt and Reznick, 2008]; [Manceau et al., 2010]; [Elmer and Meyer, 2011]). However, to our knowledge, no cases have been reported where different genetic architectures, in terms of linkage and gene effect size, underlie highly similar phenotypic variations.

Our mapping shows that the *Heliconius* “toolkit” of colour genes is used throughout the *H. numata* clade, where only the supergene architecture was previously known. Within *H. numata*, the large-effect toolkit loci not associated with the supergene play a minor role in pattern variation ([Jones et al., 2011]). The contrasting genetic architectures observed when comparing *H. numata* to other silvaniform species do not relate to differences in the identity of the colour loci themselves, but rather to large variations in the effect size of the loci participating in determining wing patterning. The increase in linkage between elements controlling different regions on the wings may explain the build-up of a large effect supergene in *H. numata*. As previously suggested, the *H. melpomene* loci *Yb*, *N* and *Sb* may be the orthologues of some elements composing the supergene in *H. numata* ([Joron et al., 2006], [Joron et al., 2011]), and our data also suggest the existence of a gene cluster at this position in *Heliconius hecale*. Those elements control variation in distinct regions of the wing (Figure 1.4B), and their co-variation in response to mimicry may participate to an initial build-up of co-adapted clusters in this region, later locked into a supergene in *H. numata* within the  $\sim 400$  Kb genomic region where recombination is suppressed by inversions (c.f. the “sieve” hypothesis; [Turner, 1977]; [Joron et al., 2011]). This may be consistent with the observation that large gene effects may often result from the aggregation of independent small effect mutations ([Martin and Orgogozo, 2013]).

#### 1.4.4 Conservation of genetic architectures does not constrain adaptation

Our results extend our knowledge of the homology of wing colour loci implicated in different adaptive radiations in the genus *Heliconius*, and show how shared genetic architectures are implicated both in mimicry convergence between species and in the diversification associated

with local adaptation ([Joron et al., 2006]; [Papa et al., 2008]; [Reed et al., 2011]; [Martin et al., 2012]). The toolkit of *Heliconius* wing-patterning genes therefore stands as an ancestral architecture shared by species with radically different wing patterns and exposed to different mimicry selection pressures. If mimetic wing patterns are considered as an integrated complex trait, variation in the distribution of individual effect sizes and interactions among the contributing genes across the radiation demonstrates how profoundly malleable these traits are. Using a conserved set of switch genes, novel phenotypes appear to be explored via the effects of those genes on phenotypic variation, presumably through an evolution of the downstream wiring with effector genes, and the possible involvement of new modifiers, rather than by the recruitment of new switch genes. The conservatism in the wing-patterning toolkit does not appear to impose strict limits on the evolution of novel phenotypes, highlighting the power of selection regimes in bringing populations to local adaptive optima.

Here, we have mainly focused on primary components of genetic architecture such as number of loci, genome position, and the distribution of gene effect sizes. However, the mapping populations restricted our capacity to investigate aspects related to interactions such as epistasis, dominance and pleiotropy, which may also respond to selection and contribute to the complexity of the adaptive responses ([Le Poul et al., 2014]). We encourage the use of detailed morphometric quantification of pattern variation on large mapping populations to examine the interaction of the components of genetic architecture and their role in adaptive evolution.

## 1.5 Supplementary information

### 1.5.1 Supplementary methods

#### • 1 • Genotyping methods

Markers were first amplified in brood parents, then in the progeny when allelic variation was found. Where possible, allelic variation was scored using amplicon length variation in 3% agarose gels stained with SYBR®Safe. In the absence of clear size variation, but where Sanger sequencing revealed an appropriate restriction enzyme site difference within the amplicon, size differences in restriction digestion profiles were used to infer genotype. In two cases, competitive allele-specific PCR KASP™ assays (LGC Genomics, UK) were used, following the manufacturer's protocol, with fluorescence measured in a CFX96 Bio-Rad Real-Time PCR machine. In other cases, direct Sanger sequencing of PCR products was used. A list of the used markers and more details about the genotyping methods are given in Table 1.3.

#### • 2 • PCR Conditions

Standard PCR master mixes and thermal cycling conditions were used. PCR mixes contained 10–50 ng of genomic DNA, 1× reaction buffer with 2 mM MgCl<sub>2</sub>, 0.1 mM each dNTP, 50 pmol of each primer and 0.25 U DreamTaq polymerase (Fermentas). Thermal cycling conditions were 94 °C for 1 min, 35 cycles of 94 °C for 15 sec, 55 °C annealing for 30 sec, 72 °C for 60 sec, and a final extension at 72 °C for minutes.

#### • 3 • Multivariate quantitative trait locus (QTL) analysis

**Mapping colour pattern loci** Taking distinct nested subsets of Principal Components (PCs) derived from Colour Pattern Modelling (CPM) may change genome association profiles slightly. Therefore, to detect the most robust QTL peaks we explored genome association profiles using

different subsets of axes by incrementing the number of added PCs. The subset of axes containing all PCs explaining more than 2% of variance in the principal component analysis (PCA) was chosen. Additionally, we explored subsets that encompassed an increasing number of components accounting for down to 1% of explained variance, which revealed additional suggestive QTLs, especially on linkage groups known to harbour wing patterning loci or QTLs in other species. The effect of an additive colour QTL was estimated using Haley-Knott regression ([Haley and Knott, 1992]; [Knott and Haley, 2000]) by fitting the multivariate linear model  $y_i|M_i \sim N_q(\mu + \sum_c x_{ic}\beta_c + \sum_j p_{ij}\beta_j, S)$ , where  $x_{ic}$  is the value of the covariate  $c$  and  $p_{ij} = Pr(g_i = j|M_i)$  is the probability of the QTL genotypes given the flanking markers for individual  $i$  and  $\beta$  are the  $q$ -dimensional effect of the covariate  $c$  or of the genotype  $j$ . These probabilities were computed using *R/qtl* ([Broman et al., 2003]) at each centimorgan along the 20 autosomal chromosomes considering a genotyping error rate of  $10^{-3}$  and the Kosambi map function ([Broman et al., 2002]). Presence of the QTL was evaluated using the Pillai trace criterion ([Pillai, 1967]). The probabilities associated with its approximated  $F$  statistics were transformed to their negative  $\log_{10}$  to make results comparable with LOD scores ([Leamy et al., 2008]). All computations for colour pattern QTL mapping were conducted in the *R/shapeQTL* package written by Nicolas Navarro and available under request. Genome-wide significance threshold. Genome-wide significance for the presence of a QTL was evaluated using a permutation approach ([Churchill and Doerge, 1994]). The trait together with its covariate (sex) was reshuffled among individuals whereas the original genotype probabilities were kept constant. The genome scan was repeated on these data and the maximal score (LOD) was recorded for each of 1000 iterations. We took the 90th quantile of that distribution as the 10% genome-wide threshold to detect QTLs, but defined QTLs with  $LOD \geq 3$  as being suggestive.

**Multiple QTL modeling** We adopted an approach for model search including or dropping QTLs without any prior on the number of QTLs per chromosome following the approach developed for univariate traits by Broman and collaborators ([Broman and Speed, 2002]; [Manichaikul et al., 2006]; [Broman and Sen, 2009]). Searches were restricted to additive QTLs only. Model choice was based on penalized scores using the  $-\log_{10}$  of the  $P$ -value minus the model complexity times the 5% genome-wide threshold.

**Interval estimates of QTL locations** To determine the credible intervals around each significant QTL peak, we used the *bayesint* function of *R/qtl* with a coverage probability of 0.95. Coverage of these intervals was proven stable and consistent across a variety of situations ([Manichaikul et al., 2006]). Intervals were computed from the chromosome profile of the QTL conditional on all other refined QTL positions.

**QTL effects, effect sizes and visualization** QTL effects were estimated conditional on the covariate (sex) and other QTLs included in the final model using multivariate regression of phenotypes on the backcross parameterization of QTL genotype probabilities. The effect of each colour pattern QTL was visualized within Color Patterning Modeling (CPM).

**Test for epistatic interactions** Epistatic interactions between the uncovered QTLs were tested for by implementing a multivariate linear model with additive pairwise interaction between the QTLs.

## 1.5.2 Supplementary tables

Table 1.1 – Summary table of major and minor effect wing colour pattern loci mapped in the genus *Heliconius*

Linkage groups	Colour locus <sup>A</sup>	Causal gene	Phenotypic effect <sup>B</sup>	Species <sup>C</sup>	Loci interacting epistatically with major effect loci	References
1	K, Khw		White/yellow switch of background elements	<i>H. melpomene/cydno</i>		[Jiggins et al., 2005b]; [Joron et al., 2006]; [Kapan, 1998]; [Kronforst et al., 2006c]; [Linares, 1997]; [Naisbit et al., 2003]
				<i>H. hecale</i>		This study
				<i>H. numata</i>		[Jones et al., 2011]
2	*		Red FW band shape Red patterns on both FW and HW	<i>H. melpomene</i>		[Baxter et al., 2008a]
				<i>H. erato</i>		[Nadeau et al., 2014]
				<i>H. melpomene</i>		[Baxter et al., 2008a]
7	Sd, HeSd, St, Ly, c-spot	<i>WntA</i>	Length and shape of yellow FW patch	<i>H. erato/himera</i>	<i>HeD, Cr</i>	[Kapan et al., 2006]; [Kapan et al., 2006]; [Mallet, 1989]; [Martin et al., 2012]; [Nadeau et al., 2014]; [Papa et al., 2013]; [Sheppard et al., 1985]
				<i>H. melpomene</i>		[Jiggins et al., 2005b]; [Kapan et al., 2006]; [Martin et al., 2012]; [Nijhout et al., 1990]; [Sheppard et al., 1985]
				<i>H. melpomene</i>		



10	<b>Ac, Ps</b>	<i>WntA</i>	Discal "hourglass" pattern of FW and width of HW bar	<i>H. cydno/pachinus</i>	[Gallant et al., 2014]; [Gilbert, 2003]; [Kapan et al., 2006]; [Martin et al., 2012]; [Nijhout et al., 1990]
	<b>HhAc</b>	<i>WntA</i>	Melanisation close to FW discal cell	<i>H. hecale</i>	This study
	<b>HiAc</b>	<i>WntA</i>	Melanisation close to FW discal cell	<i>H. ismenius</i>	This study
13	*		Yellow HW band	<i>H. erato</i>	[Nadeau et al., 2014]
			Quantitative FW and HW pattern variation	<i>H. numata</i>	[Jones et al., 2011]
			Red FW band shape	<i>H. melpomene</i>	[Baxter et al., 2008a]
	<i>Ro</i>		Rounding of the distal edge of the upper FW band	<i>H. erato</i>	[Nadeau et al., 2014]
13			Red FW band shape	<i>H. melpomene</i>	[Baxter et al., 2008a]
	<b>Supergene P</b>		Whole FW and HW variation	<i>H. numata</i>	[Jones et al., 2011]; [Joron et al., 2006]
			HW yellow bar	<i>H. erato/himera</i>	[Counterman et al., 2010]; [Kapan et al., 2006]; [Jiggins and McMillan, 1997]; [Joron et al., 2006]; [Mallet, 1989]; [Papa et al., 2008]; [Sheppard et al., 1985]; [Tobler et al., 2004]
	<b>Cr</b>		White HW margin	<i>H. erato/himera</i>	
			Medial yellow FW band	<i>H. erato/himera</i>	

15	<b>Sb</b>	<b>Yb, HmYb, Cs</b>	Yellow HW bar	<i>H. melpomene/cydrno</i>	<i>Br</i>	[Emsley, 1964]; [Ferguson et al., 2010]; [Kapan et al., 2006]; [Jiggins et al., 2005b]; [Joron et al., 2006]; [Linares, 1997]; [Mallet, 1989]; [Nadeau et al., 2014]; [Naisbit et al., 2003]; [Nijhout et al., 1990]; [Sheppard et al., 1985] [Baxter et al., 2008a]
			Pink/orange switch on ventral side of FW band	<i>H. melpomene</i>		
			White or yellow HW margin	<i>H. melpomene/cydrno</i>	Non mapped locus <i>L</i>	[Emsley, 1964]; [Ferguson et al., 2010]; [Jiggins et al., 2005b]; [Joron et al., 2006]; [Linares, 1996]; [Linares, 1997]; [Naisbit et al., 2003]
17	<b>*</b>	<b>Hspot</b>	Marginal spots on HW	<i>H. hecale</i>	<i>HhAc</i>	This study
		<b>N</b>	Yellow colour in the medial FW band	<i>H. melpomene</i>	<i>HmB, HmD</i>	[Mallet, 1989]; [Nadeau et al., 2014]; [Sheppard et al., 1985]
		<b>HhN</b>	Subapical row of spots on FW	<i>H. hecale</i>		This study
		<b>HiN</b>	Shape of distal edge of FW band	<i>H. ismenius</i>		This study
		<b>Fspot</b>	Apical row of spots on FW	<i>H. hecale</i>		This study
17	<b>*</b>		Yellow HW bar	<i>H. erato</i>		[Nadeau et al., 2014]
		<b>HeD, D, DRY</b>	<i>optix</i> Red FW “dennis”	<i>H. erato/himera</i>	<i>Sd, Cr, non mapped modifiers</i>	[Baxter et al., 2008a]; [Counterman et al., 2010]; [Jiggins and McMillan, 1997]; [Kapan et al., 2006]; [Mallet, 1989]; [Martin and Reed, 2014]; [Nadeau et al., 2014]; [Papa et al., 2008]; [Reed et al., 2011]; [Sheppard et al., 1985]
			<i>optix</i> Red HW rays	<i>H. erato/himera</i>		
			<i>optix</i> Red/yellow FW	<i>H. erato/himera</i>		

<b>HmD, D</b>	<i>optix</i>	Red HW "rays" and proximal "dennis" patches on both wings	<i>H. melpomene</i>	<i>N</i>	[Baxter et al., 2008a]; [Mallet, 1989]; [Nadeau et al., 2014]; [Reed et al., 2011]; [Sheppard et al., 1985]; [Turner and Crane, 1962]
<b>HmB, B</b>	<i>optix</i>	Red medial FW band	<i>H. melpomene</i>	<i>N</i>	[Baxter et al., 2008a]; [Mallet, 1989]; [Martin and Reed, 2014]; [Reed et al., 2011]; [Sheppard et al., 1985]; [Turner and Crane, 1962]; [Turner, 1972]
<b>Br, "forceps" shutter</b>	<i>optix</i>	Brown oval pattern on the ventral side of the HW	<i>H. cydno/pachinus</i>	<i>Yb</i>	[Chamberlain et al., 2011]; [Gilbert, 2003]; [Naisbit et al., 2003]; [Reed et al., 2011]
<b>18</b>					
<b>HhBr</b>		Shape of marginal black band on HW	<i>H. hecale</i>		This study
<b>HiBr</b>		Shape of marginal black band on HW	<i>H. ismenius</i>		This study
<b>G, basal red</b>	<i>optix</i>	Basal HW red spots and basal FW small bar	<i>melpomene/cydno/pachinus</i>		[Chamberlain et al., 2011]; [Gilbert, 2003]; [Kapan et al., 2006]; [Linares, 1996]; [Martin and Reed, 2014]; [Naisbit et al., 2003]
<b>Wh (possibly)</b>		White/yellow switch of FW band	<i>H. erato/himera</i>		[Papa et al., 2013]; [Sheppard et al., 1985]
<b>Cm</b>		HW and proximal FW melanisation	<i>H. hecale</i>		This study
		Middle FW band size and shape	<i>H. erato/himera</i>		[Papa et al., 2013]
		Quantitative FW pattern variation	<i>H. numata</i>		[Jones et al., 2011]
<b>19</b>		Quantitative FW pattern variation	<i>H. numata</i>		[Jones et al., 2011]

Sex Z chromo- some	Red FW band shape	<i>H. melpomene</i>	[Baxter et al., 2008a]
	FW band shape	<i>H. erato</i>	[Mallet, 1989]
	“Whiteness” (or “yellowness”) of FW band	<i>H. erato/himera</i>	[Papa et al., 2013]
	Quantitative FW and HW pattern variation	<i>H. numata</i>	[Jones et al., 2011]

A Some loci have alternative names separated by commas. The name of major effect loci is written in bold script. QTLs that were named in this study are written in non-italic script.

B Abbreviations FW and HW correspond to forewing(s) and hindwing(s), respectively.

C Species in which a given locus was described. Slashes separate species that were intercrossed to infer gene homology.

\* Low confidence genotype-to-phenotype association.

Table 1.2 – Sampling localities for the specimens of *Heliconius hecale* and *H. ismenius* used in the crosses

Species	Race	Country	Locality	GPS coordinates
<i>Heliconius hecale</i>	<i>melicerta</i>	Panamá	Gamboa, Colón	N9°7.257, W79°43.591
	<i>zuleika</i>	Panamá	Miramar, Bocas del Toro	N8°59.369, W82°14.597
	<i>clearei</i>	Venezuela	Tumeremo, Bolívar	N6°50.185, W61°35.554
<i>Heliconius ismenius</i>	<i>boulleti</i>	Panamá	Yaviza, Darién	N8°16.783, W77°48.589
	<i>telchinia</i>	Panamá	Miramar, Bocas del Toro	N8°59.369, W82°14.597

Table 1.3 – List of markers used to fine-map colour loci in *H. hecale* and *H. ismenius*

Marker name	Marker type (Annotation on H. melpomene genome)	Scaffold (starting position on H. melpomene genome)	Primer sequence	Genotyping method (enzyme)	Genotyped brood(s)
Around HhK					
KASP_Wg	SNP <sup>8</sup>	HE671174 (588001)	NA <sup>7</sup>	KASP™ assays <sup>6</sup>	Br112, Br117
Wingless_F1	EPIC <sup>1</sup> ( <i>HMELO22601</i> )	HE671174 (1166344)	ATGCTAGTATCGTTGCAGGC	Sanger sequencing	Br112
Wingless_R1		HE671174 (1167592)	CACAACAATGAAGCCGGCAG		
scf670889_genotype_F3	Non-coding	HE670889 (74431)	GACATCTCTGAGTGTGAC	dCAPS <sup>2</sup> ( <i>EcoRI</i> )	Br112
scf670889_genotype_R3a		HE670889 (75124)	ACGCGAGCTTCATTCCGCCGA		
KASP_Trib2	SNP <sup>8</sup>	HE670375 (50059)	NA <sup>7</sup>	KASP™ assays <sup>6</sup>	Br112
EN10_F3 <sup>4</sup>	SCNL <sup>3</sup> ( <i>HMELO11986</i> )	HE671246 (~65983)	ACCAGCTGGACATGATGAGRA	Restriction profiles ( <i>HindIII</i> )	Br112, Br117
EN10_R2 <sup>4</sup>		HE671246 (66241)	ACGRTARGCCTTCAAAGTCARGAAT		
BAP28_F2 <sup>4</sup>	SCNL <sup>3</sup> ( <i>HMELO02583</i> )	HE668177 (139034)	GATCAAAAATGTATCAGTCC	Sanger sequencing	Br112
BAP28_R2 <sup>4</sup>		HE668177 (139761)	GTTATTATCTGGAAGCTGTA		
Around HhAc/HiAc					
WntAscf_genotype_F3	Non-coding	HE668478 (318250)	CTGATACATTGATGAGGA	Sanger sequencing + Amplicon length polymorphism	Br122, Br96, Br101
WntAscf_genotype_R3		HE668478 (319228)	CTCGGAGCCACTCCTGAC		

WntAscf_genotype_F6	Non-coding intergenic (HMELO02905- HMELO02906)	HE668478 (402431)	GTCCTAGCAACTAAGTTC	Restriction profiles ( <i>TaqI</i> )	Br122
WntAscf_genotype_R6a		HE668478 (403235)	GACAKGCTTGATCATCCA		
WntAscf_genotype_F6	Non-coding intergenic (HMELO02905- HMELO02906)	HE668478 (402431)	GTCCTAGCAACTAAGTTC	Sanger sequencing	Br96, Br101
WntAscf_genotype_R6b		HE668478 (403501)	ACTTCTTCGCAGRGGTACA		
WntA_F5	EPIC <sup>1</sup> (HMELO18100)	HE668478 (481158)	CAGGGGTACGTGGTCTGT	Sanger sequencing	Br122
WntA_R5		HE668478 (482177)	CGGCTCACAAATAGTCCGGG		
Chitsynth3_F5	EPIC <sup>1</sup> (HMELO18102)	HE668478 (497654)	CCTTCGACGATGGTGTAACGTGCAAA	Sanger sequencing	Br96, Br101
Chitsynth3_R5		HE668478 (497048)	AGGCACTGGGGCAAGGTGGC		
scf669520_genotype_F2	EPIC <sup>1</sup> (HMELO04236)	HE669520 (87479)	AGTCCTSGGTCCAGCGGA	Amplicon length polymorphism	Br122
cf669520_genotype_R2		HE669520 (88848)	GGTCTTCRACCACTTCA		
<b>Around HhN/HhN</b>					
Fizzyscf_genotype_F5b	EPIC <sup>1</sup> (HMELO000008)	HE667780 (499593)	GCTGCACTGCCGCCAGCTGA	Sanger sequencing	Br122
Fizzyscf_genotype_R5		HE667780 (500190)	GCTCRTCCTCGCTCGCATTA		
Gene25_ex7_F	EPIC <sup>1</sup> (HMELO000025).	HE667780 (670937)	TCTTCCAAATTTAARCTAAYGTCG	Sanger sequencing	Br96, Br101
Gene25_ex7_R	Partially non-coding	HE667780 (670240)	TTGCGATGTCCGCACCTGACG		
Hn25_fizz_F1	EPIC <sup>1</sup> (HMELO000025)	HE667780 (680510)	CGCAACGTTATCGCCTAGAT	Sanger sequencing + Restriction profiles ( <i>XhoI</i> )	Br96, Br101, Br112
Gene25_ex1_R		HE667780 (679516)	GAACTGAAAGCGAAAGMAC		
Fizzyscf_genotype_F4	EPIC <sup>1</sup> (HMELO02028).	HE667780 (841439)	TGGTGGGAATACGCAAGGA	Sanger sequencing	Br96, Br101
Fizzyscf_genotype_R4	Partially non-coding	HE667780 (842349)	TGGACTGTCGGCTTTAAGC		

Fizyyscf_genotype_F8	EPIC <sup>1</sup> (HMELO000041)	HE667780 (1030768)	ACCTCAGAGTGTCGAGCTC	Sanger sequencing	Br96, Br101
Fizyyscf_genotype_R8		HE667780 (1031999)	GATGGTCCAATCAGACTGGC		
<b>Around HhBr/HiBr</b>					
Optixscf_genotype_F5	Non-coding intergenic + partial exons (HMELO01009-HMELO01022)	HE670865 (251935)	CATCCCGATCGCTAAGGCAC	Sanger sequencing	Br96, Br101
Optixscf_genotype_R5		HE670865 (253169)	CAACTTACASCATCAACTGAC		
Optixscf_genotype_F6	EPIC <sup>1</sup> (HMELO01018)	HE670865 (285696)	ATACAGGACTGGGCTCTGA	Sanger sequencing	Br122, Br120
Optixscf_genotype_R6		HE670865 (287142)	CCCAGCTCATTGGAAAGC		
Optixscf_genotype_F14	EPIC <sup>1</sup> (HMELO01014)	HE670865 (308062)	CATCTTTCAGGGAAGTATGG	Restriction profiles ( <i>TaqI</i> )	Br96, Br101
Optixscf_genotype_R14		HE670865 (308886)	GCACATTGCCACGGCTCG		
HmB453k_genotype_F	Non-coding	HE670865 (375602)	AGGTTTCYCSAGACGGTATCCTCA	Amplicon length polymorphism	Br122
HmB453k_genotype_R		HE670865 (376487)	CCGTTGGGTATCCGTAAAGCC		
Optixscf_genotype_F1	Non-coding	HE670865 (380464)	GAGACATRTTARTTTTCTCTGCC	Restriction profiles ( <i>SpHI-HF</i> + <i>TaqI</i> )	Br96, Br101, Br112
Optixscf_genotype_R1		HE670865 (381718)	CACCACITTTGAAAGGAATCGTA		
Optixscf_genotype_F2	Non-coding	HE670865 (388302)	GTTCTGACITCTCTGAGATCATCAT	Amplicon length polymorphism	Br122
Optixscf_genotype_R2		HE670865 (389396)	AGGCATAGACAGTTGAGCGGAGA		
Optixscf_genotype_F10	EPIC <sup>1</sup> (HMELO01038)	HE670865 (498086)	ACGCCCCATTCAGGTAGCCTC	Sanger sequencing	Br96, Br101
Optixscf_genotype_R10		HE670865 (499708)	GTGGTATTGAACGATGGTC		

<sup>1</sup> Exon-primed intron-crossing; <sup>2</sup> Derived Cleaved Amplified Polymorphic Sequences; <sup>3</sup> Single-copy nuclear loci; <sup>4</sup> Marcus Kronforst, pers. comm.; <sup>5</sup> [Martin et al., 2012]; <sup>6</sup> Kompetitive Allele Specific PCR (LGC Genomics); <sup>7</sup> Not applicable; <sup>8</sup> Single-nucleotide polymorphism



Table 1.4 – Allelic segregation of colour pattern loci in the mapping families of *H. hecale* and *H. ismenius*

<i>Heliconius hecale melicerta</i> × <i>H. hecale clearei</i> crosses							
Family	Colour locus	Paternal genotype	Maternal genotype	Number of progeny			Expected Mendelian ratio
				<i>HhK<sub>c</sub>HhK<sub>m</sub></i>	<i>HhK<sub>m</sub>HhK<sub>m</sub></i>	Not scored	$\chi^2$ P-value
Br112		<i>HhK<sub>c</sub>HhK<sub>m</sub></i>	<i>HhK<sub>m</sub>HhK<sub>m</sub></i>	84	98	0	1:1 1.0769 0.2994
Br117	<b>HhK</b>	<i>HhK<sub>c</sub>HhK<sub>m</sub></i>	<i>HhK<sub>m</sub>HhK<sub>m</sub></i>	49	58	0	1:1 0.7570 0.3843
Br112 + Br117		NA	NA	133	156	0	1:1 1.8304 0.1761
<i>Heliconius hecale melicerta</i> × <i>H. hecale zuleika</i> crosses							
				<i>HhAc<sub>2</sub>HhAc<sub>m</sub></i>	<i>HhAc<sub>m</sub>HhAc<sub>m</sub></i>	Not scored	
Br122		<i>HhAc<sub>2</sub>HhAc<sub>m</sub></i>	<i>HhAc<sub>m</sub>HhAc<sub>m</sub></i>	46	50	1	1:1 0.1667 0.6831
Br120	<b>HhAc</b>	<i>HhAc<sub>2</sub>HhAc<sub>m</sub></i>	<i>HhAc<sub>m</sub>HhAc<sub>m</sub></i>	7	12	3	1:1 1.3158 0.2513
Br120 + Br122		NA	NA	53	62	4	1:1 0.7043 0.4013
				<i>HhN<sub>2</sub>HhN<sub>2</sub></i>	<i>HhN<sub>2</sub>HhN<sub>m</sub></i>	<i>HhN<sub>m</sub>HhN<sub>m</sub></i>	Not scored
Br122		<i>HhN<sub>2</sub>HhN<sub>m</sub></i>	<i>HhN<sub>2</sub>HhN<sub>m</sub></i>	26	32	27	13 1:2:1 3.6558 0.1607
Br120	<b>HhN</b>	<i>HhN<sub>2</sub>HhN<sub>m</sub></i>	<i>HhN<sub>m</sub>HhN<sub>m</sub></i>	NA	11	7	4 1:1 0.8889 0.3458
Br120 + Br122		NA	NA	NA	NA	NA	NA NA NA

Family	Colour locus	Paternal genotype	Maternal genotype	Number of progeny			Expected Mendelian ratio	$\chi^2$	P-value
				<i>HhBr<sub>z</sub>HhBr<sub>m</sub></i>	<i>HhBr<sub>m</sub>HhBr<sub>m</sub></i>	Not scored			
Br122		<i>HhBr<sub>z</sub>HhBr<sub>m</sub></i>	<i>HhBr<sub>m</sub>HhBr<sub>m</sub></i>	49	44	5	1:1	0.2688	0.6041
Br120	<b>HhBr</b>	<i>HhBr<sub>z</sub>HhBr<sub>m</sub></i>	<i>HhBr<sub>m</sub>HhBr<sub>m</sub></i>	8	11	3	1:1	0.4737	0.4913
Br120 + Br122		NA	NA	57	55	8	1:1	0.0357	0.8501

<i>Heliconius ismenius bouletti</i> × <i>H. ismenius telchinia</i> crosses									
Family	Colour locus	Paternal genotype	Maternal genotype	Number of progeny				Expected Mendelian ratio	P-value
				<i>HiAc<sub>t</sub>HiAc<sub>b</sub></i>	<i>HiAc<sub>b</sub>HiAc<sub>b</sub></i>	Not scored			
Br96		<i>HiAc<sub>t</sub>HiAc<sub>b</sub></i>	<i>HiAc<sub>b</sub>HiAc<sub>b</sub></i>	23	13	0	1:1	2.7778	0.0956
Br101	<b>HiAc</b>	<i>HiAc<sub>t</sub>HiAc<sub>b</sub></i>	<i>HiAc<sub>b</sub>HiAc<sub>b</sub></i>	8	10	0	1:1	0.2222	0.6374
Br196 + Br101		NA	NA	31	23	0	1:1	1.1852	0.2763
Family	Colour locus	Paternal genotype	Maternal genotype	Number of progeny				Expected Mendelian ratio	P-value
				<i>HiN<sub>t</sub>HiN<sub>b</sub></i>	<i>HiN<sub>b</sub>HiN<sub>b</sub></i>	Not scored			
Br96		<i>HiN<sub>t</sub>HiN<sub>b</sub></i>	<i>HiN<sub>b</sub>HiN<sub>b</sub></i>	15	21	0	1:1	1	0.3173
Br101	<b>HiN</b>	<i>HiN<sub>t</sub>HiN<sub>b</sub></i>	<i>HiN<sub>b</sub>HiN<sub>b</sub></i>	8	10	0	1:1	0.2222	0.6374
Br196 + Br101		NA	NA	23	31	0	1:1	1.1852	0.2763
Family	Colour locus	Paternal genotype	Maternal genotype	Number of progeny				Expected Mendelian ratio	P-value
				<i>HiBr<sub>t</sub>HiBr<sub>b</sub></i>	<i>HiBr<sub>b</sub>HiBr<sub>b</sub></i>	Not scored			
Br96		<i>HiBr<sub>t</sub>HiBr<sub>b</sub></i>	<i>HiBr<sub>b</sub>HiBr<sub>b</sub></i>	16	20	0	0.4444	0.505	
Br101	<b>HiBr</b>	<i>HiBr<sub>t</sub>HiBr<sub>b</sub></i>	<i>HiBr<sub>b</sub>HiBr<sub>b</sub></i>	11	7	0	1:1	0.8889	0.3458
Br196 + Br101		NA	NA	27	27	0	1:1	0	1

Table 1.6 – Summary of RAD library sequencing and mapping statistics.

	Library 1 ( <i>melicerta</i> × <i>zuleika</i> ) <i>N</i> = 64		Library 2 ( <i>melicerta</i> × <i>clearei</i> ) <i>N</i> = 64		Library 3 ( <i>boulleti</i> × <i>telchinia</i> ) <i>N</i> = 64		All libraries combined <i>N</i> = 192	
Filtering steps	Number of reads	Percentage (%) <sup>1</sup>	Number of reads	Percentage (%) <sup>1</sup>	Number of reads	Percentage (%) <sup>1</sup>	Number of reads	Percentage (%) <sup>1</sup>
Sequenced reads	247,783,430		162,166,578		343,911,015		753,861,023	
Retained reads	188,974,386	76.27	137,315,022	84.68	290,156,532	84.37	616,445,940	81.77
Mapped reads	177,609,643	93.99	129,572,241	94.36	272,334,861	93.86	579,516,745	94.01
Reads after deduplication	23,454,960	13.21	16,954,588	13.09	26,418,741	9.70	66,828,289	11.53
		9.47 <sup>2</sup>		10.46 <sup>2</sup>		7.68 <sup>2</sup>		8.86 <sup>2</sup>

<sup>1</sup> Percentage is expressed relative to the number of reads in the previous filtering step<sup>2</sup> Percentage of reads after deduplication relative to the total number of sequenced reads

Table 1.7 – Number of per-individual 100-bp Illumina reads (sequenced and filtered through the first quality filtering steps) in Library 1 (Brood 122: *H. hecale melicerta* × *H. h. zuleika*)

Barcode	Sample	Sequenced reads	Retained reads	Mapped reads	Reads after deduplication
<b>CGATA</b>	<b>MJ11-2718<sup>1</sup></b>	<b>25732118</b>	<b>18438392</b>	<b>16778275</b>	<b>1045850</b>
ACACG	MJ11-2688	9948346	7205156	6791939	949641
CCAAC	MJ11-2779	9939526	7779076	7337787	912266
ATATC	MJ11-2731	10319068	7483572	7059846	907840
AGGAC	MJ11-2730	9181841	7292064	6918616	892152
ATGCT	MJ11-2776	9207657	7267068	6886915	877760
<b>CGGCG</b>	<b>MJ11-2463<sup>2</sup></b>	<b>8396067</b>	<b>6994052</b>	<b>6626957</b>	<b>854830</b>
AGAGT	MJ11-2729	8855104	6591020	6252333	847735
CCGGT	MJ11-2780	8858882	7125696	6548152	844874
CAGTC	MJ11-2778	8446873	6659330	6284163	838981
CAACT	MJ11-2777	8374724	6177230	5830719	806271
TTGGC	MJ11-2642	6671879	5386058	5026425	739315
AAAAA	MJ11-2685	6210214	4324324	4105322	667760
GCATT	MJ11-2612	5442139	4441378	4223045	579166
GGAAG	MJ11-2614	5164532	4049114	3848970	554353
AAGGG	MJ11-2686	5479533	2903670	2743489	509770
GCGCC	MJ11-2613	4379898	3631954	3449297	499812
GGGGA	MJ11-2616	4541978	3617850	3440234	499135
TAGCA	MJ11-2630	4546801	3690104	3502306	490533
TGGTT	MJ11-2640	4284587	3425896	3259201	475009
GAAGC	MJ11-2591	4057737	3302644	3137934	465573
ACGTA	MJ11-2728	3997431	2602282	2435934	451995
TAATG	MJ11-2627	4060992	3208364	3053018	443538
GTGTG	MJ11-2626	3929706	3096300	2943795	431993
CTAGG	MJ11-2580	3577563	2814156	2672368	403496
GAGAT	MJ11-2596	3298722	2410394	2223612	338276
CATGA	MJ11-2813	4861128	3930732	3617383	337509
TCTCT	MJ11-2963	4956151	3547176	3260669	321423
GTCAC	MJ11-2786	1928458	1584916	1495332	316396
GGTTC	MJ11-2787	1662645	1375690	1298944	283395
TATAC	MJ11-2965	2193856	1730656	1639341	275326
TCAGA	MJ11-2637	2322884	1788856	1698940	272777
TGCAA	MJ11-2848	3302057	2779298	2618276	267625
TTTTA	MJ11-2845	3255758	2436620	2304275	254513
TTAAT	MJ11-2641	2067017	1602960	1516407	247341
TGTGG	MJ11-2847	3202932	2453990	2304730	244115
GGCCT	MJ11-2788	1343268	1065984	1009162	234640

TACGT	MJ11-2966	1901846	1532234	1451874	232105
TCCTC	MJ11-2964	1598237	1272922	1205446	210738
CACAG	MJ11-2844	2758648	1958736	1844666	207073
CCCCA	MJ11-2812	2199867	1804264	1702524	202929
TTCCG	MJ11-2846	2298902	1863760	1756883	197950
TGACC	MJ11-2639	1479142	1158840	1101095	192258
CTGAA	MJ11-2581	1484154	1143548	1086926	183045
GTACA	MJ11-2625	1470511	1078542	1023942	174039
CGCGC	MJ11-2810	1402412	1212288	1132074	140201
CGTAT	MJ11-2809	1925999	1229322	1165702	138446
CTCTT	MJ11-2808	1469152	1129968	1072183	135665
TCGAG	MJ11-2638	1066706	692680	658963	114911
GTTGT	MJ11-2785	604765	449850	427501	108821
GACTA	MJ11-2804	1044231	892068	848964	107587
ATCGA	MJ11-2784	564604	432496	408847	105636
CTTCC	MJ11-2807	1015487	785930	744019	94790
GATCG	MJ11-2791	1122903	837066	725156	92291
AGCTG	MJ11-2806	873478	712370	670957	88462
GCTAA	MJ11-2789	854270	708964	671480	84591
CCTTG	MJ11-2811	922259	668524	633733	82567
AGTCA	MJ11-2805	703939	563242	533604	70742
GCCGG	MJ11-2790	548449	461576	436666	59917
ATTAG	MJ11-2782	350203	165148	156696	44637
ACTGC	MJ11-2657	20076	1760	1655	798
AATTT	MJ11-2653	49739	1630	1489	683
AACCC	MJ11-2651	9263	1276	1198	549
ACCAT	MJ11-2656	44116	1360	1289	544
<b>Total</b>		<b>247 783 430</b>	<b>188 974 386</b>	<b>177 609 643</b>	<b>23 454 960</b>

<sup>1</sup> Mother of brood 122

<sup>2</sup> Father of brood 122

Table 1.8 – Number of per-individual 100-bp Illumina reads (sequenced and filtered through the first quality filtering steps) in Library 2 (Brood 112: *H. hecale melicerta* × *H. h. clearei*)

Barcode	Sample	Sequenced reads	Retained reads	Mapped reads	Reads after deduplication
CGATA	MJ11-2706 <sup>1</sup>	31586600	26752988	24751360	1753361
CGGCG	MJ11-2296 <sup>2</sup>	8892296	7738266	7313679	1085291
CAGTC	MJ11-2711	5281174	1560946	1485986	479463
GAAGC	MJ11-2430	3056550	2112546	2007728	478822
AAGGG	MJ11-2670	5328565	2630936	2513321	469090
AGAGT	MJ11-2690	5307831	2461396	2348132	460205
ACACG	MJ11-2672	4686169	2370502	2269696	451060
GAGAT	MJ11-2432	3054033	2265006	2159296	440877
TCGAG	MJ11-2543	2905596	2304474	2192958	439031
TAGCA	MJ11-2522	2867345	2064096	1967730	438357
GCATT	MJ11-2433	2818020	1608594	1530684	424645
GCGCC	MJ11-2467	2610108	1988918	1893799	419014
TAATG	MJ11-2521	2687770	2198134	2093376	414354
ATATC	MJ11-2693	4653598	2377810	2270310	414315
GGAAG	MJ11-2483	2845371	2108804	2000793	413891
CCGGT	MJ11-2715	4349906	2407456	2292018	405284
TCAGA	MJ11-2539	2481961	697724	663884	395128
ACGTA	MJ11-2678	4390208	898550	843157	394874
CAACT	MJ11-2708	4215097	984888	936834	393694
CTGAA	MJ11-2428	2517675	1772692	1681187	393525
AGGAC	MJ11-2691	4221221	4684	4396	392870
GGGGA	MJ11-2486	2466810	4558	4030	391989
ATGCT	MJ11-2695	4271691	3956	3478	391671
CCAAC	MJ11-2714	4335602	3076	2781	383291
GTGTG	MJ11-2492	2444909	3518438	3237167	370833
AAAAA	MJ11-2648	4053702	4388290	4182964	369351
TTGGC	MJ11-2593	2102579	4031774	3846639	343279
GTACA	MJ11-2491	2004021	3780838	3582411	314747
CTAGG	MJ11-2403	1919328	4471800	4268083	308116
TGCAA	MJ11-2874	2101028	3562670	3370565	208582
TTAAT	MJ11-2592	1179074	3863334	3662897	199176
TCTCT	MJ11-2882	1925304	3602622	3422215	198716
TGGTT	MJ11-2578	1175456	3630180	3433194	182746
TATAC	MJ11-2886	1788779	4538612	4304519	181636
TCCTC	MJ11-2883	1668337	3661136	3462336	172084
TTTTA	MJ11-2840	1575227	3774034	3508621	161672
TGTGG	MJ11-2872	1528011	106116	100190	155147

TACGT	MJ11-2889	1482460	222490	206870	152859
TGACC	MJ11-2577	901106	98452	93130	149028
CCCCA	MJ11-2830	1231483	850112	802002	135213
CATGA	MJ11-2835	1213001	742494	709037	126761
TTCCG	MJ11-2871	1085307	625206	597542	113251
GTCAC	MJ11-2753	968012	625700	597286	108186
GGTTC	MJ11-2754	849310	188108	142238	95484
GGCCT	MJ11-2755	729743	277804	264490	80997
GACTA	MJ11-2769	724078	638648	610420	79232
GCTAA	MJ11-2759	708833	277024	262923	78435
CGTAT	MJ11-2799	748972	496068	472509	77493
CGCGC	MJ11-2815	571647	369196	352124	65859
AGCTG	MJ11-2793	576094	430126	410663	64824
CTCTT	MJ11-2797	514797	593006	565555	56653
CTTCC	MJ11-2795	431437	500100	476158	47334
CCTTG	MJ11-2816	377567	315946	300598	40945
GATCG	MJ11-2763	328597	1057588	1008614	37489
AGTCA	MJ11-2770	323388	1060754	978067	37018
CACAG	MJ11-2839	282425	238696	222314	31191
ATCGA	MJ11-2748	259617	1358780	1296008	29519
GCCGG	MJ11-2762	216249	937192	890200	25964
ATTAG	MJ11-2747	134276	1333234	1237185	14746
GTTGT	MJ11-2750	132182	1856788	1768236	13690
AATTT	MJ11-2609	29376	1661118	1587855	649
AACCC	MJ11-2608	8429	1458554	1391294	644
ACCAT	MJ11-2633	28437	1562614	1493977	536
ACTGC	MJ11-2636	12803	1288380	1222532	399
<b>Total</b>		<b>162 166 578</b>	<b>137 315 022</b>	<b>129 572 241</b>	<b>16 954 588</b>

<sup>1</sup> Mother of brood 112

<sup>2</sup> Father of brood 112

Table 1.9 – Number of per-individual 100-bp Illumina reads (sequenced and filtered through the first quality filtering steps) in Library 3 (Broods 96 and 101 mainly: *H. ismenius bouletti* × *H. h. telchinia*)

Barcode	Sample	Sequenced reads	Retained reads	Mapped reads	Reads after deduplication
CTAGG	MJ11-2264 <sup>1</sup>	38739993	32662854	30342291	1892527
CGATA	MJ11-2363 <sup>2</sup>	35389452	30194930	28576405	1805661
CGGCG	MJ11-2406 <sup>3</sup>	20585884	17325512	15399581	1383246
CAACT	MJ11-2402	11850348	10062816	9479636	1142718
AAGGG	MJ11-2329	8458331	6960246	6539455	918845
GCATT	MJ11-2301	12787618	10934104	10339925	816704
AAAAA	MJ11-2328	7597081	6342870	5984500	812860
TGGTT	MJ11-2318	12740143	10897012	10283756	799158
GAAGC	MJ11-2290	12163923	10553168	9936356	745201
ATATC	MJ11-2368	7975510	6518902	6197532	742741
ACACG	MJ11-2364	5624189	4709820	4491874	717336
CAGTC	MJ11-2408	5959139	5034456	4712162	701873
CCAAC	MJ11-2426	5881182	4974108	4687226	684329
ATGCT	MJ11-2369	5788804	4842640	4621616	677135
GGGGA	MJ11-2304	9563462	8086646	7622086	661028
TAGCA	MJ11-2313	8668218	7495530	6988366	589846
GGAAG	MJ11-2303	8074931	6585794	6268711	561974
TGACC	MJ11-2317	6893188	5865872	5462717	529000
TCAGA	MJ11-2315	7144977	6181174	5884549	516684
GAGAT	MJ11-2300	7201087	5766030	5457954	513365
GCGCC	MJ11-2302	6894298	5970946	5685755	513153
CCGGT	MJ11-2427	3846722	3211670	3031477	501911
GTACA	MJ11-2305	6605964	5678316	5281624	499396
CTGAA	MJ11-2314	6084170	5239820	4933966	462258
TAATG	MJ11-2312	6127570	5031616	4750616	438485
TCGAG	MJ11-2316	5605360	4749388	4471598	431093
GTGTG	MJ11-2306	5673645	4796346	4477293	427402
AGAGT	MJ11-2366	3326990	2630738	2516304	418175
TGCAA	MJ11-2687*	4703459	4046032	3857234	352986
ACGTA	MJ11-2365	2490780	2082294	1977996	349413
TCCTC	MJ11-2652*	4034566	3480000	3307773	318913
TCTCT	MJ11-2849*	4335342	3622550	3446747	318703
TGTGG	MJ11-2655*	4237022	3618720	3449153	317078
TTGGC	MJ11-2320	3534799	3039474	2833474	314358
TTCCG	MJ11-2617*	3788221	3211864	3006985	284027
TATAC	MJ11-2781*	3682044	3118546	2975126	278910



AGGAC	MJ11-2367	2103408	1772862	1663580	268682
TTAAT	MJ11-2319	2855219	2265572	2149925	253072
TTTTA	MJ11-2998	3256291	2651558	2504958	225434
GGTTC	MJ11-2379	1218766	1012952	954444	186409
TACGT	MJ11-2771**	2248345	1858028	1775254	176892
CATGA	MJ11-2574	2113463	1809542	1670742	163741
CCCCA	MJ11-2538	1707097	1472432	1387340	160200
GTCAC	MJ11-2355	942327	794328	742446	150804
CACAG	MJ11-2575	1716787	1463884	1374591	141343
GGCCT	MJ11-2380	784393	637040	599575	126435
GCTAA	MJ11-2495	768449	641738	600169	123719
AGTCA	MJ11-2531	1414800	1218734	1149939	115585
GTTGT	MJ11-2999	731729	586036	473402	100733
CGTAT	MJ11-2535	1261035	1020876	961773	100591
CTCTT	MJ11-2534	1255996	1029298	910773	97738
CCTTG	MJ11-2537	1123829	934320	878531	90555
ATCGA	MJ11-2514	594343	465982	434760	81186
CGCGC	MJ11-2536	857887	728766	683611	80286
CTTCC	MJ11-2533	983648	822396	775596	80273
GATCG	MJ11-2525	449316	374746	352131	77861
GACTA	MJ11-2530	403094	334220	318600	68845
AGCTG	MJ11-2532	573222	452954	425768	48395
GCCGG	MJ11-2496	220997	183268	171040	40981
ATTAG	MJ11-2443	136823	93956	88736	19320
ACTGC	MJ11-2327	21869	5498	5004	763
AACCC	MJ11-2324	8616	696	681	162
AATTT	MJ11-2325	54797	574	541	125
ACCAT	MJ11-2326	46057	1472	1132	121
<b>Total</b>		<b>343 911 015</b>	<b>290 156 532</b>	<b>272 334 861</b>	<b>26 418 742</b>

<sup>1</sup> Father of broods 96 and 101

<sup>2</sup> Mother of brood 96

<sup>3</sup> Mother of brood 101

\*Individuals belonging to brood 122

\*\*Individual belonging to brood 112

Table 1.10 – Summary of RAD library marker types and segregation patterns

Cross	Genotyped sites (variable and not) <sup>1</sup>	Female informative markers <sup>2</sup>	Male informative markers <sup>3</sup>	Intercross markers <sup>4</sup>	Total informative sites	Segregation patterns				Markers in maps <sup>5</sup>
						Supported by 1 marker	Supported by 2 markers	Supported by $\geq 3$ markers	Total	
<b>melicerta</b> × <b>zuleika</b>	187,559	1,066	1,106	484	2,656	346	123	182	651	526
<b>melicerta</b> × <b>clearei</b>	142,662	1,085	749	310	2,144	342	124	162	628	546
<b>boulleti</b> × <b>telchinia</b>	290,275	738	715	307	1,760	512	113	69	694	270
<b>All broods combined</b>	620,496	2,889	2,570	1,101	6,560	1,200	360	413	1,973	1,342
<b>Average across broods</b>	206,832	963	857	367	2,187	400	120	138	658	447

<sup>1</sup> Markers for which SNPs were called in  $\geq 80\%$  offspring ( $\geq 10\times$  and  $< 200 \times /249\times$  for offspring/parents respectively,  $GQ \geq 30$ ,  $MQ \geq 40$ ), missing data was imputed in Beagle and Mendelian inconsistencies were excluded

<sup>2</sup> Markers heterozygous in mother and homozygous in father, excluding those not fitting a 1:1 ratio

<sup>3</sup> Markers heterozygous in father and homozygous in mother, excluding those not fitting a 1:1 ratio

<sup>4</sup> Markers heterozygous in both parents, excluding those not fitting 1:2:1 ratio

<sup>5</sup> Markers mapped to the 20 autosomes within *Joinmap* (Van Oijen and Voorrips, 2001)

Table 1.11 – Summary of major and minor (suggestive) QTLs mapped in *H. hecale* and *H. ismenius*

Chromosome	Name <sup>A</sup>	Homologous colour locus <sup>B</sup>	Candidate gene	Phenotypic effect <sup>C</sup>	Cross	Scaffold(s) <sup>D</sup>
1	–			Variation of the FW yellow apical dots and the marginal edge of the FW band	<i>H. hecale melicerta</i> × <i>H. h. clearei</i>	HE672073- HE672064
7	–			Position of the proximal edge of the HW black marginal band	<i>H. hecale melicerta</i> × <i>H. h. zuleika</i>	HE671567
10	–			Black/orange switch near the HW costa and shape of the distal edge of the FW band	<i>H. hecale melicerta</i> × <i>H. h. clearei</i>	HE671252- HE668478
10	HhAc	Ac ( <i>H. melpomene</i> ), Sd ( <i>H. erato/himera</i> ), Ac ( <i>H. cydno/pachinus</i> )	WntA	Size and position of black patterns close to the FW discal cell	<i>H. hecale melicerta</i> × <i>H. h. zuleika</i>	HE668478 (378)
10	HiAc	Ac ( <i>H. melpomene</i> ), Sd ( <i>H. erato/himera</i> ), Ac ( <i>H. cydno/pachinus</i> )	WntA?	Size and position of black patterns close to the FW discal cell	<i>H. ismenius bouletti</i> × <i>H. i. telchinia</i>	HE668478 (519)
12	–			Black/orange switch on median region of HW	<i>H. hecale melicerta</i> × <i>H. h. clearei</i>	HE672051
13	–			HW and proximal FW melanisation, and shape of the proximal edge of the FW band	<i>H. hecale melicerta</i> × <i>H. h. clearei</i>	HE670176*-, HE671149

13	–		Shape of the proximal edge of the HW black marginal band and variation of the FW yellow apical dots	<i>H. hecale melicerta</i> × <i>H. h. zuleika</i>	HE671149- HE670666
15	<b>Fspot;</b> <i>HhN</i>	<b>Fspot</b> = ?; <i>HhN</i> = <i>N</i> ? ( <i>H. melpomene</i> )	<b>Variation of the FW yellow apical dots (Fspot)</b> , shape of the distal edge of the FW band ( <i>HhN</i> ) and melanisation of the HW and proximal FW region	<b><i>H. hecale melicerta</i> × <i>H. h. clearei</i></b>	<b>HE667780 (623)</b>
15	–		Melanisation of the HW and the proximal FW region	<i>H. hecale melicerta</i> × <i>H. h. clearei</i>	HE671193
15	<b>Hspot;</b> <i>HhN</i>	<b>Hspot</b> = <b>Sb</b> ( <i>H. melpomene</i> /cydno); <i>HhN</i> = <b>N</b> ( <i>H. melpomene</i> )	<b>Variation in the number of yellow spots on the margin of the HW (Hspot) and the presence/absence of the third row of yellow FW spots (HhN)</b>	<b><i>H. hecale melicerta</i> × <i>H. h. zuleika</i></b>	<b>HE667780 (295)</b>
15	<i>HhN</i>	<i>N</i> ( <i>H. melpomene</i> )	Variation in the distal area of the FW band ( <i>HhN</i> ) and melanisation on the HW median and distal area	<i>H. ismenius bouletti</i> × <i>H. i. telchinia</i>	HE667780
16	–		Black/orange switch near the HW costa and shape of the distal edge of the FW band	<i>H. hecale melicerta</i> × <i>H. h. clearei</i>	HE668274
17	–		Shape of the distal edge of the FW band	<i>H. ismenius bouletti</i> × <i>H. i. telchinia</i>	HE671784- HE671089
18	<b>Cm</b>	<b>B, D</b> ( <i>H. melpomene</i> ), <b>D</b> ( <i>H. erato</i> )	<b>Black/orange switch on the HW and the proximal FW region</b>	<b><i>H. hecale melicerta</i> × <i>H. h. clearei</i></b>	<b>HE670865 (629)</b>

18	HhBr	Br (H. cydno/pachinus)	optix?	Shape of the black marginal band of the HW	H. hecale melicerta × H. h. zuleika	HE670865 (652)
18	HiBr	Br (H. cydno/pachinus)	optix?	Shape of the black marginal band of the HW	H. ismenius bouletti × H. i. telchinia	HE670865 (293)

- A Dashes correspond to unnamed QTLs. Traits segregating in Mendelian ratios are in italic script. Traits written in non-italics refer simultaneously to the quantitative trait and the QTL of major effect associated with it. Major effect loci are written in bold, minor QTLs are written in non-bold. Minor effect loci usually correspond to suggestive QTLs detected using a lower threshold ( $LOD > = 3$ ) and might act as modifiers of pattern elements also controlled by major loci.
- B One of the names given to Mendelian colour loci being homologous to the QTLs in H. hecale and H. ismenius in other Heliconius species (shown in brackets and separated by a slash when their detection involved interspecific crosses). The equality symbol represents homology in cases where more than one pattern element is associated with the same QTL. Detailed information about the colour loci mapped in other species can be found in Table 1.1.
- C Heatmaps describing the phenotypic effect of each QTL are shown in Figures 1.3 and 1.12. FW stands for forewing(s) and HW for hindwing(s).
- D Most likely location of a given QTL on a genomic scaffold or between a pair of scaffolds (separated by a hyphen). For major effect loci, segregation patterns corresponding to those in Figure 1.11 are shown in brackets.

\*Unmapped scaffold.

## 1.5.3 Supplementary figures

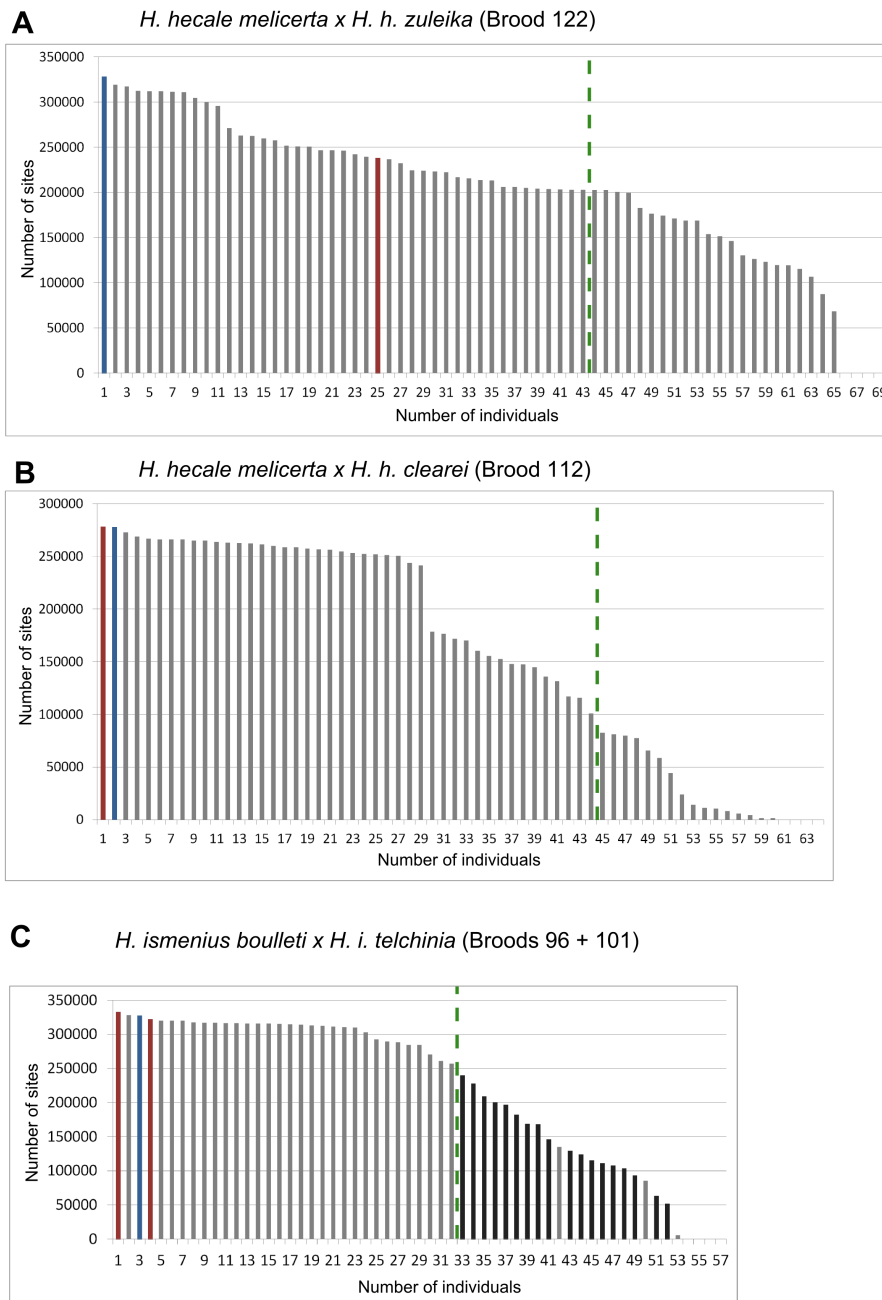


Figure 1.5 – **Number of retained positions after filtering** genotype calls for coverage ( $> 10\times$  and  $< 200\times/249\times$  for offspring/parents, respectively), genotype quality ( $GQ \geq 30$ ) and mapping quality ( $MQ \geq 40$ ) with a custom *Perl* script ([Kanchon Dasmahapatra 2012, *pers. comm.*]). (A) *H. hecale melicerta* x *H. h. zuleika*, (B) *H. hecale melicerta* x *H. h. clearei* and (C) *H. ismenius bouletti* x *H. i. telchinia* broods (grey and black bars denote brood 96 and 101, respectively). In all panels, blue bars indicate the F1 father and red bars the mother(s). Only individuals left of the vertical dashed green line were retained for mapping (41, 42 and 29 offspring plus 2 backcross parents in Broods 122, 112 and 96, respectively). Due to low sequence volumes in all Brood 101 progeny, this brood was excluded from the mapping.

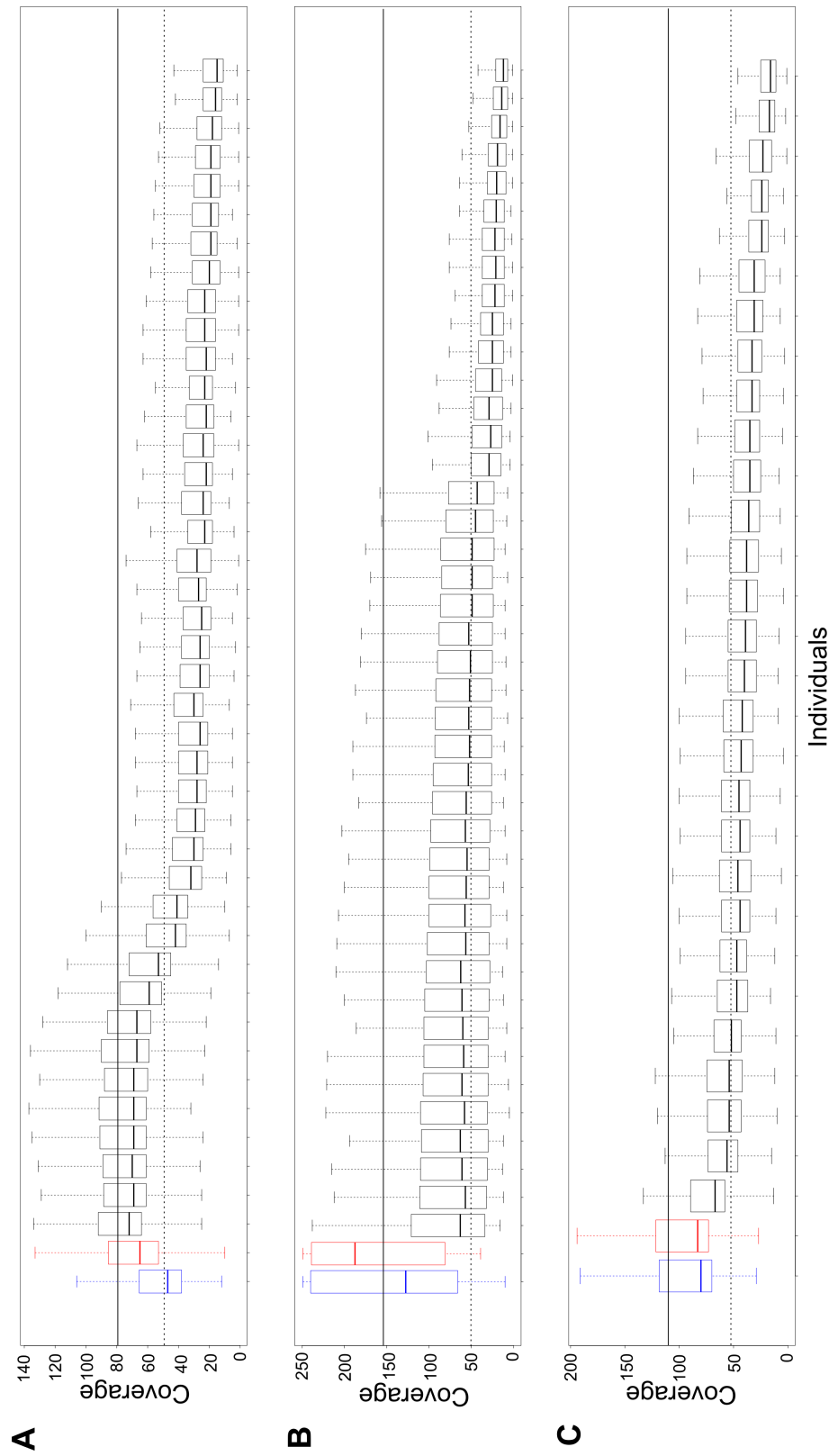


Figure 1.6 – **Distribution of sequencing depth** across the set of good quality polymorphic SNPs for the chosen subsets of individuals in each type of cross. Box and whisker plots showing the median value and first and third quartiles (boxes), and the upper and lower extreme values (whiskers) for (A) *H. hecale melicerta* × *H. h. zuleika* (Br122), (B) *H. hecale melicerta* × *H. h. clearei* (Br112) and (C) *H. ismenius bouletti* × *H. i. telchinia* (Br96). Mean coverage for parents (solid horizontal line) and offspring (dashed horizontal line) are indicated for each family. The father (blue box) and the mother (red box) are highlighted for each cross.

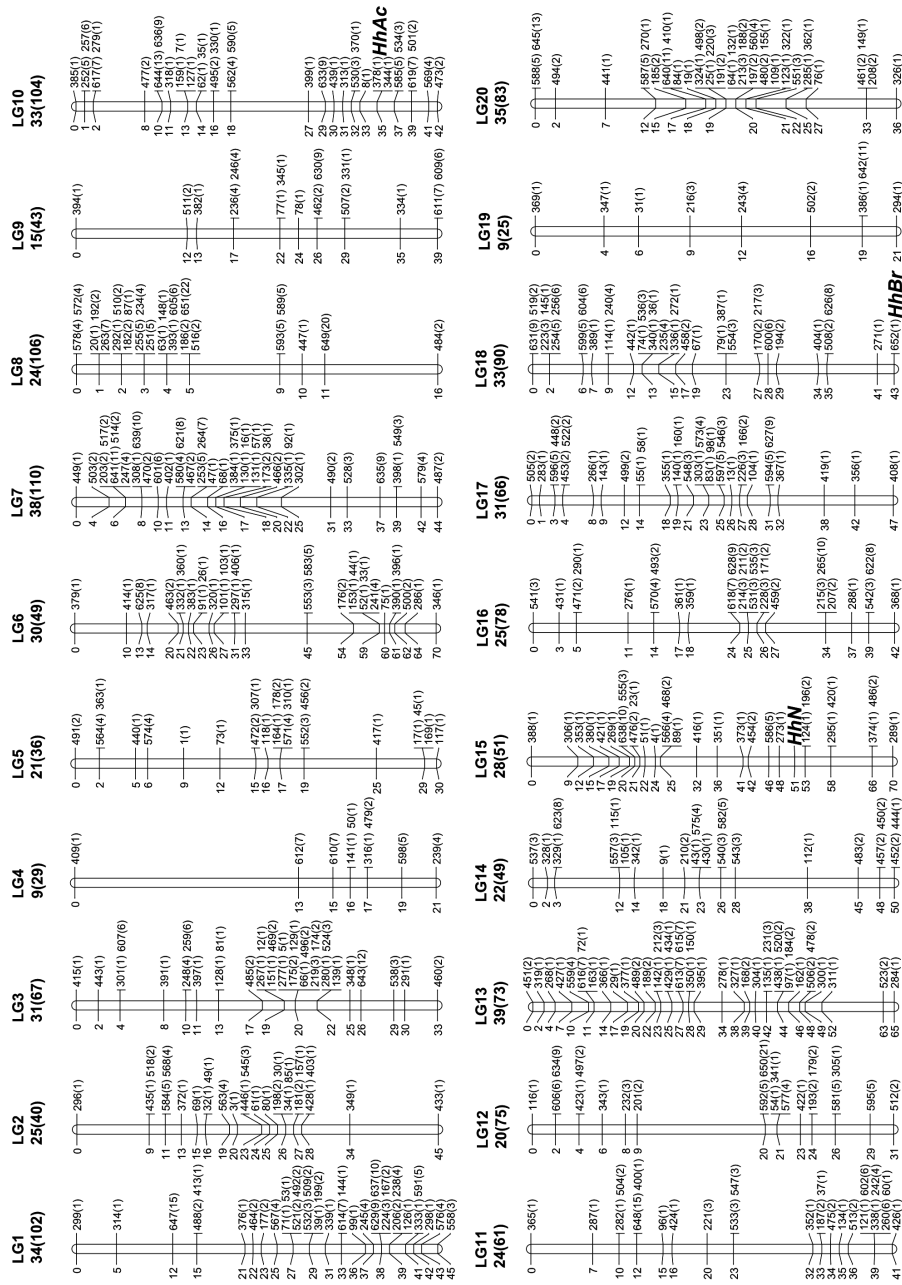


Figure 1.7 – Linkage map of the *Heliconius hecale* genome (20 autosomes) derived from the *H. hecale melicerta* × *H. hecale zuleika* cross. The position of each RAD-sequencing derived marker is shown in units of recombination frequency (centimorgans). Markers were given numerical identifiers which were in turn derived from collapsing multiple supporting SNPs with identical segregation patterns (number in brackets). The total number of markers per linkage group is shown on top of each chromosome map, with the total number of SNPs in brackets. The position on the genome of wing patterning loci *HhAc*, *HhN* and *HhBr* is shown.



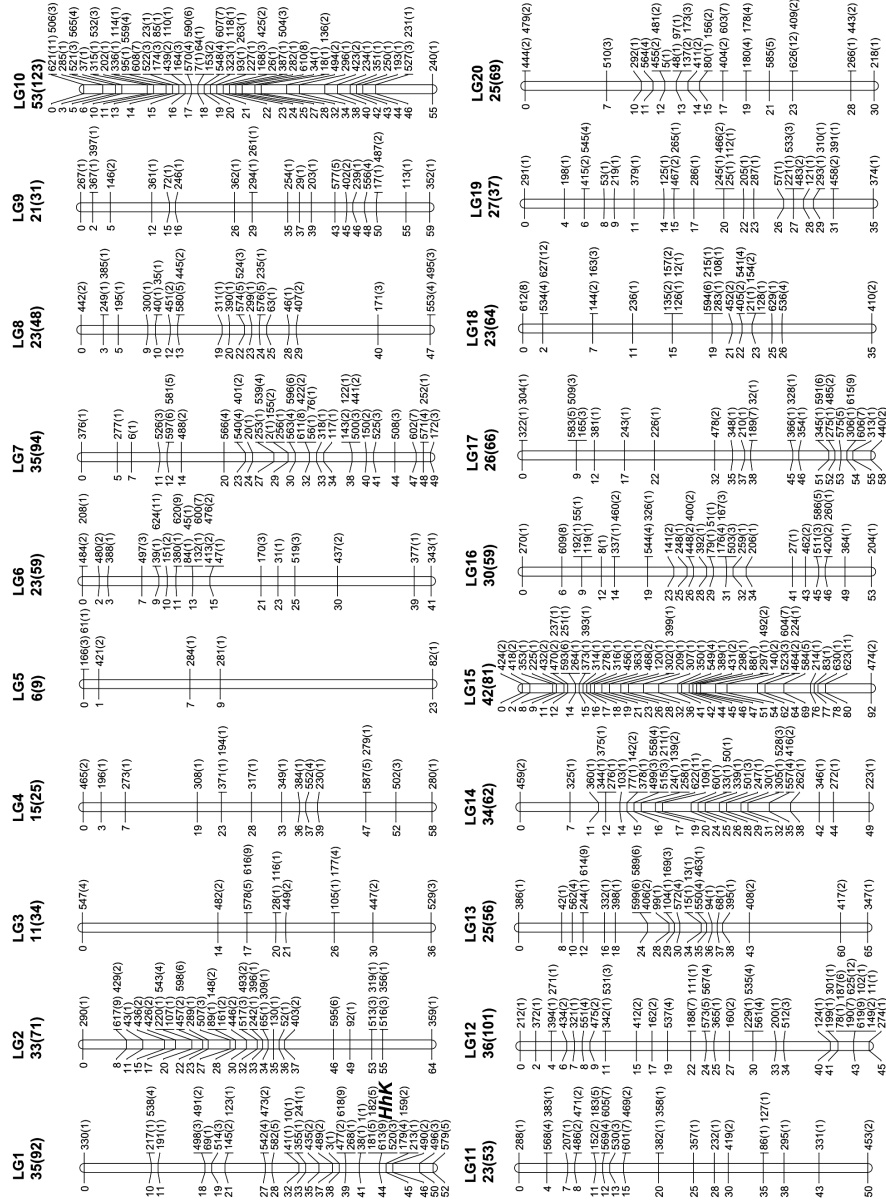


Figure 1.8 – Linkage map of the *Heliconius hecale* genome (20 autosomes) derived from the *H. hecale melicerta* × *H. hecale clearei* cross. The position of each RAD-sequencing derived marker is shown in units of recombination frequency (centimorgans). Markers were given numerical identifiers which were in turn derived from collapsing multiple supporting SNPs with identical segregation patterns (number in brackets). The total number of markers per linkage group is shown on top of each chromosome map, with the total number of SNPs in brackets. The position on LG1 of the wing patterning locus *HhK* is shown.

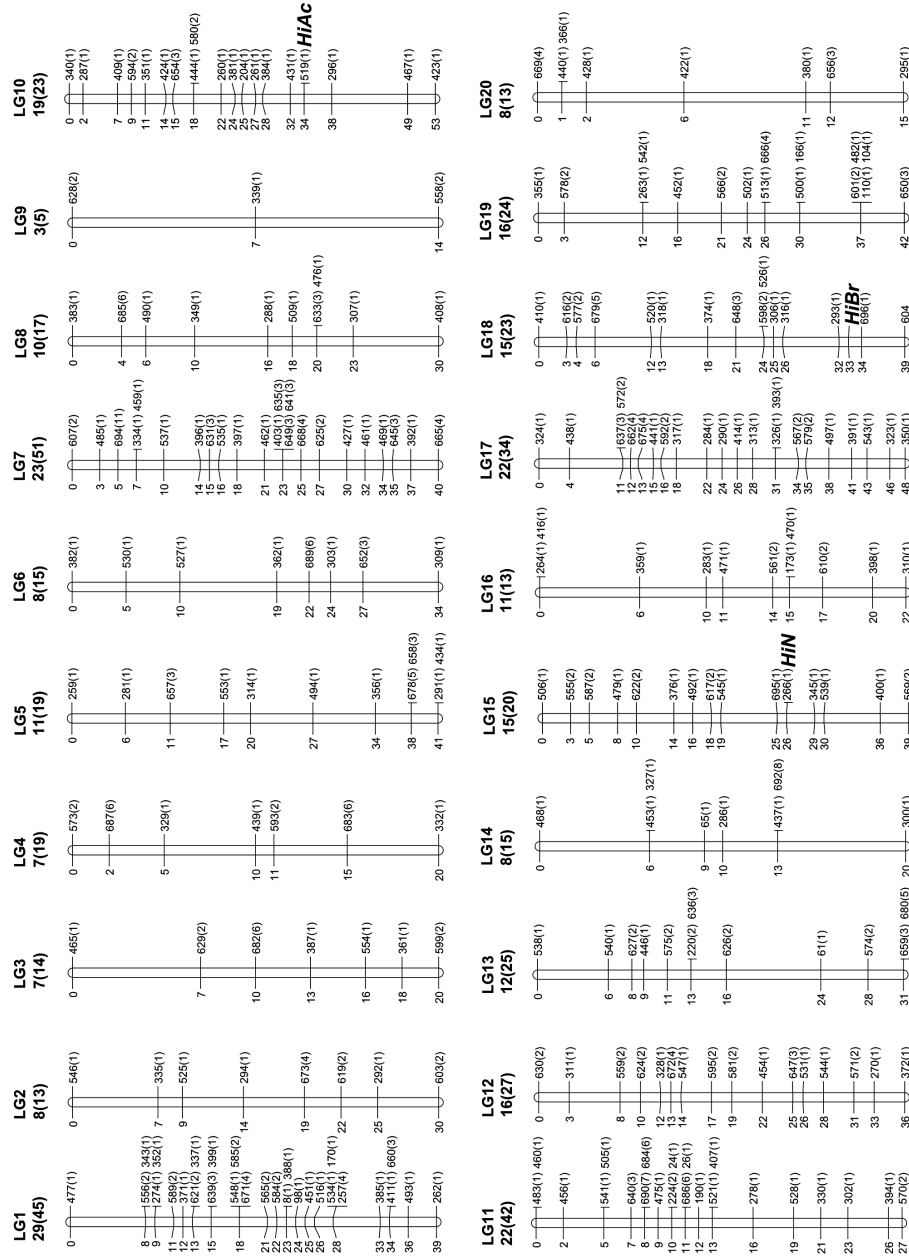


Figure 1.9 – Linkage map of the *Heliconius ismenius* genome (20 autosomes) derived from the *H. ismenius bouletti* × *H. ismenius telchinia* cross. The position of each RAD-sequencing derived marker is shown in units of recombination frequency (centimorgans). Markers were given numerical identifiers which were in turn derived from collapsing multiple supporting SNPs with identical segregation patterns (number in brackets). The total number of markers per linkage group is shown on top of each chromosome map, with the total number of SNPs in brackets. The position on the genome of wing patterning loci *HIAc*, *HIN* and *HiBr* is shown.

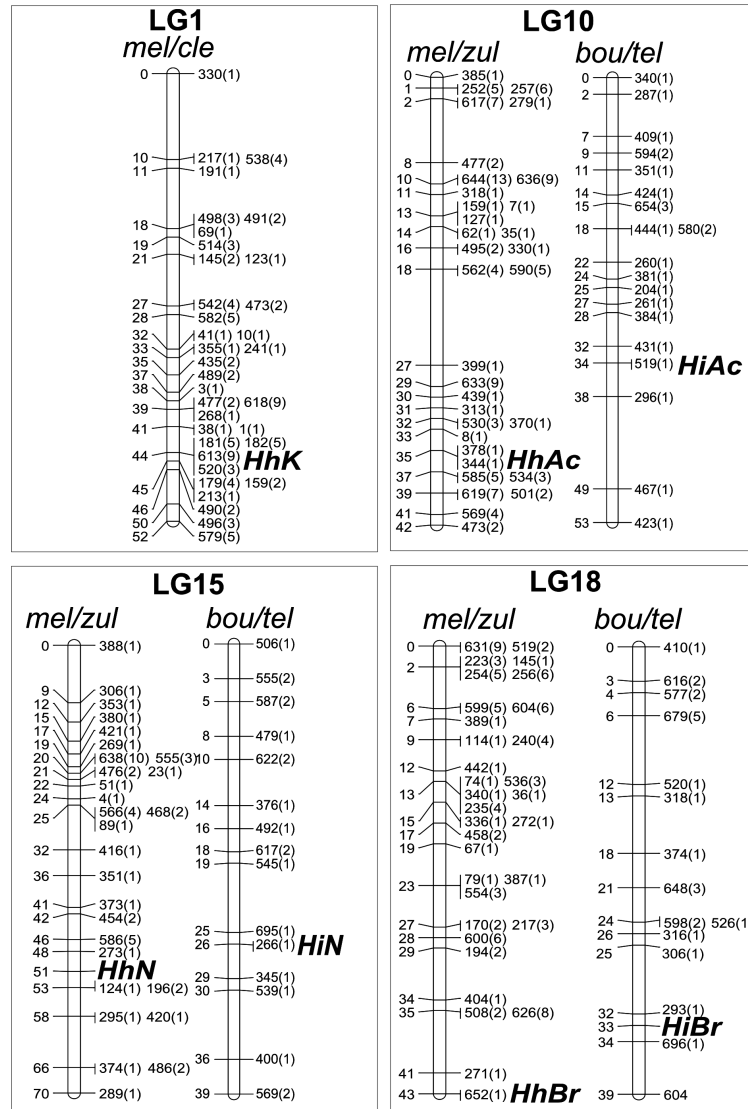


Figure 1.10 – Genomic position (in centimorgans) for each colour locus on its linkage group (LG) in *H. hecale melicerta* × *H. h. clearei* (*mel/cle*), *H. hecale melicerta* × *H. h. zuleika* (*mel/zul*), and *H. ismenius bouletti* × *H. i. telchinia* (*bou/tel*) crosses. Markers are named by their segregation patterns (seg.pat.) collapsing multiple supporting RAD markers (number in brackets). Locus *HhK* maps to LG1, together with the genomic scaffold that contains gene *wingless* (HE671174; several SNPs collapsed to seg.pats. numbers 181 and 182), but also with markers on scaffolds HE670375 and HE671357 (seg.pats. 613 + 520 and 179 + 159 + 490, respectively). Loci *HhAc* and *HiAc* map to LG10. They cluster together with markers on the *WntA* scaffold scfHE668478 (seg.pat. 378 + 344 and 519). Loci *HhN* and *HiN* map to LG15, close to the *Yb-Sb-N/Cr/P* superscaffold (scfHE667780; seg.pat. 124 + 196 and 695). Loci *HhBr* and *HiBr* map to LG18. They cluster with markers on the scaffold containing the gene *optix*, scfHE670865 (seg.pat. 652 and 293 + 696).

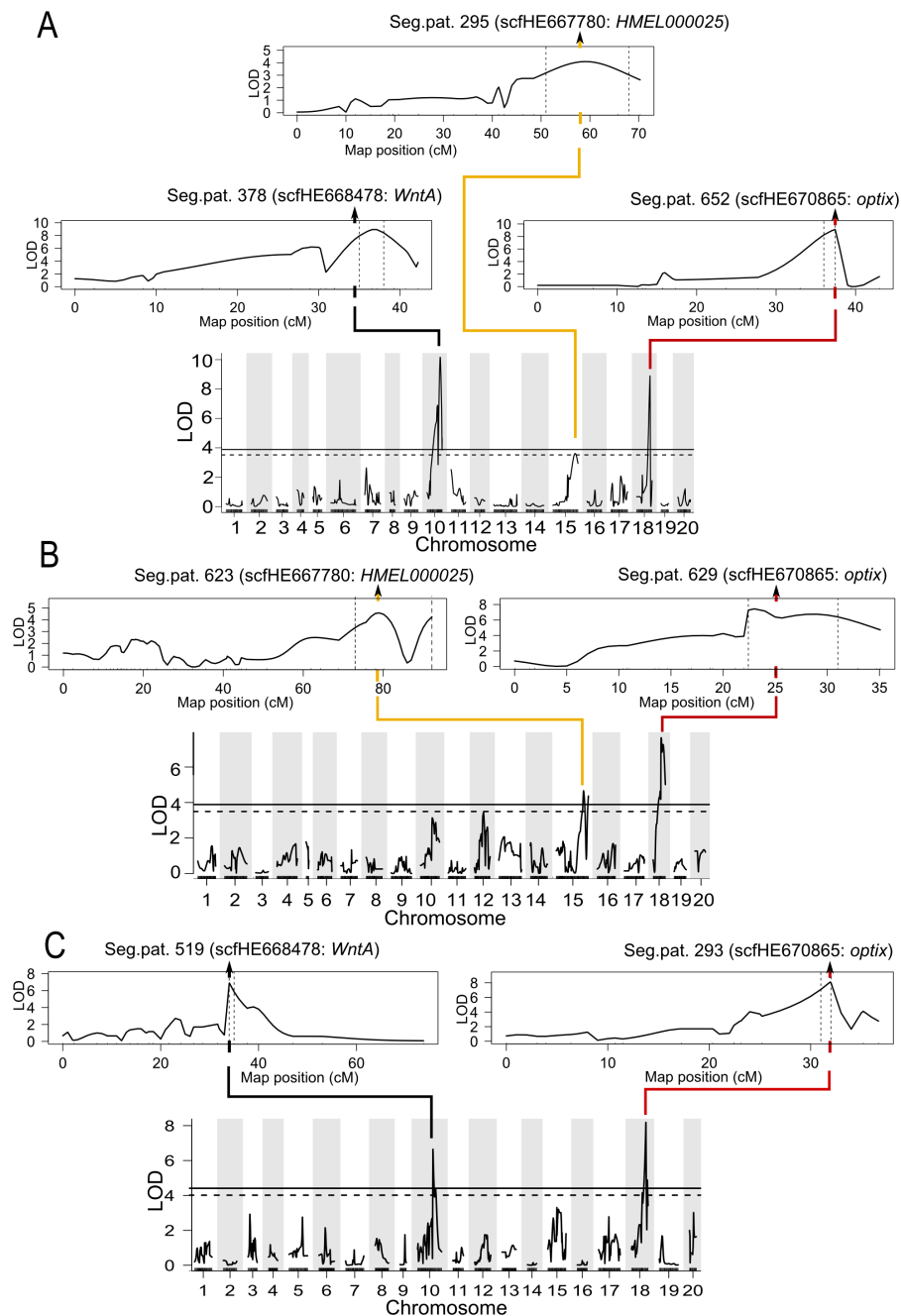
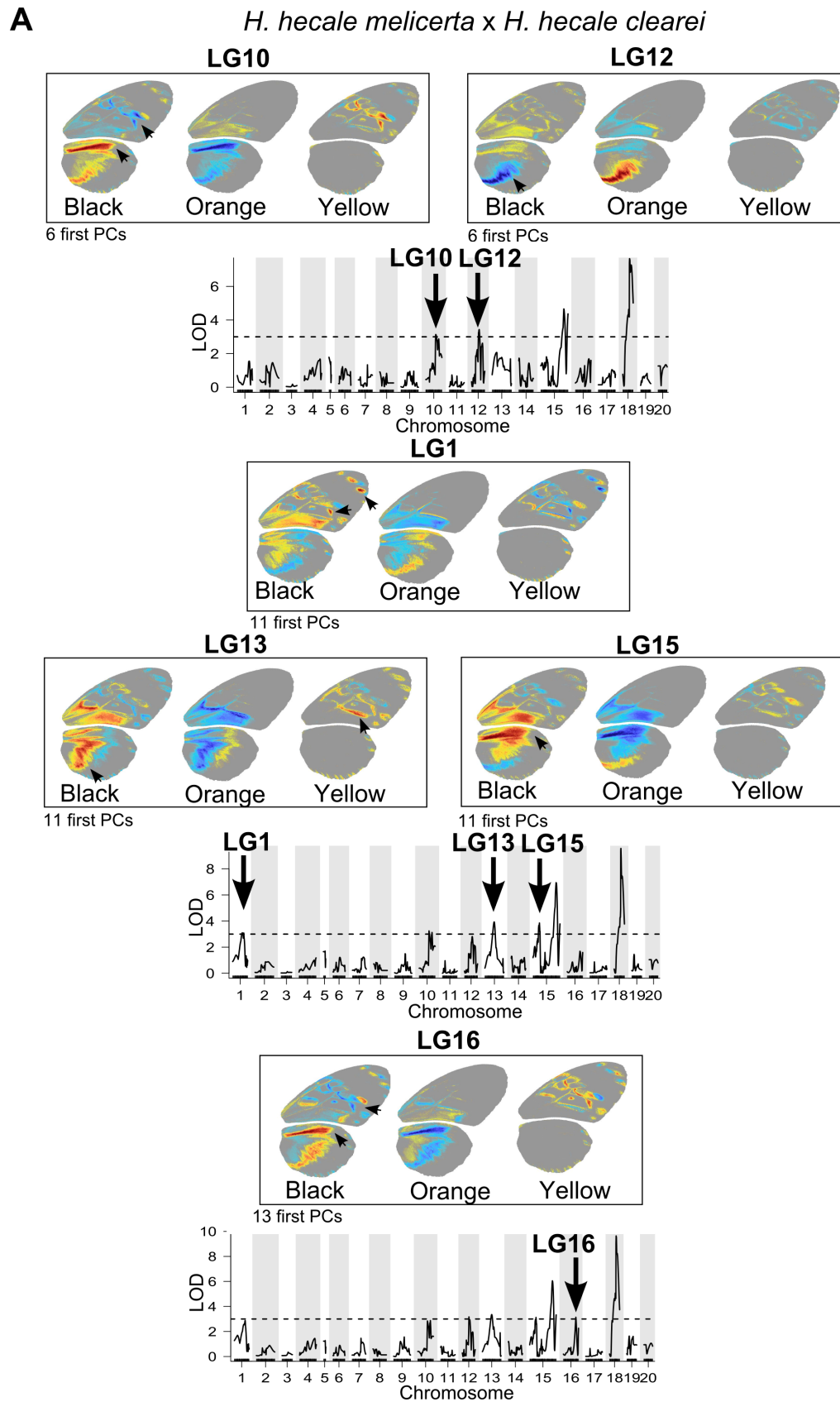


Figure 1.11 – Detail of QTL mapping around the highest peaks of LOD score in each cross. (A) *H. hecale melicerta* × *H. h. zuleika*, (B) *H. hecale melicerta* × *H. h. clearei* and (C) *H. ismenius bouletti* × *H. i. telchinia*. Credible intervals around the highest peak (see also Figure 1.3) are delimited by vertical dashed lines. *H. melpomene* genome scaffolds (scf) containing candidate colour loci (*WntA*, *HMELO000025* and *optix*) are indicated for each major QTL, with the position of the corresponding RAD marker (seg.pat.) shown by arrows. Credible intervals were determined using the *bayesint* function of *R/qtl* with a coverage probability of 0.95.



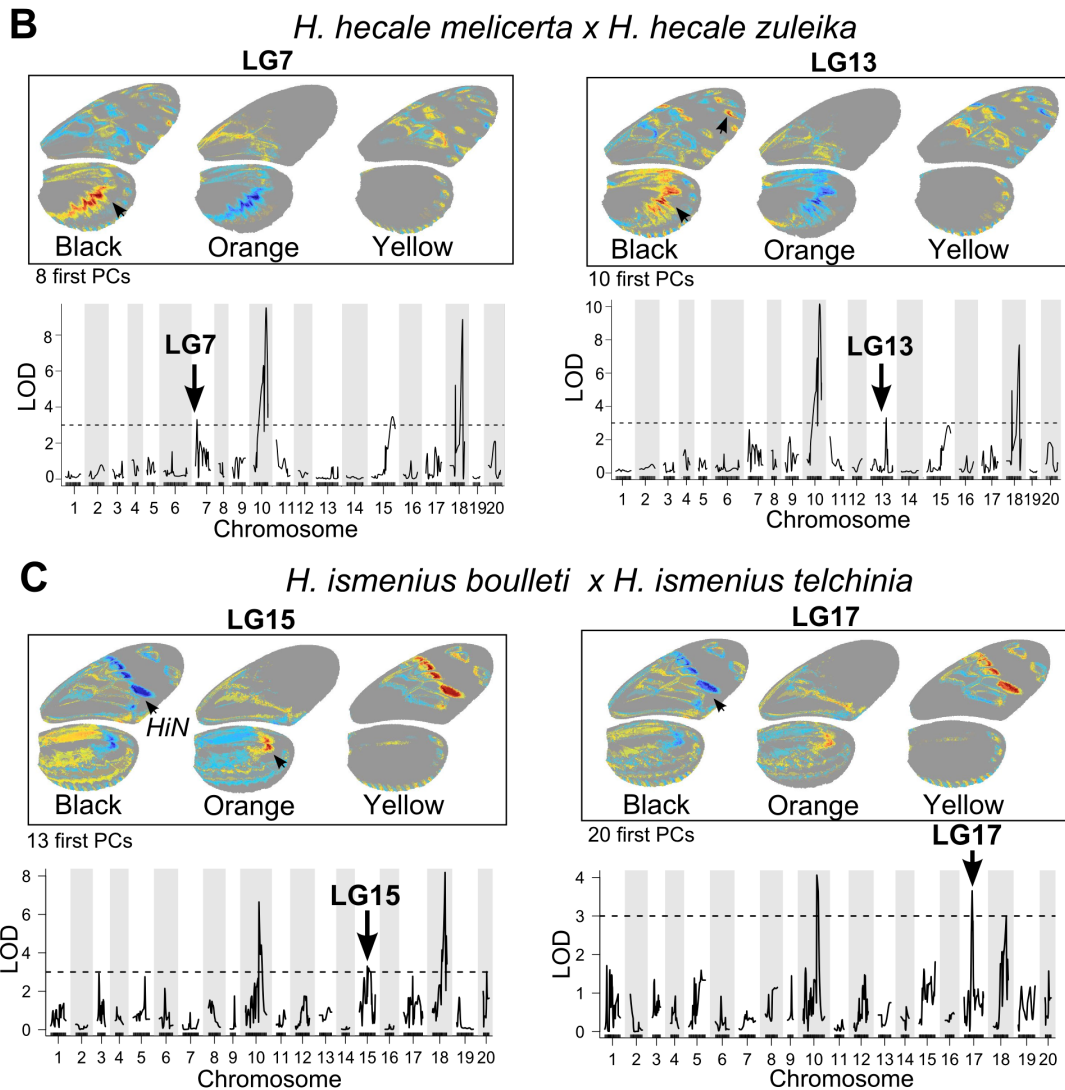


Figure 1.12 – Genomic position and phenotypic effect of suggestive QTLs identified in the *H. hecale* and *H. ismenius* crosses. (A) *H. hecale melicerta* × *H. h. clearei*, (B) *H. hecale melicerta* × *H. h. zuleika* and (C) *H. ismenius bouleti* × *H. i. telchinia*. Each panel illustrates the phenotypic effect of a specific QTL (upper pane) and the genome-wide association scans (lower pane) with the targeted QTL and linkage group indicated by black arrows. The subset of principal components used in each case is stated. Coloured wing diagrams show the spatial distribution of individual QTL effects on pattern variation extracted from multivariate wing pattern analysis. Phenotypic variation is broken down into heatmaps for each of the three main colours (black, orange and yellow), representing, for every wing position, the strength of association between colour presence and allelic transition at the QTL (from blue to red). For analytical simplicity purposes, both white and yellow elements in the *H. hecale melicerta* × *H. h. clearei* cross were considered as yellow elements. Genomic plots show the genome-wide association (LOD) scan between wing pattern variation and markers on the 20 autosomes, with the conventional association threshold of LOD = 3 shown with an horizontal dashed line. See Table 1.11 for information about the approximate genomic location of each QTL.

## Chapter 2

# The hybrid zone between two races of *Heliconius hecale*: exploring the role of wing pattern coloration in early stages of diversification

Bárbara Huber (1,2,3), Annabel Whibley (1), Benjamin Rice (2), Francisco Delgado (4), Luis Murillo (5), Adriana Tapia (2), Thierry Wirth (1,3), W. Owen McMillan (2), Violaine Llaurens (1), Mathieu Joron (1,2,6)

1. Institut de Systématique, Evolution et Biodiversité, Muséum National d'Histoire Naturelle, Paris, France
2. Smithsonian Tropical Research Institute, Balboa, Panama
3. Laboratoire Biologie Intégrative des Populations, Ecole Pratique des Hautes Etudes (EPHE), Paris, France
4. Universidad de Panamá, Santiago de Veraguas, Panamá
5. Escuela de Agronomía Universidad de Costa Rica, San José, Costa Rica
6. Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, Montpellier, France

## 2.1 Introduction

Hybrid zones offer the opportunity to study the speciation process and explore the evolution of reproductive isolation barriers between divergent but still hybridising taxa. How can reproductive barriers arise in an essentially continuous population and how do they maintain boundaries between diverged lineages? How do different pre and post-zygotic isolation factors interact to trigger diversification? And more generally, what are the mechanisms allowing certain adaptive traits to lead the radiation of whole clades? These questions have extensively been investigated but remain largely unanswered ([Coyne and Orr, 2004]). Although evolutionary mechanisms involved in maintaining and enhancing diversification in the face of gene flow might differ among different taxa, natural and sexual selection seem to play a central role in these processes ([Coyne and Orr, 1998]; [Schemske, 2000]; [Schluter, 2000]; [Malausa et al., 2005]; [Via, 2009]).

Hybrid zones exemplify distinct stages of reproductive isolation between hybridising populations ([Barton and Hewitt, 1989]; [Szymura and Barton, 1991]; [Harrison, 1993]; [Veen et al., 2001]); their study has contributed to describe speciation as a gradual and accumulative process ([Jiggins and Mallet, 2000]). Documenting different stages of the speciation continuum within a single clade allows exploring how different ecological and genetic factors cumulate and contribute to reproductive isolation between populations ([Hendry et al., 2009]; [Powell et al., 2013]). Here we focus on butterflies from the neo-tropical genus *Heliconius*, which exemplify distinct stages of this continuum and which have undergone a recent adaptive radiation of aposematic wing colour patterns acting as warning signals of their toxicity ([Brown, 1981]; [Mallet, 1993]). Several toxic species from inside and outside the genus share locally the same wing colour pattern (i.e. Müllerian mimicry), which decreases predation risk and therefore generates strong local selective pressures on wing coloration. Most *Heliconius* species display a geographic mosaic of parapatric distinctly-coloured races mimicking local communities of defended species. Hybrid zones among these geographic races have variable size depending upon the strength of disruptive selective pressures and the level of gene flow ([Mallet, 1986]; [Jiggins et al., 1997]; [Blum, 2008]; [Arias et al., 2012]; see Table 2.4). Interestingly, wing patterns are not only subjected to strong disruptive selective forces due to counter-selection on hybrid individuals displaying non-mimetic wing colour patterns (acting as post-zygotic isolation barriers; [Merrill et al., 2012]), but are also essential cues for mate preferences (constituting pre-zygotic barriers to gene flow; [Jiggins et al., 2001]; [Merrill et al., 2011a]). Wing colour patterns might thus act as “magic traits” (c.f. [Servedio et al., 2011]) facilitating ecological speciation ([Gavrilets, 2004]; [Jiggins et al., 2005a], [Jiggins, 2008]).

Recent advances in genomics have facilitated the access to whole genome data allowing investigating various models of speciation with gene flow in different organisms ([Lawniczak et al., 2010]; [Feder et al., 2012]; [Ellegren et al., 2012]), including *Heliconius* butterflies whose reference genome (*H. melpomene*) was sequenced in 2011 ([*Heliconius* Genome Consortium, 2012]). Genomic islands of divergence have been observed around wing colour loci when comparing parapatric races of *H. melpomene*, and have been suggested to be the result of strong divergent selection due to local mimicry ([Nadeau et al., 2012]; [Martin et al., 2013]; [Kronforst et al., 2013]). This high divergence in restricted fractions of the genome contrasts with high gene flow observed among races in the rest of the genome, suggesting that geographical races of *H. melpomene* are on an early stage of the speciation process. However, although mimetic wing colour patterns are an important diversification factor in *Heliconius* ([Mallet et al., 1998a]), it is unclear whether diversifying selection on this trait would be sufficient to generate speciation. Additional ecological and/or genetic factors might be required for speciation process to be



completed. In consistence with this idea, speciation has been shown to have proceeded as a multifactorial process between pairs of rarely-hybridising closely-related species like *H. cydno*/*H. pachinus*, *H. erato*/*H. himera* and *H. melpomene*/*H. cydno*, implicating divergence at host plant specificity and microhabitat preference ([Smiley, 1978]; [Mallet and Gilbert, 1995]; [Jiggins et al., 1996]). The exploration of interracial hybrid zones in *Heliconius* allows investigating how reproductive barriers cumulate, how disruptive natural and sexual selection forces on wing coloration act together, and how these selective pressures counterbalance gene flow during incipient stages of diversification.

Most research on *Heliconius* hybrid zones focuses on species in the two distantly-related clades *melpomene* and *erato* (see Table 2.4). The selective mimetic environment determining local adaptation of the colour pattern lineages in these clades is often simple and homogeneous, in part because the mimetic circles are usually not very speciose. For instance, species *H. melpomene* and *H. erato* are perfect and exclusive mimics of each other across a large portion of their distribution area. Research on this relatively simple system has allowed drawing straightforward conclusions about the importance of mimicry in the diversification of the genus ([Mallet, 2009]). Other *Heliconius* species like those in the “silvaniform” clade have been comparatively much less explored and, to our knowledge, no hybrid zone between silvaniform lineages has been investigated so far. Silvaniform species belong to “tiger-patterned” mimetic communities, which are important components of the butterfly communities in the neo-tropics and display high complexity of wing patterns, composed of a mosaic of black, orange and yellow/white elements. In this case, although mimicry rings are usually clearly defined, with several species (mainly in the Danaine subfamily) showing fine-detail similarities, mimicry is also often blurred by imperfect mimics. Such a context allows exploring to which extent the complexity of mimetic patterns and the composition of the mimetic communities determines the characteristics and the dynamic of hybrid zones in the genus *Heliconius*.

Here, we focus on the Panamanian hybrid zone between two distinctly-coloured races of the silvaniform *Heliconius* species *H. hecale*, namely *H. hecale zuleika* and *H. hecale melicerta*. We take advantage of the known genetic architecture or wing patterning in this species, which was investigated by performing crosses between these races ([Huber et al., 2015]). We implement a combination of population genetics and genomics, cline theory and mating behaviour experiments with the aim of characterising the divergence stage between the *H. hecale* races and placing them into a broad comparative context in the speciation continuum within the genus *Heliconius*. By having an integrative view on this hybrid zone, we explore how natural and sexual selection contribute to maintain phenotypic divergence in presence of gene flow in early stages of divergence. We first use population genetics and population genomics approaches to estimate the extent of genome-wide gene flow between these two Central American *H. hecale* parapatric races. We investigate whether the genomic regions associated with the colour loci are more divergent than the genomic background, a fact that would result from disruptive selective pressures for adaptation to local mimetic environments. Next, we explore the strength of natural selection acting on these loci in the field, taking profit of the known allelic segregation at each colour locus. Namely, three major colour loci, which vary in the dominance relationship of their alleles, determine wing pattern differences between *H. h. zuleika* and *H. h. melicerta* ([Huber et al., 2015]). We build phenotypic clines at those major loci along a sampled transect and compare them to evaluate what does the steepness of the colour pattern clines tell about the disruptive selection maintaining divergence between the two races. We further ask whether the width and position of the clines differ between loci. Finally, we test for male-based mate choice to determine whether there are assortative mating preferences that could be strengthening divergence between the *H. hecale* races.

## 2.2 Materials and Methods

### 2.2.1 Butterfly sampling along the Panamanian transect

Two sampling strategies were used (Figure 2.1). On the one hand, we collected a total of 333 adult butterflies of both races along a west-to-east transect spanning  $\sim 800$  km from San José (Costa Rica) to Darién (Panama) to analyse the transition zone between *Heliconius hecale melicerta* and *H. hecale zuleika*. The transect was defined all along the centre of the Panamanian Isthmus. Distances among populations on this linear transect were computed based on the location of the red dots on the map (see Figure 2.1). The average distance between populations was 133.46 km ( $\pm 83.59$  km), with the lowest distance (34.7 km) existing between populations BT (Bocas del Toro) and Ch (Chiriquí), which are actually located on distinct slopes of the mountain chain Cordillera Central. The westernmost sampled population contains almost exclusively phenotypically pure *H. hecale zuleika* and the eastern most analysed population is virtually pure for *H. hecale melicerta*. Five intermediate sites were sampled within these extremes throughout Panama. Butterflies caught in distinct sites within a diameter of 50 km were considered as belonging to the same population. Between 25 and 67 specimens were collected in each population. Most of the butterflies were collected during the years 2011 and 2012, and a few were caught in 2009 and 2013. Wings were removed and stored in envelopes for further phenotypic description, and bodies were placed in a DMSO/EDTA/High-NaCl solution for proper DNA preservation. On the other hand, for mate choice experiments *zuleika* and *melicerta* butterflies were collected from populations Bocas del Toro (BT) and Colón (Cl), respectively, and then bred at the Smithsonian Tropical Research Institute (STRI) (see [Huber et al., 2015] for more details).

### 2.2.2 Microsatellite genotyping and population genetic analysis

We examined allelic variation at eight microsatellite loci, one of them taken from Mavárez and González ([Mavárez and González, 2006]) and the rest developed for *Heliconius numata* ([Le Poul et al., 2014]), a species belonging to the silvaniform clade, like *H. hecale*. We analysed a set of 24 *H. h. zuleika* specimens from Bocas del Toro (BT) and 32 *H. h. melicerta* samples from Darién (D). Primer sequences are shown in Table 2.2. Microsatellite markers were amplified using M13-tailed specific primers labelled at the 5'-end with one out of four fluorescent dyes (YAKYE, AT565, AT550 or FAM) to allow multiplexing (Eurofins Genomics). The sequence of the used M13-tail was: 5' CACGACGTTGTAAAACGAC 3'. PCR was carried out in 10  $\mu$ l final volume reactions. They contained around 4 ng of genomic DNA, 1 $\times$  reaction buffer, 0.2 mM of each dNTP, 0.1  $\mu$ M of the forward primer, 1  $\mu$ M of the reverse primer, 1  $\mu$ M of the M13 labeled primer and 1.25 units of DreamTaq Polymerase. We performed standard PCR reactions, starting with 2 min of initial denaturation at 94 degreeCelsius, following with 44 cycles of 30 seconds at 94 degreeCelsius, 30 sec at 52 degreeCelsius and 45 sec at 72 degreeCelsius, and finishing with a final extension for 5 minutes at 72 degreeCelsius. Markers were scored on an ABI3130 sequencer (Applied Biosystems). The size of the alleles was determined in reference to the Genescan-500 LIZ dye size standard (Life Technology) and using the software *GeneMarker v2.4.0*. Statistical analysis of microsatellite data was performed using the software *GENETIX v4.05.2* ([Belkhir et al., 1996]). We surveyed within-race genetic diversity in terms of expected heterozygosity (*He*) and observed heterozygosity (*Ho*), and estimated the *Fst* statistic between populations ([Weir and Cockerham, 1984]). We also ran a factorial correspondence analysis (FCA), which projects the individuals on a factor space based on the similarity of their allelic states.

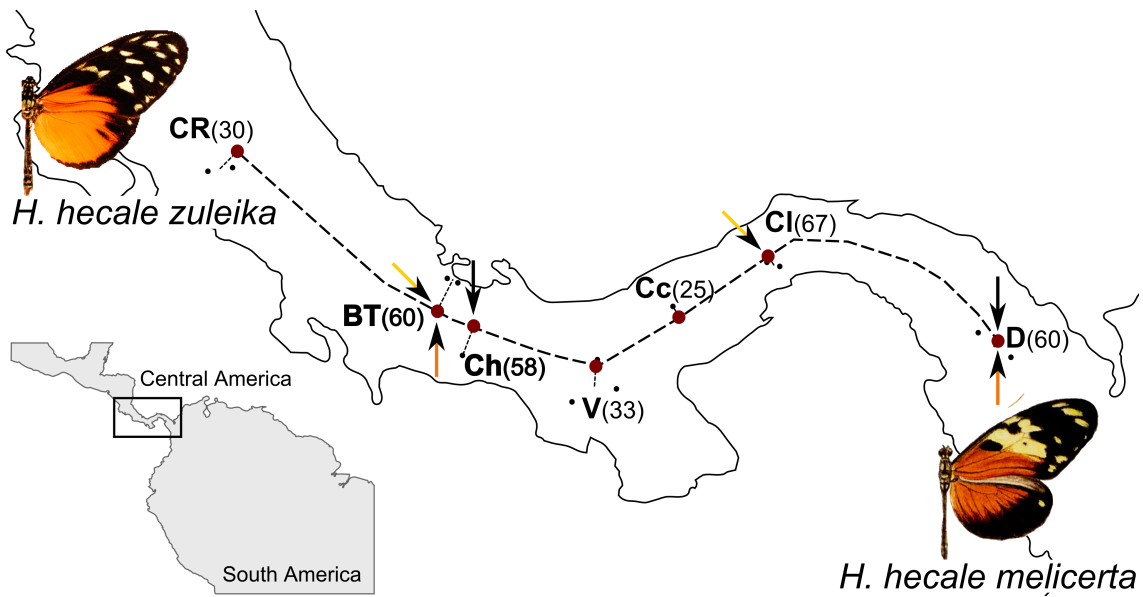


Figure 2.1 – Sampling sites (little black dots) and studied populations (big red dots) on the analysed transect (dashed line) along the transition zone between *H. hecale zuleika* and *H. h. melicerta*. CR: Costa Rica; BT: Bocas del Toro; Ch: Chiriquí; V: Veraguas; Cc: Coclé; Cl: Colón; D: Darién. The number of specimens collected in each population is shown in brackets. Orange and black arrows point to the collecting sites for specimens used for population genetics and genomics analyses, respectively. Yellow arrows indicate sampling sites where living butterflies were collected for mate choice assays. The location of Panama and the Southeast of Costa Rica are delimited in a black rectangle on the map of America in gray.

### 2.2.3 Exploration of intraspecific genome-wide differentiation on re-sequenced whole genomes

We compared full genome sequences of ten wild-caught *H. hecale* samples. We chose five phenotypically pure specimens of each colour pattern race *H. hecale zuleika* and *H. hecale melicerta*, which were collected in populations Chiriquí (Ch) and Darién (D), respectively (Figure 2.1). We extracted genomic DNA from  $\sim 1/3$  of the thorax of the preserved bodies using the DNeasy Blood and Tissue Kit (Qiagen). We prepared custom Illumina sequencing libraries for each sample separately and sequenced them to an average depth of  $24\times$  using an Illumina Hi-Seq 2000 ( $2 \times 100$  paired-end sequencing). We mapped the quality-filtered reads for each sample back to the reference genome scaffolds of *H. melpomene version 1.1* ([Heliconius Genome Consortium, 2012]) using *Stampy v1.0.17* ([Lunter and Goodson, 2011]) with default parameters except for the substitution rate set to 0.05. After removing PCR-duplicate reads using *Picard Tools v1.107* (<http://picard.sourceforge.net>) and locally realigning around indels using the *Genome Analysis Tool Kit (GATK) v2.8-1* ([DePristo et al., 2011]), we scored variable sites also using *GATK* ([DePristo et al., 2011]), setting the expected heterozygosity to 0.015. We obtained a subset of high quality genotypes for biallelic sites having  $> 5\times$  coverage,  $> 30$  genotype quality (GQ) and  $> 40$  mapping quality (MQ), using *GATK* ([DePristo et al., 2011]), and used it for subsequent analyses. This final dataset consisted of 3,206,100 SNPs. We evaluated genome-wide genetic divergence between the *H. hecale* parapatric races by estimating the *Fst* statistic (the fixation index, a measure of population differentiation; [Weir and Cockerham, 1984]). *Fst* values were calculated with an average/sliding windows approach across the genome using the VCFtools software package ([Danecek et al., 2011]). We

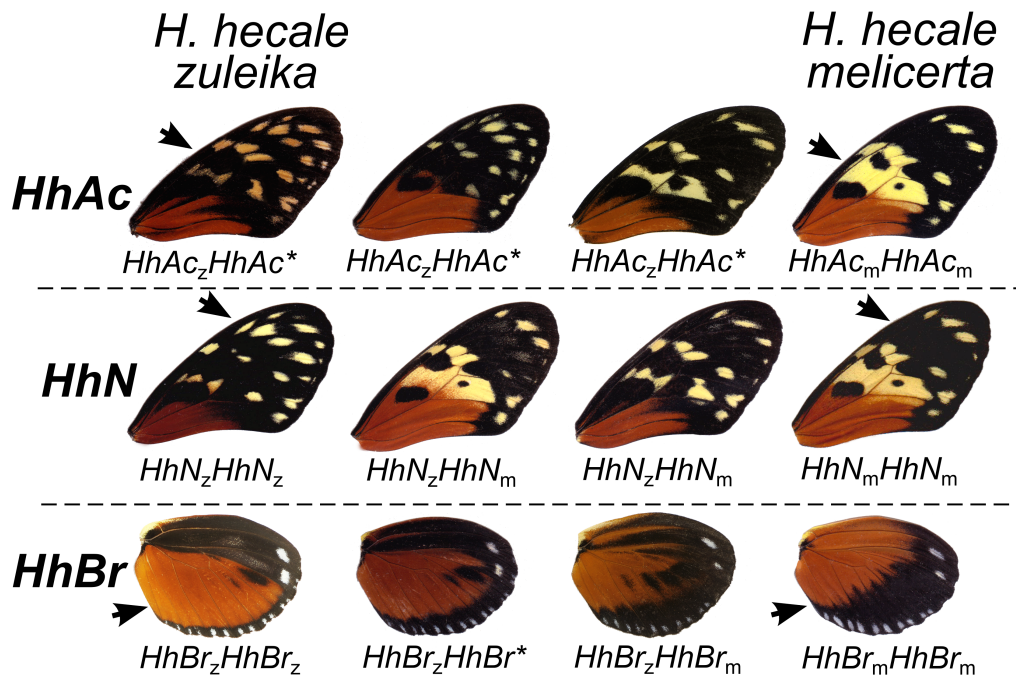


Figure 2.2 – Sample of pure and hybrid phenotypes found for the three colour genes across the Panamanian hybrid zone. Black arrows highlight the phenotypic effect of each locus, to be contrasted between the phenotypically pure *H. hecale zuleika* (column on the left) and *H. hecale melicerta* (column on the right). Under each wing picture, the genotypic state for a given locus is written, with asterisks meaning “unknown allele”.

also zoomed into the scaffolds containing the wing colour loci to explore whether or not islands of divergence existed around the candidate colour genes. Moreover, we estimated genome-wide diversity ( $\pi$  or  $P_i$ ) for each race using VCFtools ([Danecek et al., 2011]).

#### 2.2.4 Wing colour pattern analysis

A previous study has shown that the marked differences in wing colour pattern between the Panamanian races *H. hecale zuleika* and *H. h. melicerta* are largely determined by three loci of Mendelian inheritance (*HhAc*, *HhN* and *HhBr*), which are homologous to those identified in other *Heliconius* species ([Huber et al., 2015]). Therefore, we were able to score wing colour genotypes for the whole sample along the studied hybrid zone, based on visual inspection of phenotypic switches at each of these three loci (see Figure 2.2 for example phenotypes). For locus *HhAc*, heterozygotes could not be distinguished from homozygotes because of strong dominance of the *zuleika* allele; *zuleika*-like phenotypes were thus coded as one unknown allele and one *zuleika* allele. For locus *HhBr*, the *zuleika* allele is also strongly dominant over the *melicerta* allele. However, the heterozygous state was usually differentiable from the dominant recessive state. For locus *HhN*, heterozygotes could be easily distinguished from homozygotes. Nevertheless, whenever doubts existed to determine one or both allelic states of a locus for a given specimen, we coded the allele as unknown. By using such a conservative strategy we are confident about the genotypic assignment we have achieved, although we are aware of the underestimation in the number of admixed individuals in a given population.

### 2.2.5 Analysis of phenotypic cline shape

We estimated changes in the allele frequency for the three major wing colour pattern loci along the Panamanian hybrid zone (see Chapter 1). First, we calculated allele frequency in each population and integrated it with the distance of the populations along the drawn linear transect (see Figure 2.1). We then mathematically described the geographic structure of the studied hybrid zone by modelling the shape of the cline for each wing colour trait separately. We generated maximum-likelihood curves for each phenotype by using the algorithms proposed by [Szymura and Barton, 1986] [Szymura and Barton, 1991] and implemented in the *R* package *HZAR* ([Derryberry et al., 2014]). We fit a sigmoidal curve at the centre of the cline using the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm ([Metropolis et al., 1953]; [Hastings, 1970]). This method allowed searching for the maximum-likelihood estimates of clines parameters, which could then be compared among clines. Following [Mullen and Hoekstra, 2008] we implemented a model that allowed four parameters to vary for each character: cline centre ( $c$ ), cline width ( $w$ ), minimum and maximum frequencies at the tail ends of the cline ( $p_{\min}$  and  $p_{\max}$ , respectively).

### 2.2.6 Phenotypic hybrid index

To evaluate the distribution of genotypic classes (i.e. phenotypically pure versus admixed individuals) in any give population across the transition zone, we derived a “phenotypic hybrid index” similar to that used by [Jiggins et al., 1997] and [Arias et al., 2012]. We used the allele scoring for the three colour loci in each sampled individual. The mean value over all allelic states yielded a score between 0 and 1, where 0 corresponds to an entirely *melicerta*-like butterfly and 1 to a phenotypically pure *zuleika* phenotype. In contrast, specimens with a score between these extremes are hybrid for at least one of the three loci. For instance, a three locus genotype [*HhAczHhAc*\* *HhNmHhNm* *HhBrzHhBrm*] would be equivalent to  $[1 + \text{“unknown”} + 0 + 0 + 1 + 0]$  and would give that individual a phenotypic hybrid index of  $2/5 = 0.4$ .

### 2.2.7 Mate choice experiments between the two races

To understand the mate preferences operating on the phenotypes, we tested whether *H. hecale* show mating preferences toward wing colour pattern. First we tested for male choice only by allowing mature males of each race ( $N = 31$  of *H. h. melicerta* and  $N = 32$  of *H. h. zuleika*) to choose between female wing models of both races. Males, which were reared to sexual maturity in the insectaries or collected in the field, were tested on different days to register the total number of approaches and courtship events in three independent trials (technical replicates). Models consisted of real non washed female wings and were connected to a plastic wire to be moved in a fluttering motion inside a  $2 \times 2 \times 2$  m cage (used device: Butterfly 2000, Adriana Tapia’s design). Courting events were registered during 5 minutes, switching the position of the two models in the middle of this period to avoid biases due to their position in the cage or to male learning. A likelihood approach ([Edwards, 1972]; [Jiggins et al., 2001]; [Kronforst et al., 2006c]; [Merrill et al., 2011b]) was used to estimate the relative probability of males courting the *zuleika* wings rather than the *melicerta* wings. The total number of events directed towards *zuleika* models by male  $x$  ( $n_{x,zul}$ ) and to *melicerta* by the same male ( $n_{x,mel}$ ), in addition to the probability of courting *melicerta* wings by male  $y$  ( $P_{y,mel}$ ), were included in the likelihood function:

$$\ln(L) = \sum n_{x,zul} \ln(P_{y,mel}) + n_{x,mel} \ln(1 - P_{y,mel})$$

The solver option in Excel (Microsoft) was used to estimate the probability of male courtship by numerically searching for values of  $P$  that maximized  $\ln(L)$ . The support limits (asymptotically equivalent to 95% confidence intervals) were assessed by looking for values that decreased  $\ln(L)$  by two units ([Edwards, 1972]). To compare the courtship events of the males of both species, we produced a model where the relative probabilities were established equally ( $P_{mel} = P_{zul}$ ) and a model where the probabilities for each species were calculated separately ( $P_{mel} \neq P_{zul}$ ). These were compared by performing a likelihood ratio test assuming that the test statistic  $G = 2\Delta \ln(L)$  is  $\chi^2$ -distributed and using one degree of freedom.

To jointly test male and female preferences, we performed 25 “tetrad” experiments ([Jiggins et al., 2001]), exploring the extent of intraspecific assortative mating. A single recently emerged virgin female and a mature male (more than 5 days old) of each race were placed in a  $2 \times 2 \times 2$  m cage and the total number of events in each main step of courtship was registered during 30 minutes or until the first couple of butterflies mated. A successful courtship sequence in *H. hecale* consists of four main steps, which are described in detail in Chapter 3: *localization* (the male approaches the female), *hovering* (the male moves his wings fast on top of the female, close to her head), *attempting* (the male bends his abdomen towards the female’s) and *success* (mating). However, the female can refuse the male at any point (more commonly around the two last steps of courtship) by adopting a *female rejection* posture, where she opens her wings flat against the substrate and raises up her abdomen while displaying her abdominal glands. In cases in which both couples mated simultaneously half a mating was recorded for each of them.

## 2.3 Results

### 2.3.1 Extreme *H. hecale* populations lack differentiation across of the transition zone

Genotype data from eight relatively variable microsatellites (9 alleles per marker in average; see Table 2.2) showed no genetic differentiation when comparing the *zuleika*-like population from Bocas del Toro (BT) to the *melicerta*-like population from Darién (D). The factorial correspondence analysis (FCA) showed no genetic structure. More precisely, the specimens belonging to the two *H. hecale* colour races were not at all discriminated in the factor space (Figure 2.3), suggesting that the genetic composition similarity is not greater between than within races. The estimated *Fst* value based on microsatellites was not significantly different from zero (0.009, 95% confidence interval calculated on 1000 bootstraps from the markers:  $-0.003$  to  $0.026$ ). The analysis of race differentiation based on re-sequenced whole-genomes from populations Chiriquí (Ch) and Darién (D) also showed a very low genome-wide *Fst* value ( $0.02 \pm 0.07$ ).

### 2.3.2 The profile of differentiation between the parapatric races is highly variable throughout the genome

Sliding window analyses revealed a strikingly high number of peaks having *Fst* values exceeding the genomic threshold (corresponding to the genome-wide mean), and drawing a highly heterogeneous landscape of genomic differentiation between the races of *H. hecale* studied (Figure 2.4A). Differentiation peaks exhibited, with some exceptions, rather low *Fst* values (mainly ranging from 0.2 to 0.3). The differentiation profile was largely robust to changes in window size, although with somewhat shallower peaks (results not shown). Because *Fst* is a measure of relative divergence which can be inflated by low within-group nucleotide diversity

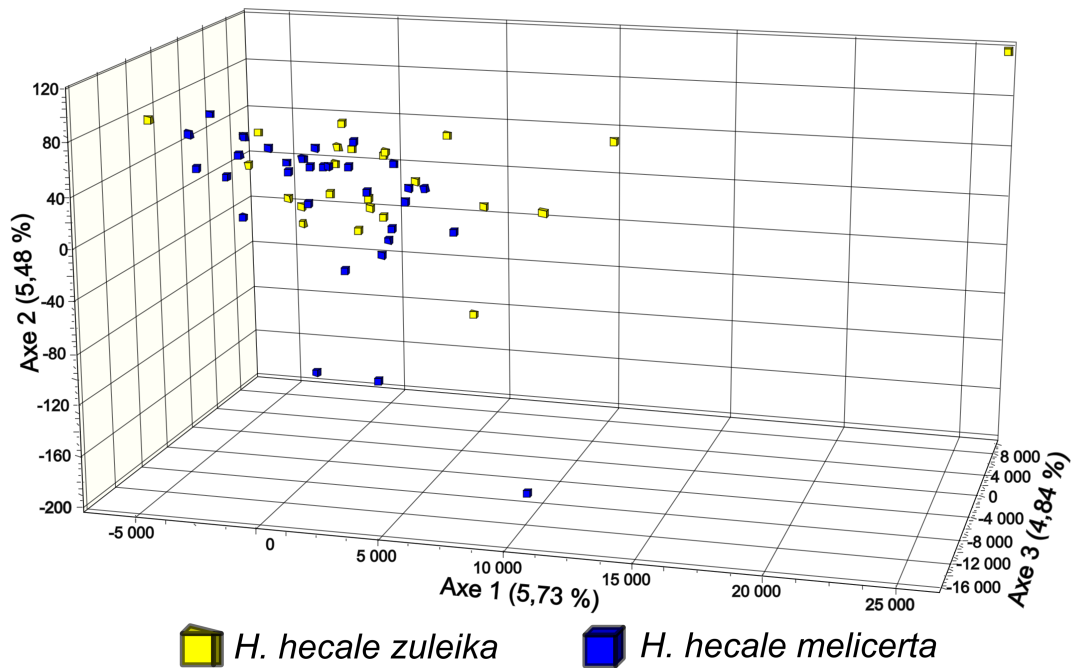


Figure 2.3 – **Factorial Correspondence Analysis (FCA) of the microsatellite data.** *H. hecale zuleika* (in yellow boxes) and *H. hecale melicerta* specimens (in yellow boxes) were collected in Bocas del Toro and Darién populations, respectively.

([Charlesworth, 1998]), some of the observed differentiation peaks might be the by-product of the low nucleotide diversity observed in *H. hecale* ( $P_i = 0.007$ ). Accordingly with this idea, genome-wide plots of  $F_{st}$  and  $P_i$  show that these summary statistics are locally inversely correlated (Figure 2.8).

### 2.3.3 Wing patterning loci are major genomic islands of differentiation between *H. h. zuleika* and *H. h. melicerta*

Some of the highest divergence peaks were related to the wing colour loci mapped in *H. hecale* (Figure 2.4B), and presumably result from the strong selective forces for mimicry. The highest peak of differentiation across the genome ( $F_{st} \sim 0.6$ ) is located close to colour locus *HhBr*, which sits in an interval encompassing *optix* ([Huber et al., 2015]); its phenotypic effect can be seen in Figure 2.2). *Optix* is known as a major colour patterning gene whose expression prefigures the variety of red wing elements in other *Heliconius* species (e.g. *H. melpomene*, *H. cydno* and *H. erato*; [Reed et al., 2011]; [Martin et al., 2014]), and functional sites appear to sit in its regulatory region ([Reed et al., 2011]; [Pardo-Diaz et al., 2012]; [Heliconius Genome Consortium, 2012]; [Supple et al., 2013]). Similarly, we detected a clear peak of differentiation close and upstream to a gene named *WntA* and nowhere else within the scaffold containing this gene (Figure 2.4A). *WntA* was found to control melanic patterns near to the forewing discal cell in several *Heliconius* species ([Martin et al., 2012]) and also in *H. hecale* ([Huber et al., 2015]; Figure 2.2 shows the phenotypic effect of *HhAc*, this locus being homologous to *WntA*). In contrast to the results we just described, no clear molecular signals of selection were detected on the genomic scaffold containing a third colour region in *H. hecale*, *HhN* (Figure

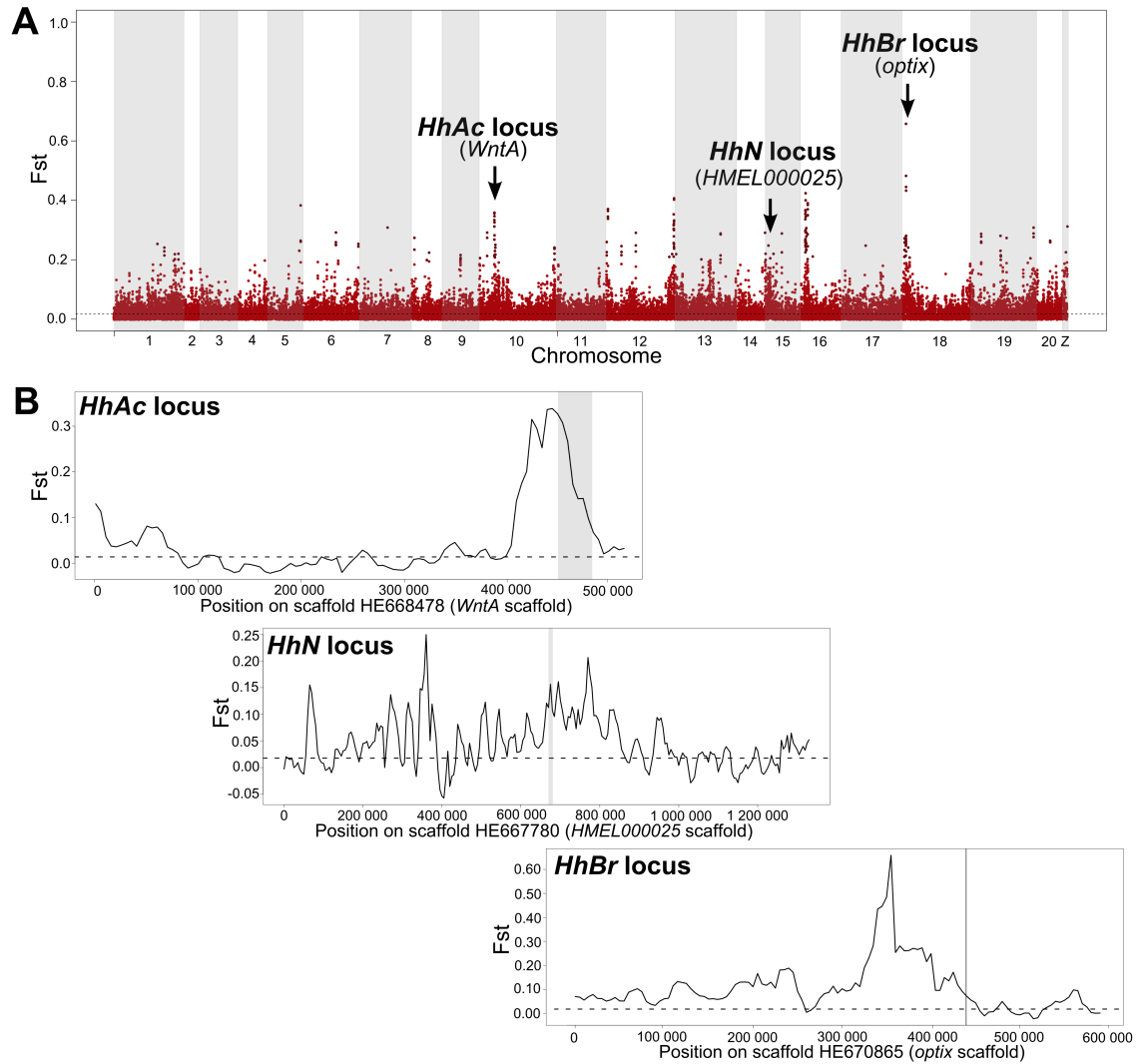


Figure 2.4 – Genome-wide genetic differentiation ( $F_{st}$ ) between *H. hecale zuleika* and *H. h. melicerta*. (A) Divergence landscape across the 21 chromosomes. The genomic location of colour loci on the genome is highlighted with black arrows. Candidate genes for each locus are shown in brackets under the colour locus name. (B)  $F_{st}$  statistic along the genomic scaffolds containing the colour loci. The position of candidate genes (*WntA*, *HMELO00025* and *optix*) is indicated in each box by a vertical gray box or a line. Statistics were calculated for overlapping 20-kb windows moving in 5kb intervals. The genomic threshold of differentiation is set by a horizontal dashed line. Five genomes of each race were used for comparison.



Table 2.1 – Maximum likelihood parameter estimates for the phenotypic clines of the three colour loci determining most of wing pattern differences between *H. hecale zuleika* and *H. h. melicerta*

Colour locus	Width (km)	Centre (km)	pmin	pmax
<b>Hybrid index score</b> <b>(three loci together)</b>	91.52 (63.10-196.17)	421.15 (402.58-435.64)	0.04 (0.01-0.07)	0.95 (0.92-0.98)
<b>HhAcz</b>	70.41 (5.15-122.29)	425.02 (409.55-458.82)	0.00 (0.00-0.02)	0.94 (0.89-0.96)
<b>HhNz</b>	16.83 (1.47-100.03)	386.11 (379.41-421.25)	0.08 (0.05-0.12)	0.92 (0.89-0.95)
<b>HhBrz</b>	186.27 (146.65-219.76)	425.25 (405.32-440.33)	0.00 (0.00-0.02)	0.99 (0.96-1.00)

Two log-likelihood unit support limits are shown in brackets. Cline width corresponds to 1/maximum slope. The center of the cline is measured as the distance from the westernmost locality in Figure 2.1. Parameters *p*min and *p*max are the minimum and maximum estimated frequency at the eastern end and western end of the cline, respectively.

2.4B; [Huber et al., 2015]). This region assembles a cluster of loci (*Yb-Sb-N/Cr/P*) controlling several white/yellow wing colour elements in different *Heliconius* species ([Joron et al., 2006]; [Ferguson et al., 2010]). A strong candidate gene within this region is *HMELO00025*, named *poik* ([Nadeau et al., 2015]) which location is shown in Figure 2.4B. As observed in this figure, no clear differentiation peak is associated to this gene.

### 2.3.4 Phenotypic clines of the three colour loci are consistent but not concordant in the *H. hecale* hybrid zone

Mathematical cline fitting worked well under the chosen model (all four parameters were allowed to vary), both at the centre and the tails of the sigmoidal clines (Figure 2.5). We characterised the width and the centre of the clines for the three colour loci together (based on the hybrid index score; Figure 2.5A and Table 2.1) and for each colour loci separately (Figure 2.5B and Table 2.1). A maximum-likelihood (ML) estimate of the centre of the hybrid zone accounting for joint allelic variation at the distinct loci located it at 421.15 km from the Costa Rican population, that is to say, between populations Veraguas (V) and Coclé (Cc) (Figure 2.5A). However, hybrid specimens are found far outside this centre (see phenotypic hybrid index pie-charts in Figure 2.5A). Namely, a few specimens having a hybrid phenotype for one of the loci (results not shown) were even collected in the populations from the extremes of the sampled transect (Costa Rica and Darién), despite the high phenotypic homogeneity in these populations. The width of the hybrid zone is of 91.52 km (Table 2.1). The location of the cline centres for the colour loci separately (Figure 2.5B) does not differ significantly, and is almost identical among loci *HhAc* and *HhBr* (425.02 km and 425.25 km, respectively), but slightly distinct between any of these and *HhN* (386.11 km) (Table 2.1). Hence, these three clines are consistent with each other. In contrast, ML estimates of cline width are highly variable among the three colour loci, ranging from 16.83 km for *HhN* to 186.27 km for *HhBr* (Table 2.1). The difference in clines width among loci suggests contrasted levels of selection on the different colour pattern elements.

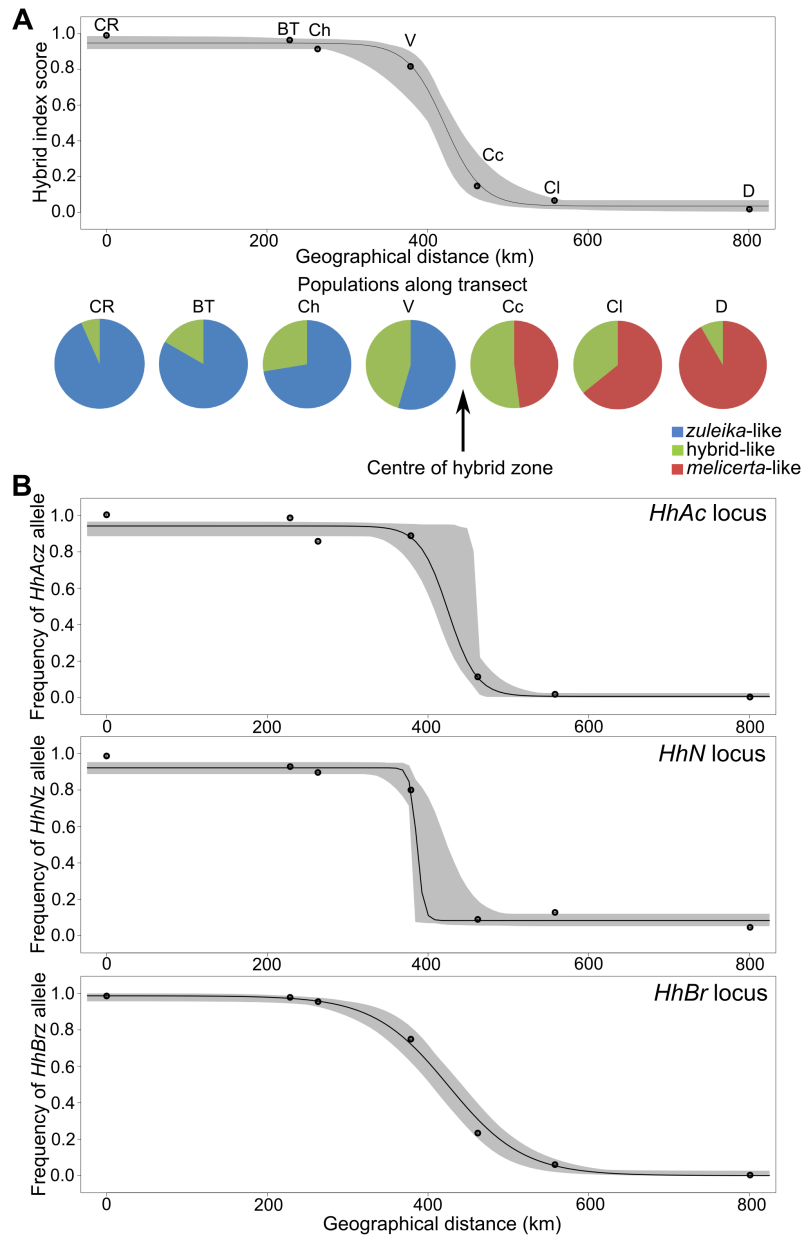


Figure 2.5 – **Phenotypic clines for the three wing colour loci and distribution of genotypic classes along the studied transect between *H. hecale zuleika* and *H. hecale melicerta*.** (A) Cline (upper panel) and phenotypic hybrid index (lower panel) based on the genotypes of the three colour loci (*HhAc*, *HhN* and *HhBr*) together. Cline is drawn as explained in B, but using the per population phenotypic hybrid index based on the three loci together. This index is based on the frequency of pure *H. hecale zuleika* individuals (*zuleika*-like), phenotypically pure *H. h. melicerta* (*melicerta*-like) and specimens of mixed ancestry for at least one of the colour genes (hybrid-like). (B) Plot of the fitted cline (line) to the observed frequency data (black dots) over the associated 95% credible cline region (gray area) for each colour loci separately. The frequency of the *H. hecale zuleika* alleles is plotted. A model where all four parameters (cline width, cline centre, *p*<sub>min</sub> and *p*<sub>max</sub>) were set free was used. Analyses were performed within the *HZAR* package of *R*. The name of the populations along the transect are as follows: Costa Rica (CR), Bocas del Toro (BT), Chiriquí (Ch), Veraguas (V), Coclé (Cc), Colón (Cl) and Darién (D). The location of the centre of the hybrid zone is shown with an up-pointing black arrow.

### 2.3.5 *H. hecale melicerta* males court preferentially their own race, but *H. hecale zuleika* males do not

In order to evaluate pre-zygotic isolation between the two races of *H. hecale* as a result of assortative mating, we tested males of the two studied races for their courtship preference, using female wing models and real females of both races. We recorded a total of 837 and 608 courtship events directed toward *H. hecale melicerta* and *H. h. zuleika* female wing models, respectively, by 31 *H. h. melicerta* males (Table 2.3). For 32 tested *H. h. zuleika* males, a total of 422 and 455 events were registered toward the *H. hecale melicerta* and *H. h. zuleika* models, respectively (Table 2.3). By implementing a likelihood approach based on these data, we found that *H. h. zuleika* males did not show significant assortative mating whereas *H. h. melicerta* males significantly preferred to court the wing models of their own race rather over the *H. h. zuleika* models ( $G = 21.439$ ,  $P < 0.001$ ; see Figure 2.6).

Among the 25 tetrad mate-choice experiments, 76% of them ( $N = 19$ ) resulted in mating (Table 2.3). The results of these trials confirmed the asymmetric assortative mating trend observed in the experiments using models (Figure 2.6). Matings happened significantly more frequently within *H. h. melicerta* pairs (8.5 out of 19) than within *H. h. zuleika* pairs (1.5 out of 19) ( $X^2 = 4.90$ ,  $df = 1$ ,  $p = 0.03$ ). We also registered more matings between *H. h. melicerta* males and females than between *H. h. melicerta* males and *H. h. zuleika* females (8.5 versus 3) although this difference was not significant ( $X^2 = 2.63$ ,  $df = 1$ ,  $p = 0.10$ ). Also, more matings happened between *H. h. zuleika* males and *H. h. melicerta* females than between *H. h. zuleika* males and *H. h. zuleika* females (6 versus 1.5), but this difference is not significant ( $X^2 = 2.70$ ,  $df = 1$ ,  $p = 0.10$ ). Thus, consistently with the male choice results based on wing models, *H. h. zuleika* males do not seem to care about mating with their own race.

## 2.4 Discussion

### 2.4.1 Extensive gene flow exists between parapatric races of *H. hecale*

Using both hyper-variable microsatellite markers and a hyper-dense array of markers, we observed no genetic structure between parapatric races *H. hecale zuleika* and *H. h. melicerta*, meaning that their gene pools are actively mixing. However, cases of even weaker genomic structure between species have been reported (e.g. [Puebla et al., 2014]). Thus, the lack of genetic structure between the *H. hecale* races may not be sufficient to assess their current stage of speciation.

The low differentiation found here across the transition zone between two races of *H. hecale* diverging in wing pattern and mimetic alliance is reminiscent of the lack of genetic structure between parapatric races of *H. melpomene*, based on RAD markers ([Nadeau et al., 2012]) or whole-genome sequencing ([Martin et al., 2013]). In contrast, across Costa Rica, phenotypically monomorphic populations of *H. erato petiverana* located tens of kilometres apart were found to be strongly differentiated, based on AFLP markers ([Kronforst and Gilbert, 2008]), mitochondrial DNA and nuclear genes ([Hill et al., 2013]). However, Kronforst and Gilbert ([Kronforst et al., 2007]) did not detect genetic discontinuities among populations of *H. hecale zuleika* in Costa Rica. The strong contrast observed between the absence of geographical genetic structure in *H. hecale* and *H. melpomene* and the mosaic of genetically distinct populations within *H. erato* suggests potential differences in dispersion abilities, local ecological adaptations and demographic histories across the genus *Heliconius*.

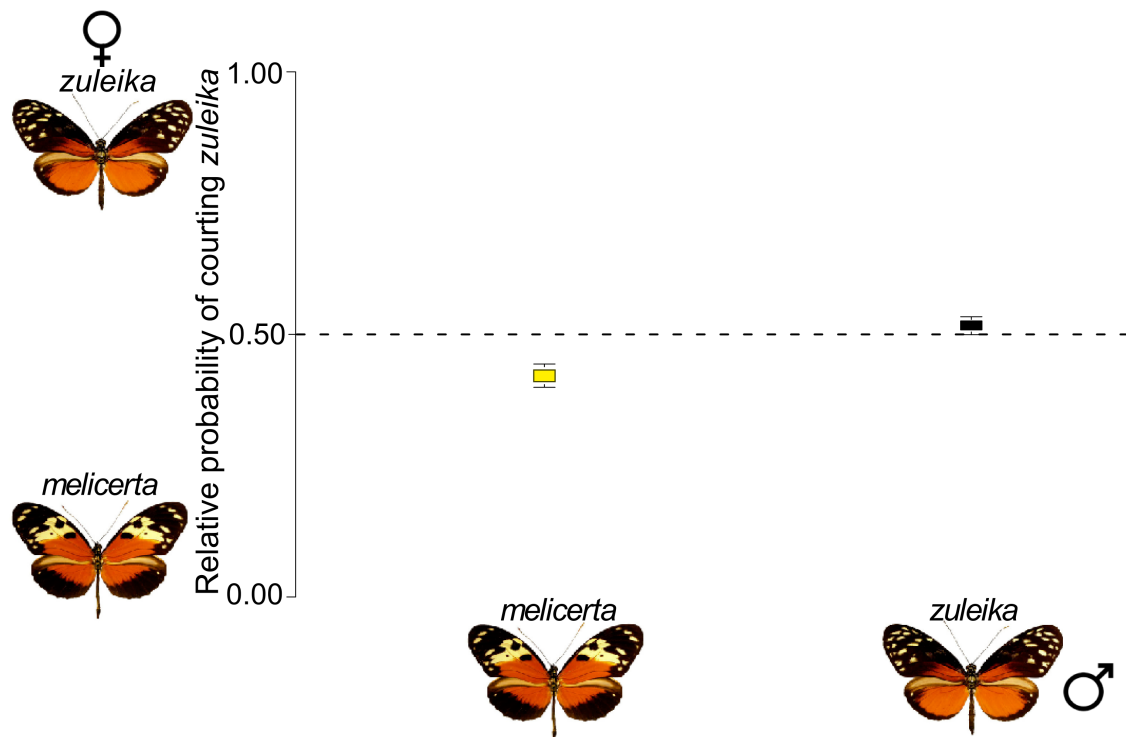


Figure 2.6 – Male choice of *H. hecale melicerta* and *H. hecale zuleika* between female wing models of both races. Values are estimated relative to the probability of courting *H. hecale zuleika* models, where 1 means a complete choice for *zuleika* and 0 a complete choice for *melicerta*. The 0.5 value means no preference. Several males of each race ( $N = 32$  *zuleika* and  $N = 31$  *melicerta*) were tested. Error bars show support limits equivalent to 95% confidence intervals.

### 2.4.2 Peaks of differentiation are scattered along the genome and include the colour loci

Despite the overall low divergence observed between *H. hecale zuleika* and *H. h. melicerta*, the sliding-window approach revealed high heterogeneity in the *Fst* estimates throughout the genome, which contrasts with the expectations for early stages of diversification under a verbal model of diversification and speciation with gene flow (see [Feder et al., 2012]). Also, the high number of shallow differentiated regions we found in the intraspecific *H. hecale* comparison contrasts with the few genomic islands of divergence (c.f. [Turner et al., 2005]) detected between parapatric races of *H. melpomene*, exclusively located around the wing colour pattern loci ([Martin et al., 2013]). Islands of divergence are usually thought to be regions under disruptive selection that contrast with the surrounding genome, where alleles migrate freely and have their frequencies homogenised across populations ([Feder et al., 2012]). These peaks of differentiation have been proposed to have a central role in early stages of speciation ([Harr, 2006]; [Carneiro et al., 2010]; [Nadeau et al., 2012]; [Feder et al., 2012]) and to be frequently associated with speciation genes ([Nosil and Schluter, 2011]). However, the causes of these divergence patterns are now heavily debated, and features of the genomic architecture (e.g. low recombination rates and low nucleotide diversity) have been suggested to account for the differentiation outliers ([Noor and Bennett, 2009]; [Bierne et al., 2011]; [Cruickshank and Hahn, 2014]). Recently, Cruickshank and Hahn ([Cruickshank and Hahn, 2014]) reanalysed published datasets of studies reporting genomic islands of divergence and found that the estimation of the  $d_{XY}$  statistic, an absolute measure of divergence which does not depend on within group diversity, failed to recover such islands detected with *Fst*-based methods. We believe that some of the islands found here display inflated *Fst* because of low nucleotide diversity ( $\pi$ ) ([Charlesworth, 1998]; [Cruickshank and Hahn, 2014]). We did actually find a negative correlation between *Fst* and  $\pi$  (see Figure 2.8), and we observed relatively low overall nucleotide diversity in *H. hecale* as compared to *H. melpomene* ( $\pi$  averaged between different *H. melpomene* subspecies = 0.012; [Martin et al., 2013]). It is difficult to assess which ones among the differentiation peaks we detected actually represent genomic regions responding to disruptive then to stabilizing selective forces. We suggest that some of these might be the result of selective forces whereas others might be statistical artefacts, also including false positive outliers derived from probabilistic events on a low number of samples and variable markers within a window (see [Willing et al., 2012]).

We observed that some of the peaks of divergence are linked to the genomic regions having a major effect on wing-pattern differences between the races of *H. hecale* ([Huber et al., 2015]), and presumed to be under strong selective pressures for mimicry. Similar results were found when comparing parapatric races of other *Heliconius* species ([Nadeau et al., 2012]; [Martin et al., 2013]). Here, the highest peak of divergence is located around the genomic region that was suggested to contain the *cis*-regulatory elements of *optix* (Figure 2.4B; [Pardo-Diaz et al., 2012]; [Supple et al., 2013]). Strangely, *in situ* hybridizations failed to associate *optix* expression to colour pattern in *H. hecale fornarina* ([Martin et al., 2014]), but the differentiation peak we found strongly suggest a potential role of the *cis*-regulatory region of *optix* in the control pattern variation in *H. hecale*, like in other *Heliconius* species ([Reed et al., 2011]; [Pardo-Diaz et al., 2012]; [Supple et al., 2013]; [Martin et al., 2014]). This confirms that regulatory variations of gene *optix* underlie an enormous phenotypic variability in the distribution of red and orange pattern elements across the genus.

We also found a high peak of differentiation close to the candidate colour gene *WntA*, which determines the position and size of melanic patterns in the forewing median region in *Heliconius*

and several other Nymphalid species ([Martin et al., 2012]; [Gallant et al., 2014]; [Martin and Reed, 2014]). In *H. hecale*, median black pattern variation was mapped to a narrow interval around gene *WntA*, and the association of *WntA* expression with pattern determination was also confirmed in this species by positional expression assays ([Huber et al., 2015]). Here, the unique peak of differentiation detected on the genome scaffold containing *WntA* falls upstream of the coding region, roughly corresponding to the region containing SNPs fully associated with variations between *H. cydno galanthus* and *H. pachinus* ([Gallant et al., 2014]).

Surprisingly, no simple signal of differentiation was detected around the third main colour locus (*HhN*) ([Huber et al., 2015]). Instead, a complex profile of alternate peaks and troughs of differentiation was observed, a pattern similar to the one observed at the same complex locus (the *Yb-Sb-N* cluster) when comparing the parapatric races of *H. melpomene amaryllis* and *H. m. aglaope* in Peru ([Nadeau et al., 2012]). Although the peaks observed in *H. hecale* did not really correspond to candidate colour loci (or their putative regulating region) within this interval (i.e. *HME000025* or *poik*; [Wu et al., 2010]; [Nadeau et al., 2015]), this kind of signal might respond to the high complexity of a region encompassing several colour loci homologous to those composing the supergene *P*, which controls local polymorphism of wing colour in the closely-related species *H. numata* ([Huber et al., 2015]; [Joron et al., 2006], [Joron et al., 2011]), resulting in a differentiation signal that is difficult to interpret.

Overall, our analyses show that some peaks of differentiation correspond to the genomic regions known to be involved in wing colour pattern variations for mimicry in *H. hecale* and other *Heliconius* species, which confirms their validity as regions potentially under strong divergent then purifying selection. Our results therefore confirm the importance of wing colour loci in local adaptation and population differentiation in this clade. However, we ignore whether other peaks of divergence correspond to distinct traits that have not been studied far, but that might be divergent between the studied races.

### 2.4.3 A hybrid zone of intermediary width exists between the races of *H. hecale*

In the populations sampled in the extremes of this hybrid zone most individuals displayed the “pure” phenotype, suggesting strong disruptive selection acting on colour patterns due to mimicry. However, we collected a few specimens that were heterozygous for one of the wing colour pattern loci. At a first glance, this showed a rather large hybrid zone, with individuals of mixed genotype found over a region of  $\sim 800$  km. [Mallet, 1993] referred to races of *H. hecale* as an example of weakly differentiated forms separated by broad clines, at the lower extremity of the continuum of geographic differentiation levels displayed by the different *Heliconius* species. However, by calculating average hybrid index scores which take into account the allelic state at all three colour loci simultaneously we showed that this hybrid zone is  $\sim 92$  km wide. This transition zone is of intermediary width when compared to other intraspecific hybrid zones studied in other *Heliconius* species. Spanning all across the spectrum mentioned by Mallet ([Mallet, 1993]), hybrid zones between races of *H. erato* from four distinct localities show widths as variable as 10, 30, 80 and 140 km ([Mallet et al., 1998a]). Overall, hybrid zones in the *erato* and *melpomene* clades have been reported to be 10 – 200 km wide ([Brown and Mielke, 1972]; [Turner, 1971]; [Benson, 1982]; [Mallet, 1986]; [Merchán et al., 2005]). However, a detailed analysis of the transects, of the reproductive barriers between lineages and of the genetic architecture of colour variation across the hybrid zone, have only been made in a limited number of cases (see Table 2.4 for a list of well understood hybrid zones in *Heliconius*).

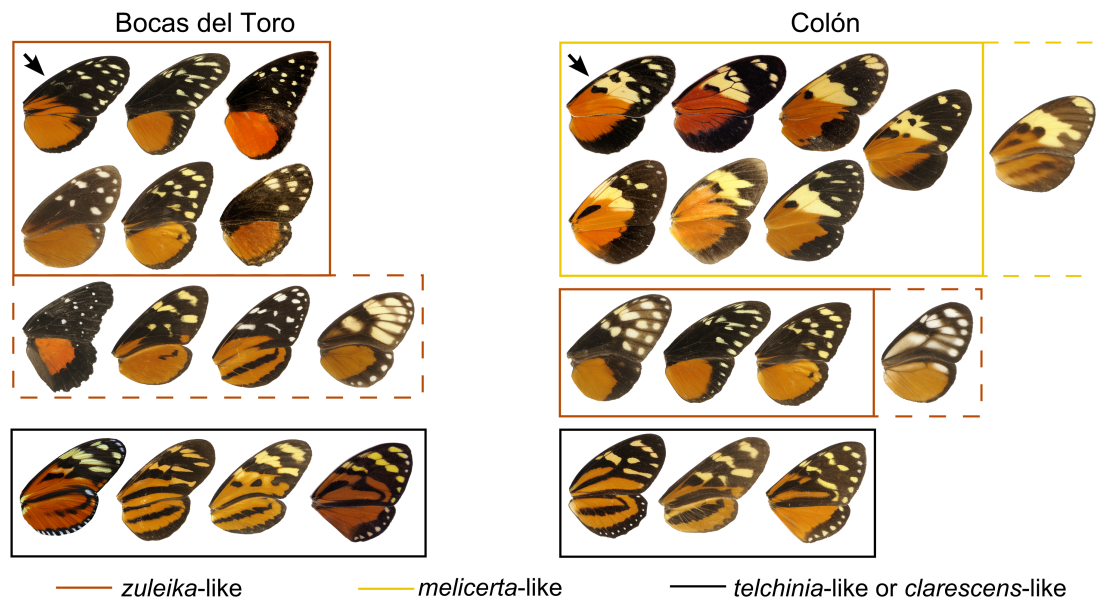


Figure 2.7 – **Partial representation of tiger-patterned mimicry rings in Panama.** This figure includes preliminary presence/absence collecting data, and is not representative of the relative abundance of each mimicry ring. Co-mimics of *H. hecale melicerta* (indicated by an arrow in the yellow box) are preponderant in the eastern population, Colón, and limited to this population. In contrast, co-mimics of *H. hecale zuleika* (indicated by an arrow in the upper-left orange box) are infrequently observed in Colón, and preponderantly found in the western population, Bocas del Toro. A third mimicry ring (black boxes) is also present in both populations, but only rarely in the Eastern one. We called it *telchinia*-like or *clarescens*-like because it contains races *H. ismenius telchinia* and *H. ismenius clarescens*, only distinguishable on the basis of the presence and absence of the hindwing medial black band, respectively. A hindwing black band of intermediary size is present in the leftmost *H. ismenius* butterfly in the black box from Bocas del Toro. By-eye designated “perfect” and “imperfect” mimics are enclosed in boxes with a continuous and discontinuous border, respectively.

The width of the colour clines is inversely proportional to the strength of selection for mimicry, thus, it depends upon the mimetic community. Tiger-patterned mimetic environments are highly variable and speciose (see Figure 2.7) and contrast with other simpler and species-poorer mimicry rings in the clade (i.e. the one formed by races of *H. melpomene* and *H. erato*). Hence, selection against some hybrids between the *H. hecale* races may not be very efficient. Mallet et al. ([Mallet et al., 1998b]) and Jiggins et al. ([Jiggins et al., 1996]) stated that the narrowness of hybrid zones may be inversely associated with the number of wing pattern elements that define the differences between forms as perceived by predators. When comparing to the best documented hybrid zones in the clade, *H. hecale melicerta* and *zuleika* are separated by the widest hybrid zone (Table 2.4), in spite of being distinguishable on the basis of three major loci. The intricate composition of wing elements of different sizes observed in tiger-patterned butterflies might be prone to be highly variable ([Le Poul, 2014]). Overall, the intermediate cline width estimated for *H. hecale* is consistent with the moderate number of genomic regions with a significant effect on wing colour pattern detected so far, and with the complexity of tiger patterns and tiger mimetic communities.

The clinal allele distributions and hence the evidence of restricted gene flow at wing colour loci strongly contrast with the absence of genetic structure at genome-wide molecular markers. This confirms that colour loci control characters directly involved in local adaptation, which is

concordant with the colour loci-related molecular signals of selection detected in this and other studies ([Nadeau et al., 2012]; [Supple et al., 2013]; [Martin et al., 2013]; [Kronforst et al., 2013]).

#### 2.4.4 Variation in the strength of selection among the three main colour pattern loci of *H. hecale*?

Over all three colour loci, the width of the Panamanian *H. hecale* hybrid zone suggests predator-operated selective forces of intermediate intensity to be acting against specimens of admixed ancestry. However, the cline shape for the three loci differs, which contrasts with the common finding of similar clines at different loci, chromosomes and characters involved in hybrid zones ([Barton and Hewitt, 1985], [Barton and Hewitt, 1989]). Cline discrepancy associated with differential introgression among distinct loci across the same hybrid zone has been commonly attributed to gene-specific natural selection ([McDonald, 1994]), but also to genetic drift ([Bierne et al., 2003]) or hybrid zone movement ([Stuckas et al., 2009]). Differential selective pressures would more readily influence deviations from concordance (i.e. equal widths or shapes) than from coincidence (i.e. equal positions), since only the shape of the cline is directly associated with the selection coefficient ([Barton and Gale, 1993]). In the *H. hecale* hybrid zone, the width of the phenotypic clines is variable, with loci *HhN* (controlling submarginal yellow elements in the forewing) and *HhBr* (controlling the shape of the black hindwing marginal band) showing the steepest and the flattest hybrid zones, respectively, and locus *HhAc* (controlling median forewing melanisation patterns) displaying an intermediary width. For locus *HhAc* and to a lesser extent for locus *HhBr*, the frequency of heterozygous might have been underestimated on the left tail of the cline, artificially decreasing the estimated width of the clines. The causative SNPs are yet to be identified for these loci and would allow a precise genotyping of the wild samples in order to refine the width of the clines. Nevertheless, the difference displayed by the three colour loci suggests varying selection on different elements of the wing colour pattern. Such variation could reflect mimetic community variations and local predators' cognitive capacities. Frequency-dependent selection for mimicry is critically determined by other toxic species, and some pattern elements might be more variable than others across all species involved, which may account for differences in the strength of selection on distinct elements of the wing. This could also depend on the perceptual effect of each locus for predators. For instance, birds visual systems are thought to perceive the contrast between black and yellow in *Heliconius* as more striking than black vs. orange ([Llaurens et al., 2014]), which might result in stronger selection on the variations of black and yellow elements controlled by *HhN* and *HhAc* than on the black and orange elements controlled by *HhBr*.

In contrast to the different cline shapes among colour loci, cline centres of the three loci are concordant. Hence, linkage disequilibrium between loci could potentially result from the dispersal of combined parental alleles at each locus into the centre of the hybrid zone ([Barton and Hewitt, 1985]), although this remains to be properly tested. Thus, in spite of the high genome-wide gene flow between *H. hecale* races, strong selection maintaining the association between loci (spatial coupling of clines; c.f. ([Abbott et al., 2013])) could presumably be overcoming the effect of recombination on continuously breaking down allelic associations between these adaptive loci. Spatially coupled barriers may favour both long-term maintenance of the hybrid zone between the *H. hecale* races or further divergence.



### 2.4.5 Asymmetrical male choice between the *H. hecale* races

Disruptive selection resulting from local mimicry might promote assortative mating in *H. hecale*. Here we detect male preferences for their own colour pattern in the *melicerta* race. This finding is in accordance with previous work in other *Heliconius* species where intraspecific assortative mating was also detected ([Jiggins et al., 2004]; [Merrill et al., 2011a]). Our results extend over previous work on *Heliconius* and other butterflies to confirm that mate choice based on wing coloration is widespread as one of the mechanisms maintaining or enhancing divergence between lineages ([Rutowski, 1977b]; [Silberglied and Taylor, 1978]; [Jiggins et al., 2001]; [Fordyce et al., 2002]; [Sweeney et al., 2003]; [Robertson and Monteiro, 2005]; [Kemp, 2007]). Moreover, in *H. hecale*, like in other *Heliconius* species, male-operated colour-based assortative mating seems to evolve readily, and whether it is really a first step leading to the accumulation of other barriers is unknown. However, it seems to evolve prior to genomic incompatibilities and to habitat specialisation, alongside with divergence in wing colour patterning ([Naisbit et al., 2002]; [Jiggins et al., 2004]).

However, we detected no preferential mating within *H. hecale zuleika*. Within the *melpomene* clade, ([Merrill et al., 2011a]) also observed that male mating preferences were subtle between distinct forms of one species, and was asymmetric both in cases of local polymorphism (between morphs of *H. cydno alithea*) and parapatric variation (between races *H. melpomene aglaope* and *H. m. amaryllis*). Our results show that assortative mating is partial between *zuleika* and *melicerta* races, supporting the hypothesis that premating isolation increases along a continuum starting with polymorphic populations and ending with well isolated sympatric species with strong symmetrical assortative mating, and hence, that the *H. hecale* races stand at the beginning of this speciation continuum.

*H. hecale melicerta* acts as the choosier race in our study, and bears recessive alleles at most major colour loci. Similarly, in the polymorphic race *H. cydno alithea* in Ecuador, homozygous recessive yellow males were shown to court yellow females preferentially, whereas no choice was exerted by white males bearing at least one dominant allele ([Kronforst et al., 2006c]; [Chamberlain et al., 2009]; [Merrill et al., 2011a]). It is possible that the recessive allele, only expressed in homozygotes, is associated with stronger assortative preference than the dominant allele which is expressed in both homozygous and heterozygous offspring. The relationship between dominance of pattern elements and choosiness remains to be further investigated, as this relationship might depend on the linkage between colour pattern genes and preferences ([Merrill et al., 2011b]).

### 2.4.6 Influence of mate choice on cline movement?

Given that *zuleika* alleles are largely dominant to *melicerta* alleles, we would predict that the *zuleika/melicerta* cline in *H. hecale* should move towards the East of Panama, as a consequence of the so-called dominance drive ([Barton and Hewitt, 1985]), suggested to be one of the factors yielding the mobility of the hybrid zone between races *H. erato hydara* and *H. e. petiverana* in Panama ([Mallet, 1986]; [Blum, 2002]; see Table 2.4). Conversely, asymmetry in mate choice would favour a westward hybrid zone movement responding to asymmetric introgression. Namely, asymmetric mate choice has been reported to lead to hybrid zone shifts, especially in birds ([Parsons et al., 1993]; [Pearson, 2000]; [Bronson et al., 2003]; [Stein and Uy, 2006]). Historical data suggests that the *H. hecale* hybrid zone has been actually moving towards the west of Panama ([Mallet, 1993]; [Rosser et al., 2012]). The syntype of *H. h. zuleika* was collected around the Panama Canal in 1877 ([Rosser et al., 2012]) and this region was defined as the hybrid zone between *H. h. melicerta* and *H. h. zuleika* during the building of the Panama

Canal,  $\sim 100$  years ago ([Brown, 1976]). Co-mimics of *H. h. zuleika* are still observed in this locality, although much rarer than *melicerta* co-mimics (Figure 2.7). We showed that nowadays this population (here called Colón) is almost monomorphic for *H. h. melicerta*, suggesting that the hybrid zone moved perhaps over 100 km westward during the 20th century. Although formal evidence is needed to validate the westward movement of this hybrid zone and to test its direct link with asymmetric mate preference, our results suggest that sexual selection on wing colour pattern might be one of the mechanisms shaping population dynamics and diversification. We ignore, however, if the mimetic community as a whole may have been spreading westward, this way influencing the spreading of the *melicerta* allele.

At their current stage, neither habitat specialisation nor hybrid genomic incompatibilities seem to be acting against gene flow between the two *H. hecale* races. On the one hand, no clear habitat discontinuity seems to exist between the East and the West of Panama, while the Atlantic and Pacific slopes of the Cordillera Central are well known to have contrasted precipitation profiles ([Pyke et al., 2001]). Actually, in the *Heliconius* clade, only some rare documented cases exist reporting macrohabitat differences to be causing genetic or phenotypic ruptures between parapatric races, thus playing a role in diversification ([Blum, 2002]; [Arias et al., 2008]). On the other hand, the races studied here are not known to show any kind of intrinsic genomic incompatibilities, the hybrids being completely interfertile ([Huber et al., 2015]). Accordingly, to date, no cases of hybrid sterility have been reported between *Heliconius* parapatric subspecies. Overall, here we demonstrate the central role of wing colour pattern genes in shaping divergence at early stages of speciation in the silvaniform clade, extending over the work done so far in other clades and other types of wing patterns within the genus *Heliconius*. Ecological data are needed to understand other factors implicated at the beginning of speciation in the clade and their genetic underpinnings, which would also contribute to make the functional link to the observed genomic islands of divergence.

## 2.5 Supplementary information

### 2.5.1 Supplementary tables

Table 2.2 – Characterisation of microsatellite markers used to explore the genetic structure between *H. hecale zuleika* and *H. hecale melicerta*.

Marker name	Forward primer	Reverse primer	Repeat pattern	Number of alleles	Size range (bp)	N <sup>2</sup>		He <sup>3</sup>		Ho <sup>4</sup>	
						zuleika	melicerta	zuleika	melicerta	zuleika	melicerta
NumP2	TGATGTTCAACTTTTATATCTTTTG*	TGAGATGGCATGACTCACAT	ACAT	15	6	146–192	23	0.1635	0.1201	0.0870	0.1250
NumP10	TCGTAGGTATTCGGAGAACG*	CTCTGCGTTCCCATTAAGAA	TGG	15	4	325–229	24	0.7075	0.6795	0.4583	0.4138
NumP11	ACCACATGGGGGTCTAAAGT*	CGAACTTCGTTGCACTCT	CA	15	4	99–145	23	0.4197	0.4087	0.3913	0.4375
Num3-2	GGTCCGATTTCAATTCCT	TGCCATGGCGTTATTATATC*	AC	3	10	155–220	22	0.7531	0.7798	0.7273	0.5625
Num2-2	GTGAAACAACATCGCGTCA	GATTGTAAGTGCCTTCGGAGT*	CGGAGT	2	27	173–304	16	0.9238	0.9287	0.8750	0.7727
Num17-1	ATTTCGGCATACTGGGTGTA*	CCACGGTTAAGGAGAATACG	AC	17	16	177–201	21	0.8719	0.8909	0.8571	0.8571
Num18-2	TTGTTACATAAGGCAATTCAAA	TTGTCACCTGTTTATTACCACAGA*	?	18	3	206–212	22	0.3048	0.0624	0.0909	0.0645
Hm06 <sup>5</sup>	AAATAGTGTGCGGCGGAATA*	TGGAGTAGAAAATGCGGGTTTA	CA	3	4	233–244	21	0.2846	0.5864	0.2381	0.5185
All loci (average)				NA	9	NA	22	0.5536	0.5571	0.4656	0.4690

<sup>1</sup> Linkage group

<sup>2</sup> Number of individuals genotyped

<sup>3</sup> Expected heterozygosity

<sup>4</sup> Observed heterozygosity

<sup>5</sup> From [Mavárez and González, 2006]

\* Primers M13-tailed on the 5' extreme (used M13-tail: 5' CACGACGTTGTAAACGAC 3')

Table 2.3 – Number of courtship events directed toward *H. hecale melicerta* (*melicerta*) and *H. hecale zuleika* (*zuleika*) female wing models by males of both species ( $N = 31$  *melicerta* and  $N = 32$  *zuleika*)

Male	Female		Male	Female	
	<i>melicerta</i>	<i>zuleika</i>		<i>melicerta</i>	<i>zuleika</i>
<i>melicerta 1</i>	5	3	<i>zuleika 1</i>	14	6
<i>melicerta 2</i>	8	16	<i>zuleika 2</i>	32	22
<i>melicerta 3</i>	18	14	<i>zuleika 3</i>	11	13
<i>melicerta 4</i>	16	18	<i>zuleika 4</i>	2	5
<i>melicerta 5</i>	39	29	<i>zuleika 5</i>	1	10
<i>melicerta 6</i>	51	37	<i>zuleika 6</i>	16	19
<i>melicerta 7</i>	31	12	<i>zuleika 7</i>	8	14
<i>melicerta 8</i>	32	16	<i>zuleika 8</i>	7	10
<i>melicerta 9</i>	44	26	<i>zuleika 9</i>	8	13
<i>melicerta 10</i>	31	21	<i>zuleika 10</i>	22	24
<i>melicerta 11</i>	22	26	<i>zuleika 11</i>	5	8
<i>melicerta 12</i>	13	11	<i>zuleika 12</i>	10	21
<i>melicerta 13</i>	9	10	<i>zuleika 13</i>	9	7
<i>melicerta 14</i>	11	13	<i>zuleika 14</i>	1	2
<i>melicerta 15</i>	16	21	<i>zuleika 15</i>	4	7
<i>melicerta 16</i>	15	13	<i>zuleika 16</i>	27	23
<i>melicerta 17</i>	15	6	<i>zuleika 17</i>	13	16
<i>melicerta 18</i>	31	26	<i>zuleika 18</i>	20	25
<i>melicerta 19</i>	22	18	<i>zuleika 19</i>	4	7
<i>melicerta 20</i>	20	12	<i>zuleika 20</i>	15	12
<i>melicerta 21</i>	43	27	<i>zuleika 21</i>	6	5
<i>melicerta 22</i>	40	22	<i>zuleika 22</i>	14	17
<i>melicerta 23</i>	52	35	<i>zuleika 23</i>	9	10
<i>melicerta 24</i>	39	20	<i>zuleika 24</i>	9	13
<i>melicerta 25</i>	34	24	<i>zuleika 25</i>	16	19
<i>melicerta 26</i>	53	25	<i>zuleika 26</i>	7	3
<i>melicerta 27</i>	21	20	<i>zuleika 27</i>	36	37
<i>melicerta 28</i>	27	14	<i>zuleika 28</i>	49	45
<i>melicerta 29</i>	17	14	<i>zuleika 29</i>	18	22
<i>melicerta 30</i>	55	50	<i>zuleika 30</i>	9	6
<i>melicerta 31</i>	7	9	<i>zuleika 31</i>	20	14
			<i>zuleika 32</i>	28	28
<b>Total</b>	<b>837</b>	<b>608</b>	<b>Total</b>	<b>422</b>	<b>455</b>
<b>Average per male</b>	27.00	19.61	<b>Average per male</b>	13.61	14.68

Table 2.4 – Summary of the best documented interracial and interspecific hybrid zones in *Heliconius* butterflies

Parapatric taxa	Type of wing pattern*	Number of major loci involved	Width of hybrid zone (km)	Genetic structure or disruption detected? (kind of markers)	Habitat specialisation differences? Ecotonal divergence?	Evidences of assortative mating?	Difference in clines shape?	Hybrid genetic incompatibilities?	Shifting hybrid zone? (time in years recorded)	References***
<i>H. cydno</i> <i>H. c. weymeri</i>	Black and white-yellow patterns	4	15	Yes (microsats and AFLP, nuclear markers, mtDNA)	No**	Yes (asym)	NR	No**	No(10)	Arias et al. 2012, Merchán et al. 2005
<i>H. melpomene</i> <i>H. m. amaryllis</i>	Postman / rayed	4	~12	No (allozymes**, nuclear and mitochondrial markers)	No**	Yes (asym)	No	No	No (25)	Merrill et al. 2011; Mallet 1989; Mallet 1990; Mallet and Barton (1989); Rosser et al. 2014; Brower 1996; Mallet 2009
<i>H. melpomene</i> <i>H. m. rosina</i> <i>H. m. melpomene</i>	Postman with HW yellow band / postman without HW yellow band	1	80	No** (nuclear and mitochondrial markers)	No**	No	NA	No	NR	Mallet 1986; Jiggins et al. 2004; Hines et al. 2011
<i>H. erato</i> <i>H. e. favorinus</i> <i>H. e. emma</i>	Postman / rayed	3	~10	No (allozymes, mtDNA)	No**	No**	No	No	No (25)	Mallet 1990; Mallet 1989; Brower 1996; Mallet and Barton 1989; Rosser et al. 2014
<i>H. erato hydarata</i> <i>H. e. erato</i>	Postman / rayed	2	30-40	Yes, weak structure (microsats)	Yes	NR	Yes, clines position varied	NR	No (31)	Blum 2002; Blum 2008; Benson 1982

<i>H. erato petiverana</i> / <i>H. e. hydara</i>	Postman with HW yellow band / postman without HW yellow band	1	80	No** (nuclear and mitochondrial markers)	No**	Yes** (only petiverana males tested)	NA	No	Yes (17)	Mallet 1986; Estrada and Jiggins 2008; Hines et al. 2011
<i>H. erato venus</i> / <i>H. e. chestertonii</i>	Postman / iridescent black with HW yellow bar	2	~4**	Yes (microsats, mtDNA)	Yes	Yes (symm)	NR	Yes	NR, hybrid position is tonal	zone Arias et al. 2008, Munoz et al. eco-2010
<i>H. erato cyrbia</i> / <i>H. himera</i>	Yellow and red banded FW and HW , respectively / postman, iridescent black	2	5	Yes (allozyme, mtDNA)	Yes	Yes (sym)	NR	No	NR, hybrid position is tonal	zone Jiggins et al. 1996, 1997; eco- Jiggins McMillan et al. 1997; Jiggins and McMillan 1997; Mallet et al. 1998
<i>H. hecale melicerta</i> / <i>H. h. zuleika</i>	Variation of tiger patterns	3	92	No (microsats and whole-genome resequencing)	No**	Yes (asym)	Yes, clines position is the same	No	Yes?	This study

\* FW is the abbreviation for forewing, HW is the abbreviation for hindwing

\*\* Based on preliminary observations

\*\*\* References correspond to the works concerning the direct study of the parapatric lineages. Look for references therein to get information about colour loci mapping

NA: it does not apply

NR: not explicitly reported

sym: Symetrical

asym: Asymetrical

## 2.5.2 Supplementary figures

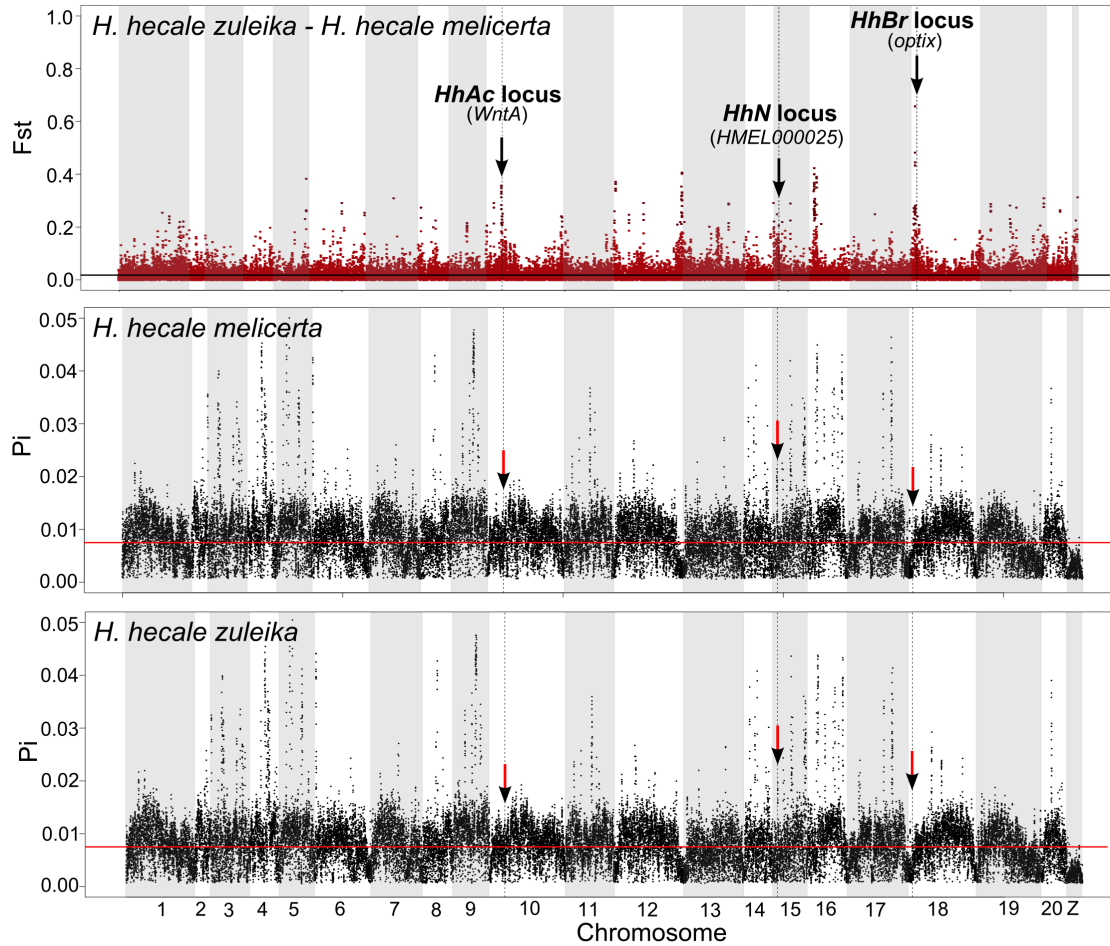


Figure 2.8 – Genome-wide differentiation ( $F_{st}$ ) between *H. hecale zuleika* and *H. h. melicerta* and nucleotide diversity ( $P_i$  or  $\pi$ ) of both races. Statistics were calculated for overlapping 20-kb windows with increasing steps of 5kb. Chromosomes are indicated on the  $X$ -axis with alternating gray and white vertical bands. The genomic threshold for a given plotted statistic is represented as a horizontal solid line. The position of the three wing colour loci is indicated by vertical arrows, with the name of the identified causal genes or candidate genes in brackets.

## Chapter 3

# Coexisting closely-related species under the same disguise: insights into species recognition issues between *Heliconius hecale* and *H. ismenius*

Bárbara Huber (1,2,6), Annabel Whibley (1), Catalina Estrada (2,3), Darha Solano (2), Robert Orpet(2), W. Owen McMillan (2), Brigitte Frérot (4), Mathieu Joron (1,2,6)

1. Institut de Systématique, Evolution, et Biodiversité, UMR 7205 CNRS, Muséum National d'Histoire Naturelle, Paris, France
2. The Smithsonian Tropical Research Institute, Balboa, República de Panamá
3. Imperial College London, Silwood Park, London, UK
4. Institute of Ecology and Environmental Sciences, Centre INRA, Versailles, France
5. Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, Montpellier, France
6. Laboratoire Biologie Intégrative des Populations, Ecole Pratique des Hautes Etudes (EPHE), Paris, France



## Abstract

Characterising genetic exchange between lineages at different stages of the speciation continuum is important to understand the evolutionary history of reproductive isolation. *Heliconius* butterflies have undergone an adaptive radiation in wing colour patterns as a result of natural selection for local mimicry and are also now well known for assortative mating based on wing pattern. Because of the double effect of wing pattern shifts on sexual isolation and on selection against non-mimetic hybrids, wing patterns are often considered “magic traits” acting against hybridisation and facilitating speciation. Here, we investigate the relative importance of wing pattern and chemical differentiation in determining hybridisation between closely-related species *H. hecale* and *H. ismenius* in geographic regions where they are mimetic and regions where they are not. We test for higher signatures of genomic introgression where species share wing patterns, but show that the genomic landscape of differentiation is similar in both geographic regions. Nevertheless, we show that major wing pattern alleles still appear to be introgressed between co-mimics, suggesting a key role for adaptive introgression in the context of mimicry evolution, but not resulting in a more permeable species barrier. This, and behavioural assays suggest that other cues are important in determining strong sexual isolation. We test for chemical differentiation between species and identify important differences in putative pheromones which we hypothesise mediate mate choice and the maintenance of species differences. Our results show that evidence for strong nucleotide differentiation and clear adaptive introgression at wing pattern loci may be partly misleading about their functional role in ecological speciation. In a clade where assortative mating is readily triggered by wing pattern differences, our results show that this alone may not lead to speciation and that the accumulation of other barriers to gene flow may be important in the completion of speciation.

## 3.1 Introduction

Few concepts have been as controversial in the field of evolutionary biology as the species concept. Nevertheless, it underlies our understanding of life diversification and more specifically, of the speciation process. Mayr’s depiction of the living world as discrete clusters of reproductively isolated entities called species, has widely been substituted by the idea that distinct species are characterised by substantial but incomplete reproductive isolation ([Coyne and Orr, 2004]). However, despite the permeability in their boundaries, species are still discrete entities that can be distinguished on a morphological, genetic, ecological and/or behavioural basis (e.g. [Noor et al., 2000]; [Saint-Laurent et al., 2003]; [Kraus et al., 2012]). This relaxed biological species concept allows seeing speciation as a continuous process where divergence accumulates in different ways including neutral divergence, local adaptation and coevolution, with introgressive hybridisation occurring inevitably throughout most speciation events, excepting cases of instantaneous or strictly allopatric speciation. Although a consensus exists now about the importance of hybridisation on diversification ([Stebbins, 1959]; [Arnold, 1992]; [Dowling et al., 1997]; [Seehausen, 2004]; [Barton, 2013]; [Abbott et al., 2013]), understanding how differences between populations can accumulate and how the species boundaries can be kept in spite of the homogenising effect of gene flow remains largely unsolved.

Complete reproductive isolation is not a requirement for diversification and not even for speciation to be achieved. However, divergence and speciation necessarily involve acquiring reproductive barriers, which are features able to decrease gene flow between lineages. Isolating barriers cumulate and/or get stronger with the advancement of the speciation process ([Coyne and Orr, 2004]), and this actually determines the gradual nature of this process. Following

Seehausen et al. ([Seehausen et al., 2014]), reproductive isolation can be conveniently classified into extrinsic post-zygotic isolation (resulting from divergent ecological selection and depending on the interaction with the environment), intrinsic post-zygotic isolation (deriving from genetic incompatibilities) and pre-zygotic isolation (responding to the effects of divergent mate choice, habitat preference, reproductive timing and gametic incompatibility). How these isolating mechanisms evolve and accumulate, and what is the relative importance of given barriers in relation to others, are questions that need to be tackled in order to understand the process of divergence with gene flow. Also, pinpointing traits that actually or potentially prevent gene flow between coexisting lineages helps understanding biological diversification. Although the barriers we detect at present are not necessarily the ones that were implicated in the initial reduction of gene flow between populations, and although their current importance may distort their historical importance during the speciation process ([Coyne and Orr, 2004]), they still inform about the evolution of reproductive isolation between taxa.

In recent years, the emerging fields of speciation genomics and population genomics have allowed the investigation of the genetic underpinnings of diversification with gene flow. Namely the search for barrier loci (i.e. loci underlying traits involved in reproductive isolation) and the more broad exploration of the genomic architecture of divergence (i.e. the number, size and distribution of the most divergent regions), even in non-model taxa ([Hohenlohe et al., 2010]; [Lawniczak et al., 2010]; [Ellegren et al., 2012]; [Kronforst et al., 2013]; [Seehausen et al., 2014]). Also, the possibility of evaluating the importance of genetic exchange in the evolutionary history of differentiating populations and lineages at different stages of the speciation continuum, has also been enhanced by genomics. For instance, in butterflies of the neotropical genus *Heliconius*, [Martin et al., 2013] found that gene flow has persisted throughout the  $\sim 2$  million years of divergence between sympatric sister species *H. melpomene* and *H. cydno*. Currently, across the range of those two species, a high fraction of the genome clusters by geography rather than by species, showing rampant admixture. Nevertheless, *H. melpomene* and *H. cydno* remain “good” species, with distinct ecologies and strong barriers to gene flow, including both pre- and post-zygotic isolation ([Jiggins et al., 2001]; [Naisbit et al., 2002]; [Merrill et al., 2012]). These results confirm pervasive gene flow between diverging species despite the rarity of hybrids in nature ([Gilbert, 2003]; [Bull et al., 2006]; [Kronforst et al., 2006b]; [Heliconius Genome Consortium, 2012]). *Heliconius* butterflies hence provide a great study system to investigate how divergence proceeds in presence of gene flow.

*Heliconius* butterflies have undergone a recent adaptive radiation in wing colour patterns as a result of natural selection for local mimicry. In this and other adaptive radiations, rapid diversification was proposed to be associated with high rates of hybridisation ([Seehausen, 2004]). Hybridisation would play a creative role by favouring the generation of genetic diversity and enhancing the response to selection. In *Heliconius*, colourful wing patterns serve as warning signals of their unpalatability to predators that learn the appearance of good versus defended prey. Warning patterns are often shared locally by several toxic species, an adaptive convergence called Müllerian mimicry. Hybrid intermediates tend to bear non mimetic patterns which are under stronger risk of attack than parental forms ([Merrill et al., 2012]), which represents an extrinsic post-zygotic barrier to gene flow. Moreover, wing patterns are involved in male-operated mate choice, with males preferentially courting females carrying their own wing colour pattern. Assortative mating based on wing pattern was proposed to be important in the speciation between *H. melpomene* and *H. cydno*, or *H. erato* and *H. himera* ([Mallet et al., 1998b]; [Jiggins et al., 2001]), but also operates between geographic races of *H. melpomene* ([Jiggins et al., 2004]; [Merrill et al., 2011a]). Wing coloration differences therefore also represent a pre-zygotic isolation barrier. The contribution of female choice based on wing

coloration is largely unknown (but see [Boyko, 2005]). Also, the role of chemical signals as pre-zygotic isolating factors promoting diversification and speciation in *Heliconius* butterflies has been until recently virtually unexplored in the genus ([Mérot et al., 2015]; [Estrada et al., 2011]), despite its widespread importance in several species of Lepidoptera, where it is implicated in intraspecific and interspecific communication as well as in sexual selection ([Vane-Wright and Boppre, 1993]; [Millar, 2000]; [Jurenka et al., 2003]; [Schulz et al., 2008]; [Nieberding et al., 2008]).

As a consequence of the combination of local convergence for mimicry and broad-scale geographic divergence, *Heliconius* butterflies represent a rich array of phenotypically differentiated lineages spanning the speciation continuum from polymorphic populations to parapatric races, and to fully isolated species in sympatry. Assortative mating based on wing patterns increases throughout this continuum ([Merrill et al., 2011a]). Genomic surveys of differentiation among *Heliconius* species and subspecies have given insights into the extent of gene flow, and into the rate and dynamics of genomic divergence during speciation ([Nadeau et al., 2012], [Nadeau et al., 2013]; [Martin et al., 2013]; [Kronforst et al., 2013]) motivating further research on how genomic divergence is built through time.

Here, we investigate whether introgressive hybridisation is common throughout the genus *Heliconius*. We focus on closely-related species *Heliconius hecale* and *H. ismenius*, which diverged  $\sim 3$  million years ago ([Kozak et al., 2015]) and whose distributions overlap in the northwest of South America and the south of Central America. This allows us to evaluate gene flow at a putatively more advanced stage in speciation than previously reported. *H. hecale* and *H. ismenius* belong to the so-called silvaniform clade in which species mostly participate in “tiger-patterned” mimicry rings with members of other butterfly families, especially Danainae and Arctiidae. Whole genome resequencing complements previous work on early stages of divergence in the genus, and allows exploring how the accumulation of reproductive isolation accumulates through the speciation process.

By playing a role in survival and in mate choice, mimicry shifts may favour speciation ([Jiggins et al., 2001]; [Mallet, 2010]) and sister species often belong to distinct mimicry rings (but see [Brower, 1996]; [Giraldo et al., 2008]; [Mérot et al., 2013]). However, closely-related species *H. hecale* and *H. ismenius* share the same wing patterns in certain parts of their range. In eastern Panama, western Colombia and Venezuela *H. h. melicerta* and *H. i. bouletti* are extremely similar and join the same mimicry ring with many other species. In contrast, in western Panama, the two species (*H. h. zuleika* and *H. i. clarescens*) have distinct patterns, each one joining a distinct mimicry ring. Here, based on the prediction that wing coloration differences should not operate as barriers to gene flow, we use behavioural, chemical and genomic approaches to test whether species boundaries are more permeable to gene flow where the two species share the same wing pattern, and ask which recognition cues may be operating in keeping species separate.

## 3.2 Materials and methods

### 3.2.1 Specimens for genomics

18 specimens of the *H. hecale* and *H. ismenius* were collected from two populations in Panama (Figure 3.1 and Table 3.2). These species are excellent co-mimics of each other In Eastern Panama (subspecies *H. h. melicerta* and *H. i. bouletti*), but display distinct patterns in Western Panama where they join different mimicry rings (subspecies *H. h. zuleika* and *H. i. clarescens*). Three to five specimens of each subspecies were sampled from Darién (East of Panama) and

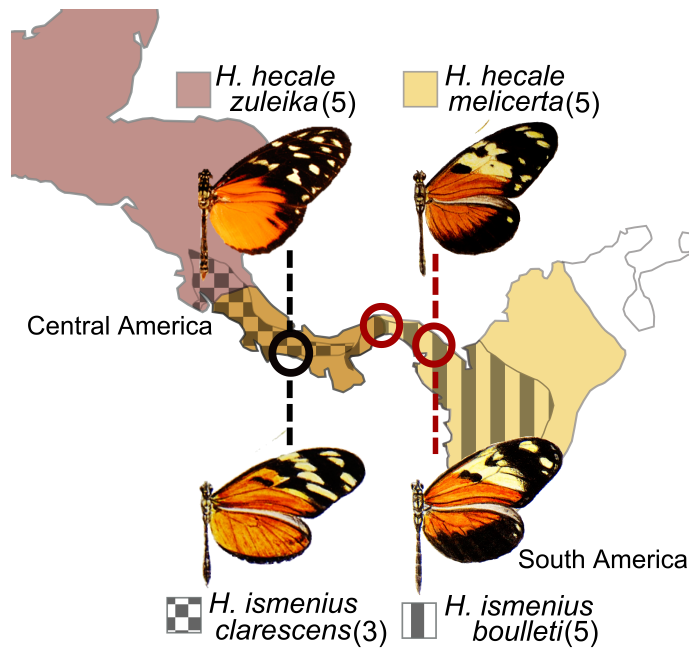


Figure 3.1 – **Distribution (shaded areas) and sampling localities (circles) of the subspecies of *Heliconius hecale* and *H. ismenius* involved in this study.** The non mimetic pair of races on the left (*H. hecale zuleika* and *H. ismenius clarescens*) was sampled from the locality displayed as a black circle (Chiriquí, West of Panama). The mimetic pair of races on the right was sampled from two populations shown in red (Colón and Darién, Center and East of Panama, respectively). A number of specimens per subspecies (in brackets) were sampled in the populations indicated by a broken line and used for whole-genome resequencing. The butterflies for the stocks (for behavioral experiments and chemical analyses) were collected in the three localities.

near David, Chiriquí (West of Panama) (Figure 3.1). Bodies were preserved in a NaCl-saturated DMSO solution and wings in paper envelopes for phenotypic analysis.

### 3.2.2 Butterflies collection and rearing

For behavioural and chemical data we obtained butterflies of the subspecies *H. hecale melicerta* and *H. ismenius bouletti* from Darién (East of Panama) and from around Gamboa, Colón (Central Panama), and individuals of *H. hecale zuleika* from Chiriquí (Figure 3.1). Stocks of these races were reared in cages of  $2 \times 2 \times 2$  m at the Smithsonian Tropical Research Institute insectaries in Gamboa, Panama. *Passiflora* plants were used as oviposition and larval hostplants, and *Lantana* sp., *Psychotria* sp., *Gurania* sp. and *Psiguria* sp. were used as adult food plants. Artificial solution of sugar and pollen was provided as a nutritional complement.

### 3.2.3 Whole-genome resequencing and bioinformatics analysis

DNA was isolated from the preserved bodies using the DNeasy Blood and Tissue Kit (Qiagen). Separate Illumina paired-end libraries were generated according to the manufacturer's protocol (Illumina Inc.). Each library was shotgun sequenced to an average coverage of  $\sim 25\times$  on an Illumina Hi-Seq 2000 with  $2 \times 100$ -base read length. Seven samples belonging to the *H. melpomene* clade ([Martin et al., 2013]) were used as an outgroup (Table 3.2). Quality-filtered sequence reads were mapped to the reference genome scaffolds of *H. melpomene* version 1.1

([Heliconius Genome Consortium, 2012]) using *Stampy v1.0.17* ([Lunter and Goodson, 2011]) with default parameters except for setting the substitution rate to 0.05 SAM/BAM file conversion, analysis, and filtering were performed using *SAMtools v0.1.19* ([Li et al., 2009]) and *Picard Tools v1.107* (<http://picard.sourceforge.net>). PCR-duplicate reads were removed using *Picard Tools v1.107*. Local realignment around indels was performed using the *Genome Analysis Tool Kit (GATK) v2.8-1* ([DePristo et al., 2011]). Genotypes were called across all individuals using a multiallelic calling function in *GATK v2.8-1* with an expected heterozygosity of 0.015. A dataset of high quality genotypes for biallelic sites having  $> 5\times$  coverage in  $> 80\%$  of the samples,  $> 30$  genotype quality (GQ) and  $> 40$  mapping quality (MQ) was obtained and used for subsequent analyses. GATK was used for these filtering steps. A subset of unlinked sites, called the *trimmed* dataset hereafter (24513 SNPs), was obtained from the dataset containing the 18 samples, by picking one variable site every  $\sim 1$ kb. For most comparative analyses, however, we used a subset of 12 samples (3 samples of each studied race having higher coverage, which we called the *restricted* dataset; see Table 3.2) to minimise possible biases due to the unequal number of re-sequenced genomes analysed. The restricted dataset consisted of  $\sim 3.2$  millions high quality SNPs.

### 3.2.4 Statistical tests for interspecific gene flow

Three distinct and complementary analyses were performed to test for genome-wide gene flow between species *H. hecale* and *H. ismenius* as a whole, between the pair of co-mimic races and between the pair of non-mimetic races. First, we tested for genetic admixture between the individuals of the two species by applying the multilocus Bayesian clustering software *Structure v2.3.4* ([Pritchard et al., 2000]) on the trimmed dataset. Admixture analyses were performed between all 10 hecale and 8 ismenius samples, and between both pairs of sympatric races (*H. hecale melicerta* vs. *H. ismenius bouletti* and *H. hecale zuleika* vs. *H. ismenius clarescens*). We used an initial run with  $K = 1$  to estimate the allele frequency distribution parameter,  $\lambda$ . Next, we ran short clustering runs (1000 burn-in, 10000 data collection) with the estimated value of  $\lambda$ , for different numbers of clusters ( $K = 2 - 6$ ), ancestry type (with vs. without admixture) and allele frequency estimation (correlated vs. independent). Since the Markov chain convergence was achieved with the preliminary number of updates and since no important differences were observed between the tested sets of models, our long runs consisted of 10000 updates of the chain after an initial burn-in of 10000 on  $K = 2 - 4$  clusters, using an admixture model and a correlated allele frequency estimation. For each tested set of parameters, 3 iterations were run. Second, we tested for genetic differentiation between the races in sympatry using estimates of the fixation index, *Fst* ([Weir and Cockerham, 1984]), and average pairwise nucleotide divergence,  $d_{XY}$ . For this and all analyses to be described next, we used the restricted dataset. *Fst* values were calculated for non-overlapping windows across the genome using the VCFtools package ([Danecek et al., 2011]). Comparisons between the sympatric races of both species were contrasted to intraspecific comparisons. *Fst* is a relative measure of differentiation and is influenced by diversity within species. In contrast,  $d_{XY}$  is an absolute measure of divergence. We estimated  $d_{XY}$  for non-overlapping windows on the 141 longest scaffolds, covering a total of  $\sim 1/3$  of the genome. We also estimated genome-wide diversity ( $\pi$  or  $Pi$ ) and *Tajima's D* ([Tajima, 1989]) for each window using VCFtools. Finally, we compared the extent of gene flow between the pairs of sympatric races using a method that relies on the relative abundance of alternative biallelic polymorphisms, named “ABBAs” and “BABAs” ([Green et al., 2010]; [Durand et al., 2011]; see Figure 3.4). Briefly, given three taxa P1, P2, P3, where P1 and P2 are sister taxa, and an outgroup O (topology [(P1,P2),P3],O), denoting A the ancestral state

of a polymorphism and B the derived allele, we look for positions retaining a polymorphism in the crown group. Sites showing the pattern AABA, ABAA, BAAA and BBAA are sorted by lineage, and cannot inform on putative gene flow; in contrast sites showing ABBA or BABA denote an incongruent lineage sorting. Ancestral polymorphism may be retained in balanced proportions between ABBAs and BABAs, while an imbalance is expected with gene flow between P3 and either P1 or P2, posterior to the split of P1 and P2. Alternatively, an excess of ABBAs reflects a higher gene flow between P3 and P1. Here we carried out this test using two distinct configurations allowing assessing gene flow between the co-mimics *H. hecale melicerta* and *H. ismenius bouletti* (top of Figure 3.4A) on the one hand, and between the non mimetic pair *H. hecale zuleika* and *H. ismenius clarescens* on the other hand (bottom of Figure 3.4A). ABBA and BABA patterns were quantified throughout the genome in non-overlapping windows. We compared these patterns by estimating two statistics, Patterson's  $D$  ([Green et al., 2010]; [Durand et al., 2011]), based on allele frequencies at each variable site, and  $f$  ([Green et al., 2010], modified by [Martin et al., 2013]), reflecting the fraction of the genome that has been shared. A 1Mb block jack-knifing approach was implemented to calculate the mean and variance of  $D$  and  $f$ , and to test whether  $D$  differed significantly from zero. For the  $d_{XY}$  estimation, ABBA-BABA counts and comparisons, we used custom scripts developed by [Martin et al., 2015].

### 3.2.5 Courtship description

Videos of the courting and mating behaviour of *H. hecale* butterflies were taken until mating happened (successful courtship,  $N = 4$ ) or until at least one whole sequence had been registered (unsuccessful courtships,  $N = 13$ ). These were used to define the sequence of behaviours and their relative importance in courtship. The behaviours that were always recorded along the entire sequence (starting with *localization* and ending with *success* or *female rejection*) were referred to as *main* events. We called *minor* events behaviours that were less conspicuous but linked to the *main* events, and therefore always present in the courtship sequence. Finally, we described *atypical* events rarely repeated across replicates. The classification of the events of courtship in *H. hecale* has resulted from the multiple behaviour experiments performed in this and on previous studies concerning one other species in the genus ([Klein and Araújo, 2010]) and other species in the subfamily Heliconiinae ([Crane, 1955], [Crane, 1957]; [Rutowski and Schaefer, 1984]; [Mega and Araújo, 2010]).

### 3.2.6 Analysis of interspecific male-female encounters

To explore the signals involved in species recognition between *H. hecale melicerta* and *H. ismenius bouletti*, we performed no-choice heterospecific encounter experiments. Three mature males of one species were put in a  $2 \times 2 \times 2$ m cage with one newly emerged (a few hours old) female of the other species. Fifteen-minute videos were taken to register all male courtship events and female response. Nine videos for heterospecific encounters involving *H. hecale melicerta* males and 7 videos involving *H. ismenius bouletti* males were analysed to register the number main courtship events occurring. Homospecific no-choice trials were run as control using *H. hecale* ( $N = 24$ ). The total number of events in each main courtship step was recorded during 15 minutes of observation. The intraspecific trials were slightly different from the interspecific experiments, since they involved two males and two females of each of the two races *H. h. melicerta* and *H. h. zuleika*. Thus, these two types of experiments are not properly comparable against each other in terms of the number of events recorded, but they still provide

us with a good understanding of the signals that may be important for species recognition in our study species.

### 3.2.7 Male choice using female wings models

To determine the importance of female wing cues in eliciting courtship in *H. hecale melicerta* and *H. ismenius bouletti* males, we performed choice courtship experiments where males of one species were presented with female wing models of both species. We registered *localization* events (short approaches of males to models) and *hovering* (fast wing movement and nearly stationary flight next to the models for more than 3 seconds) separately. Several males reared to sexual maturity in the insectaries or collected in the field were tested on different days to obtain independent behavioural registrations for each male (i.e. technical replicates). A total of 42 *H. h. melicerta* and 35 *H. i. bouletti* were characterised by 3 registrations, plus 4 *H. h. melicerta* and 3 *H. i. bouletti* males which had only two registrations each. Models were made of real unwashed female wings, connected to a plastic wire and moved in a fluttering motion inside a  $2 \times 2 \times 2$  m cage (used device: Butterfly 2000, Adriana Tapia). Courting events were registered during 5 minutes, inverting models in the middle. The relative probability of males courting *bouletti* female models rather than *melicerta* models was estimated using a likelihood approach ([Edwards, 1972]; [Jiggins et al., 2001]). *Localization* and *hovering* were analysed separately. The probability of courtship by male  $y$  towards *melicerta* wings ( $P_{y\text{mel}}$ ) and the total number of events of male  $x$  directed towards *bouletti* models ( $n_{x\text{bou}}$ ) and to *melicerta* models ( $n_{x\text{mel}}$ ), in addition to the probability of courting *melicerta* wings by male  $y$  ( $P_{y\text{mel}}$ ), were included in the likelihood function as follows:

$$\ln(L) = \sum n_{x\text{bou}} \ln(P_{y\text{mel}}) + n_{x\text{mel}} \ln(1 - P_{y\text{mel}})$$

The solver option in Excel (Microsoft) was used to estimate the probabilities of male courtship by numerically searching for values of  $P_{y\text{mel}}$  that maximized  $\ln(L)$ . The support limits (asymptotically equivalent to 95% confidence intervals) were assessed by looking for values that decreased  $\ln(L)$  by two units ([Edwards, 1972]). To compare courtship across species, we compared two alternative models, one with equal relative probabilities ( $P_{\text{mel}} = P_{\text{bou}}$ ) and one with probabilities calculated separately for each species ( $P_{\text{mel}} \neq P_{\text{bou}}$ ). Comparison was performed by a likelihood ratio test assuming a  $\chi^2$ -distribution with one degree of freedom for the test statistic  $G = 2\Delta \ln(L)$ .

### 3.2.8 Testing for female choice based on colour in *H. hecale*

To assess whether visual cues are important in mating acceptance by *Heliconius* females, we performed experiments involving a virgin *H. hecale* female (subspecies *melicerta* or *zuleika*) and two mature males (at least 5 days old) of their own subspecies with two different colouration treatments (Figure 3.8B). *Treated* males were painted with black ink, using a Sharpie pen, to cover the white and yellow patches of dorsal and ventral forewings, following Kemp ([Kemp, 2007]). *Sham* males were used as controls and were painted on the black region of the forewing, covering a similar surface as for treated males but avoiding pattern modification. Experiments were run in  $1 \times 2 \times 2$  m cages. In this case females were over one day old, since older females are more active and may be better able to reject courting males than recently emerged females. A total of 26 trials were performed (13 involving *H. h. zuleika* females and 13 involving *melicerta* females) resulting in 20 successful courtships. Direct observation of sexual behaviour was carried out until mating occurred or for a maximum period of 3 hours, and all *main* courting events were registered.

### 3.2.9 Chemical analyses

To evaluate chemical signalling differences between *H. hecale melicerta* and *H. ismenius bouletti*, chemical compounds were extracted from three different tissues. The wing regions where forewing and hindwing overlap (hereafter denoted OW), and male abdominal glands (male claspers, C) have been proposed to contain androconia (the glands for the production of male pheromones) and to be important in the release of volatile signals by *Heliconius* males ([Emsley, 1963]; [Brown, 1981]; [Schulz et al., 2008]; [Estrada et al., 2011]). Our description of courtship behaviour suggest they might have a similar role in *H. hecale* and *H. ismenius*. Abdominal glands from virgin females (G hereafter) were also analysed. A third tissue, the cuticle of the body close to the base of the wings (B), was analysed because of the presence of small bristles which might be involved in the production of chemical signals. Tissues from 3 – 7 individuals were analysed separately for females and mature males of both species. All extractions were done right after sacrificing the individuals. Pieces of tissue were immersed into 130 – 200  $\mu$ l of hexane containing dodecane 100 ng/ $\mu$ l as internal standard. Suitable glass vials were used and the extractions were not stopped. Elution samples were analysed with gas chromatography coupled to mass spectrometry (GC-MS) with a Bruker Autosampler SCION SQ 436-GC. We used a non-polar Rtx-5 MS fused silica capillary column (0.32 I.D., 30 m long, 0.25  $\mu$ m thick film) and helium as the gas carrier. Volumes of 0.5  $\mu$ l were injected in a split-less mode with the injector temperature at 250 degreeCelsius. We used peak area as a measure of a compound's abundance relative to the concentration of dodecane in the sample, fixing the latter to 100%. Peaks that could not be quantified because of their low concentration were treated as missing data for a given extract. Since nothing is known about the sensitivity of chemical receptors in these species of butterflies, we refrained from excluding low-concentration peaks from our descriptions. Quantification values were log-transformed [ $\log_{10}(x+1)$ ] to avoid losing information when comparing peaks differing importantly in abundance, as described by [Lecocq et al., 2013].

### 3.2.10 Statistical analysis of chemical data

To allow for a general visual description of the chemical similarities and differences between species and sexes, we generated heatmap-like graphs on the corrected concentration values of the distinct peaks in the chemical blends of each tissue separately, using *Matlab R2012a Student Version*. Since chemical bouquets may be perceived as composite signals, we performed multivariate analysis on the concentration of all compounds present in each tissue separately, where each compound is treated as a variable. We analysed the data of both species and sexes together on the one hand, and separating by sex on the other hand. Statistical analyses were performed in *R* 3.0.2 ([www.r-project.org](http://www.r-project.org)). We visualised the compositional similarity between the chemical cocktails by performing non-supervised non-metric multidimensional scaling (NMDS) ordination implemented in the *vegan*, *ecodist*, *MASS* and *ade4* packages in *R*. Prior to NMDS, the Bray-Curtis distance matrix of the chemical composition was computed. The appropriate number of dimensions was chosen for each dataset by plotting a “scree diagram” (stress versus number of dimensions). The significance of chemical differences between categories (i.e. species and sex) was assessed by a discriminant analysis on the Bray-Curtis similarity matrix, using a permutational multivariate analysis of variance (perMANOVA; [Anderson, 2001]) implemented in the *vegan* package, with 1000 permutations. Whenever the categories were significantly different, partial least squares discriminant analysis (PLS-DA) was applied to the dataset in order to identify variables (i.e. compounds) determining the separation of groups and thus being most diagnostic of a given category. *R* packages *mixOmics*



and *DiscriMiner* were used for this purpose. We also estimated the *indval* index, which gives the probability of association between a compound and a given category. More precisely, we computed indicator values that suggest which compounds are indicative of a given category. For this analysis we followed [Heuskin et al., 2014], and used the *indval* function in the R package *labdsv*. Finally, a comparison of the per-compound concentration was performed by using a Kruskal-Wallis test coupled with a pairwise Wilcoxon rank-sum test, which *p*-values were corrected for multiple testing using the Bonferroni correction. So far, only major peaks of interest were identified by comparison of mass spectra (within the NIST library) and retention indices with those of authentic reference samples. Retention indices (RIs) were calculated relative to the retention times of a mix of alkane standards (C10-C26). Other peaks of interest remain to be identified to allow performing biological assays to assess their role in species recognition between *H. hecale* and *H. ismenius*.

### 3.3 Results

#### 3.3.1 Structure of the genomic differentiation between *H. hecale* and *H. ismenius*

For the three datasets analysed, we found an inferred optimum of clusters of 2, suggesting strong genetic differentiation between *H. hecale* and *H. ismenius* (Figure 3.2A). A high genome-wide average *Fst* value ( $0.37 \pm 0.11$ ) was found between the two species. This value contrasts with the *Fst* calculated between the parapatric races of any of the studied species, which was close to zero (Figure 3.2B). Comparing the two *H. hecale* subspecies yielded a higher *Fst* value ( $0.02 \pm 0.07$ ) than comparing the *H. ismenius* subspecies ( $-0.02 \pm 0.06$ ). The values of *Fst* are heterogeneous along the genome (Figure 3.3). The position of the highest peaks of differentiation is conserved among the two performed comparisons, and some of them are located around the colour genes (shown by vertical dashed lines in Figure 3.3). However, the peaks of differentiation are correlated with troughs of diversity throughout the genome (Figure 3.3). The overall nucleotide diversity is very low in these species ( $Pi = 0.007$  in *H. hecale* and  $Pi = 0.003$  in *H. ismenius*; see Figures 3.3 and 3.12A), thus the estimated genome-wide *Fst* values might also be inflated. However, we estimated a high value of genome-wide  $d_{XY}$  (an absolute measure of divergence) that averaged  $0.04 \pm 0.01$ , based on a sample of scaffolds corresponding to  $\sim 1/3$  of the genome.

#### 3.3.2 Signals of higher interspecific genetic admixture between the mimetic pair than between the non mimetic pair were not consistently found

We tested whether *H. hecale melicerta* and *H. ismenius bouletti* exchange genes more frequently than *H. h. zuleika* and *H. i. clarescens*, given the wing pattern resemblance existing between the former pair but absent between the latter pair. We did not find strong evidences supporting this hypothesis (see Figure 3.2A described above and results below). However, a slightly lower interspecific *Fst* distribution was observed between the mimetic pair of races than between the non-mimetic pair (Figure 3.2B). Divergence between the co-mimics is lower ( $0.432 \pm 0.098$ ) than between the non mimetic races ( $0.463 \pm 0.098$ ) (two-sample Z-test:  $Z = -47.052$ ,  $P < 2.2 \times 10^{-16}$ ; Kolmogorov-Smirnov test:  $D = 0.163$ ,  $P < 2.2 \times 10^{-16}$ ; Figure 3.2B). The *Fst* statistic is dependent on the allele frequency at a locus ([Jakobsson et al.,

2013]). We observe differences in the distribution of allele frequencies (measured as genome-wide *Tajima's D*; Figure 3.12B) between the races of *H. hecale*. Namely, subspecies *H. h. zuleika* shows a higher value of *Tajima's D* than *H. h. melicerta* (even though the means do not differ significantly with a two-sample Z-test:  $Z = 0$ ,  $P = 1$ , the density plots are significantly distinct under a Kolmogorov-Smirnov test:  $D = 0.083$ ,  $P < 2.2 \times 10^{-16}$ ). Thus, although the estimated *Tajima's D* values are close to zero (Figure 3.12B), the slight difference observed on the distribution of allele frequencies between the *H. hecale* subspecies might be affecting the *Fst* values. The difference in divergence observed between the mimetic and non-mimetic pairs studied system is not found when estimating absolute divergence  $d_{XY}$  values (Figure 3.2C). We do not find such a trend neither when comparing the extent of gene flow between the two pairs of sympatric species with an ABBA-BABA test. The *Patterson's D* and *f* statistics calculated on the totality of the genome did not detect an excess of ABBA sites with respect to BABA sites in any of the two configurations in Figure 3.4A (Table 3.1). However, when looking into the distribution of both kinds of polymorphic sites tallied across the genome, local biases could be detected. When testing for allele sharing between the mimetic pair of races (configuration on the top of Figure 3.4A) one particular region showed a striking excess of ABBA over BABA sites (Figure 3.4B), close to and downstream the wing colour gene *optix* (Figure 3.4C). Associated to the ABBA to BABA excess around *optix*, we found a decrease in the value of divergence,  $d_{XY}$  (Figure 3.4C).

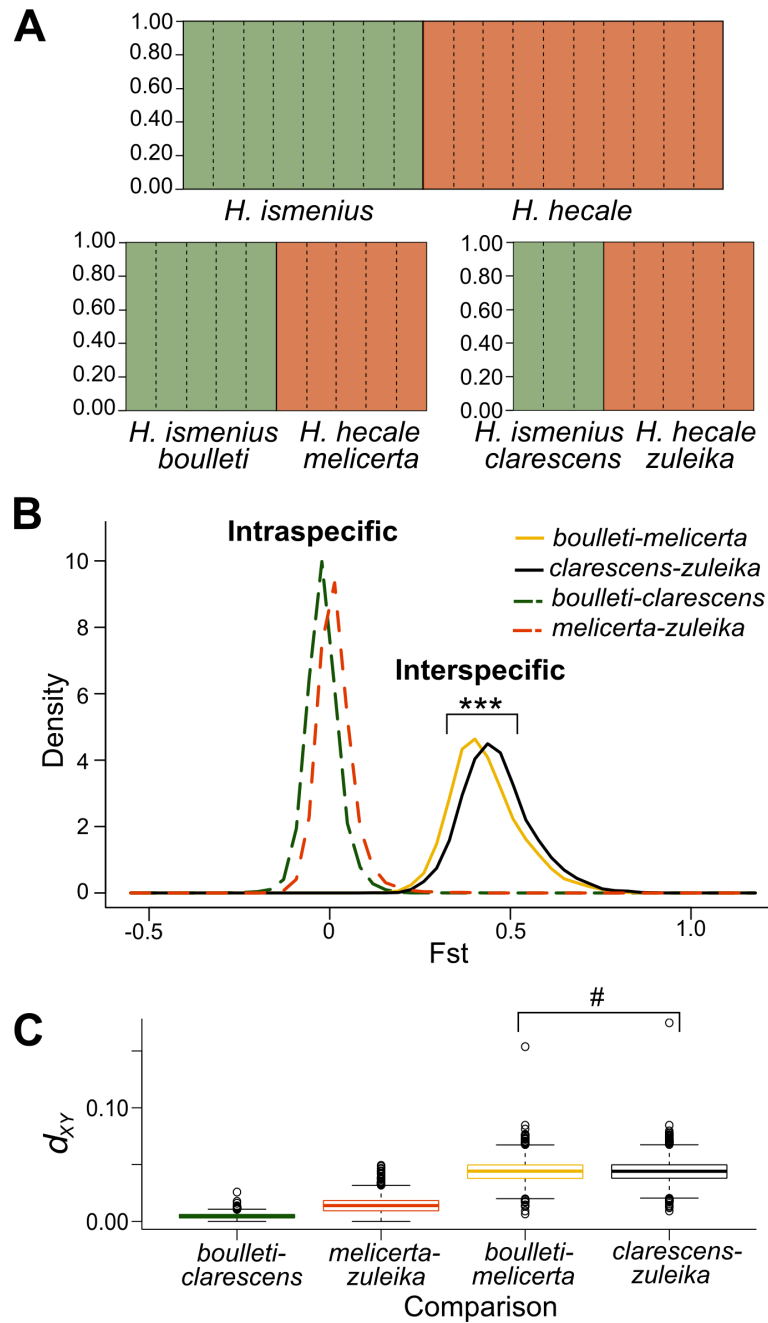


Figure 3.2 – Results of the genomic analyses used to test for gene flow between species *H. hecale* and *H. ismenius*. (A) Structure analysis showing complete genetic differentiation ( $K = 2$  and no admixture) between the two species as a whole (top), between the co-mimics *H. h. melicerta* and *H. i. boulleti* (bottom left) and between the non mimetic races *H. h. zuleika* and *H. i. clarescens* (bottom right). Different individuals are separated by dashed vertical lines. (B) Density plots of pairwise  $F_{st}$  values for non-overlapping 25-kb windows. The mimetic races (*H. h. melicerta*-*H. i. boulleti*) are less divergent than the non-mimetic races (*H. h. zuleika*-*H. i. clarescens*). The high  $F_{st}$  values for interspecific comparisons contrast with  $F_{st}$  values close to zero for the intraspecific comparisons. The *H. hecale* comparisons show slightly higher  $F_{st}$  values than the *H. ismenius* comparisons. (C) Boxplots of the average pairwise nucleotide divergence ( $d_{xy}$ ) for non-overlapping 25-kb windows. No difference is observed when contrasting the two pairs of sympatric races. Again, divergence is more important between the *H. hecale* races than between the *H. ismenius* subspecies. Significance codes: \*\*\* $p < 0.001$ , # $p > 0.05$ .

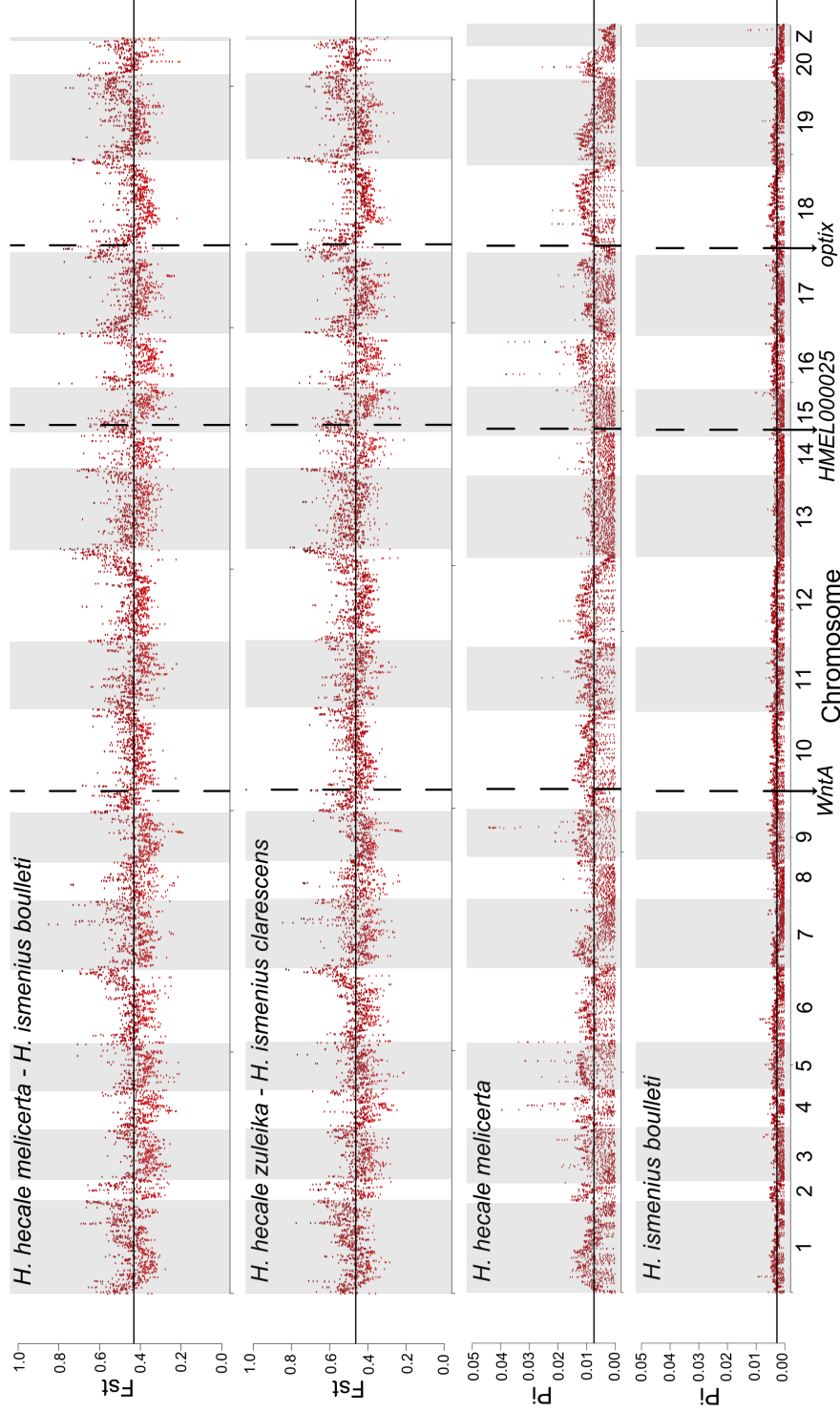


Figure 3.3 – **Genome-wide genetic differentiation ( $F_{st}$ ) between *H. hecale* and *H. ismenius* and nucleotide diversity ( $\pi$  or  $\pi$ ) of the mimetic sympatric races.** Statistics were calculated for non-overlapping 25-kb windows. The comparison between the mimetic races (first plot on top) and non-mimetic races (second plot from top to bottom) shows a highly similar profile of differentiation. Nucleotide diversity is plotted for races *H. h. melicerta* and *H. H. i. bouletti* only, since subspecies *H. h. zuleika* and *H. i. clarescens* display an almost identical profile. Chromosomes are indicated on the  $X$ -axis with alternating gray and white vertical bands. The genomic threshold for a given plotted statistic is represented as a horizontal black line. The position of three identified wing colour genes, *WntA*, the candidate *HMEL000025* and *optix*, on chromosomes 10, 15 and 18, respectively, is indicated by vertical dashed lines.

Table 3.1 – Genome-wide statistics derived from ABBA-BABA tests

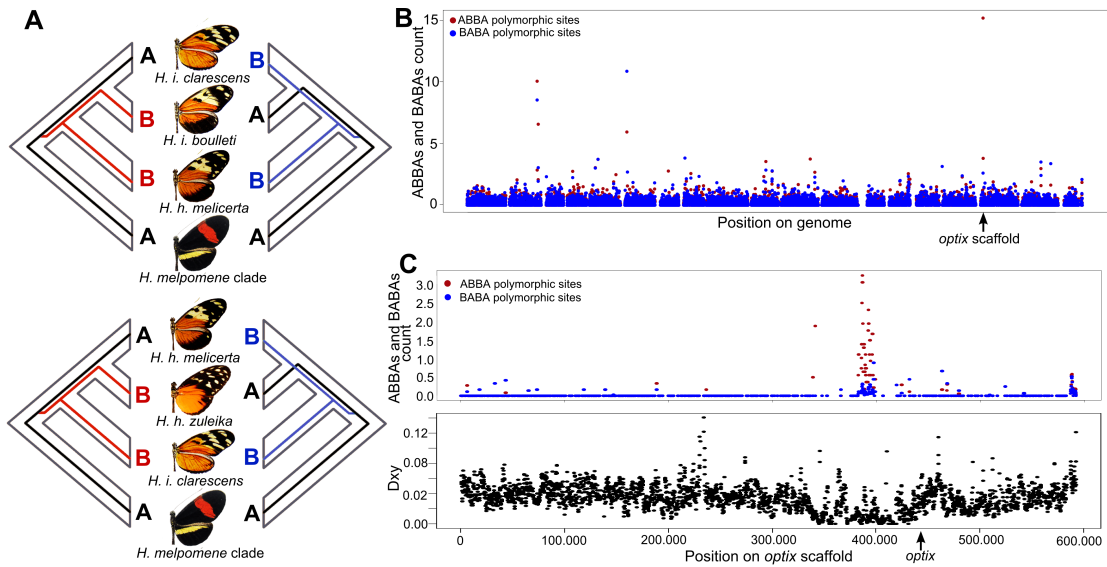
Pa <sup>a</sup>	P2 <sup>a</sup>	P3 <sup>a</sup>	Patterson's D ± standard error	Z-score <sup>b</sup>	p-value <sup>b</sup>	f statistic (%) ± standard error
<i>H. ismenius clarescens</i>	<i>H. ismenius bouletti</i>	<i>H. hecale melicerta</i>	-0,004 ± 0,005	-0,735	0,462	-0,010 ± 0,014
<i>H. hecale melicerta</i>	<i>H. hecale zuleika</i>	<i>H. ismenius clarescens</i>	0,003 ± 0,002	1,494	0,135	0,039 ± 0,026
<i>H. hecale zuleika</i>	<i>H. hecale melicerta</i>	<i>H. ismenius bouletti</i>	-0,003 ± 0,002	-1,444	0,149	-0,038 ± 0,026
<i>H. ismenius bouletti</i>	<i>H. ismenius clarescens</i>	<i>H. hecale zuleika</i>	0,005 ± 0,006	0,891	0,373	0,012 ± 0,014

<sup>a</sup> P1, P2 and P3 represent the three populations used for the ABBA-BABA test (see Methods for details). The first and second rows correspond to the configurations mainly tested (upper and lower configurations in Figure 3.4A, respectively). The two lower rows are alternative ways to resolve the comparison between the mimetic and non-mimetic pair of sympatric species

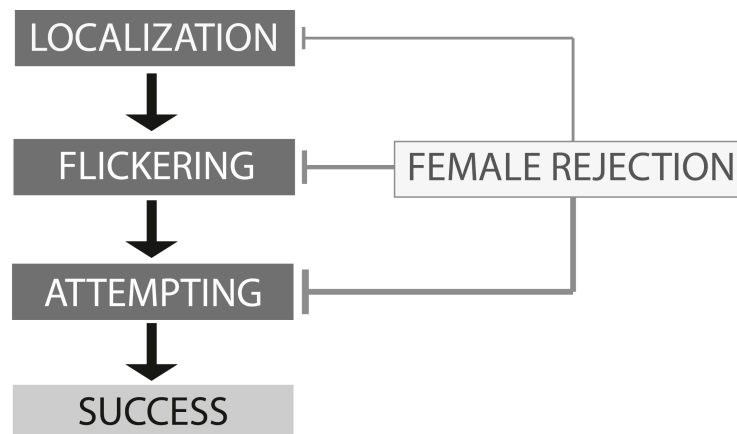
<sup>b</sup> Z-score and p-value

### 3.3.3 Description of the courtship behaviour of *Heliconius hecale*

*H. hecale* displays a relatively simple sequence of major behavioural units. Nevertheless, in accordance to the results of Crane, ([Crane, 1957]) for related species, rare or small behavioural events are quite variable, depending on the receptivity of the female. In captivity, the aerial phase (both sexes flying) is very short and simple or absent, in contrast to its complexity in some species in other groups of butterflies ([Tinbergen et al., 1942]; [Pliske, 1975]; [Brower, 1996]). The prolonged aerial-ground phase (male flying and female sitting) defines this species as being patrolling, according to [Scott, 1972]. The ground-ground phase (male and female on the surface) represents the last part of the sequence. Figure 3.5 shows five *main* events registered in the *H. hecale* courtship, three involving male behaviour (*localization*, *hovering* and *attempting*), female behaviour (*female rejection*) and both sexes behaviour (*success*). During *localization*, the male approaches the resting female. If the male is interested in the female, he hovers above her for several seconds (*hovering*) and lands besides the female to bend the tip of his abdomen towards hers (*attempting*). The female can be receptive to the male or she can reject him by adopting the mate-refusal posture ([Obara, 1964]) at any point of the male courtship. These major behavioural units are explained in detail in the ethogram in Table 3.3, where alternative names used by different authors are also included. Here we mainly use the nomenclature that Nieberding et al. ([Nieberding et al., 2008]) used in their analysis of the courtship behaviour of *Bicyclus anynana* (Nymphalidae: Satyrinae). Table 3.3 shows the *minor* events recorded (short behavioural units) and *atypical* events (events that happen occasionally) observed in the male behaviour. Importantly, some *atypical* events started by males and implicating direct contact between sensorial organs (antennae and legs with wings and head) were registered between males and females. Table 3.3 contains a description of the *main* and *minor* events that involve females only or both males and females simultaneously. The limited number of videos did not allow us to determine the importance of the *minor* and *atypical* steps in courtship, but they are mentioned here since they are consistently supported by previous studies. Moreover, Table 3.3 comprises a partial list of butterfly species for which the same behavioural units have been reported.



**Figure 3.4 – Four-clade ABBA-BABA test of introgression.** (A) Alternative analysed configurations tested for differential gene flow between the two parts of our study system. Namely, introgression between the co-mimics *H. h. melicerta* and *H. i. bouletti* and among the non mimetic races *H. h. zuleika* and *H. i. clarescens* are tested for in the configurations on top and bottom, respectively. (B) Count of ABBA and BABA polymorphic sites (on the configuration on the top) across the genome in non-overlapping 25-kb windows. A striking bias towards an excess of ABBAs is observed on the *optix* scaffold. (C) Number of sites with ABBA and BABA patterns and average pairwise nucleotide divergence ( $d_{XY}$ ) across the *optix* scaffold. The location of gene *optix* is shown by an arrow. The plotted values were calculated per 1 kb windows, moving in increments of 200bp.



**Figure 3.5 – Synthetic diagram of the main steps of courting and mating behaviour of *Heliconius hecale*.** Successful and unsuccessful courting sequences are summarized here. Steps in dark grey, light grey and white boxes represent male-related, female-related and male and female-related behaviours, respectively. *Female rejection* can happen at any point of the series but it is more frequent at later stages.

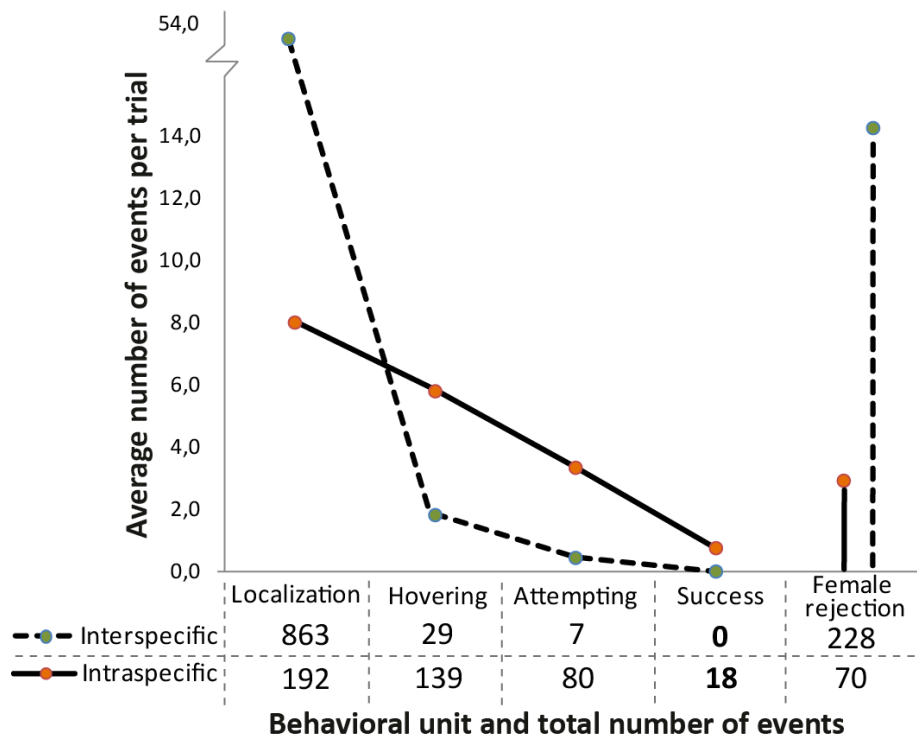


Figure 3.6 – **Comparison of interspecific and intraspecific male-female encounters.** The number of events in each *main* step of courtship is shown for interspecific trials (dashed line; *H. hecale melicerta* and *H. ismenius bouletti* tested together,  $N = 16$ ) and intraspecific trials (solid line; only *H. hecale* tested,  $N = 24$ ). The table and the Y-axis show the total number of events at each behavioral unit (registered across the totality of the trials) and the number of events at each step averaged by the number of trials, respectively.

3.3.4 Interspecific recognition between co-mimics primarily involves short range cues

None of the interspecific male-female encounters led to mating (0 out of 16 trials; table below Figure 3.6), whereas in 75% of the intraspecific control trials, mating occurred within the first 15 minutes (18 out of 24 trials; table below Figure 3.6). This result highlights the fact that the used experimental conditions were suitable for sexual behaviour leading to mating. We observed highly distinct trends in the courtship sequence between the two types of encounters tested (Figure 3.6). For intraspecific trials, we noticed a gradual transition between the *main* steps of courtship. More precisely, 72.4% of the *localization* events were followed by *hovering* events, and 57.6% of the *hovering* events elicited *attempting* behaviours in males. Finally, 22.5% of *attempting* events led to mating. In contrast, in interspecific experiments, males repeatedly approached the female (high number of *localization* events), but very few of these events were followed by active courtship (i.e. *hovering* and *attempting*). This indicates that heterospecific confusion is common based on long range cues, but species discrimination is very efficient at short range. Figure 3.6 also shows that the average number of *female rejection* events per trial, which encompass refusal at any point of courtship by males, is higher in inter than in intraspecific experiments. The totality of the interspecific *attempting* events ( $N = 7$ ), which were exerted consistently by at least four distinct males across the three different trials, were strongly rejected by the tested female (results not shown).

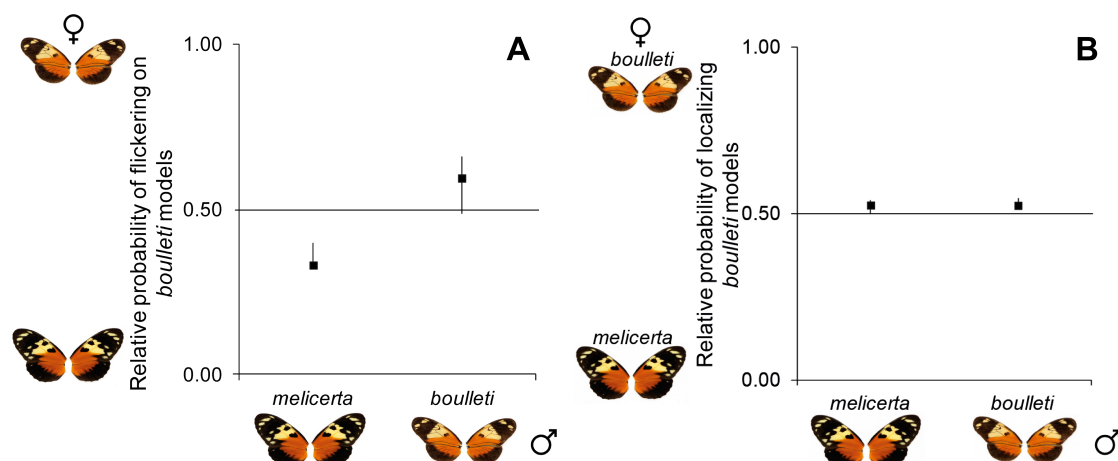


Figure 3.7 – Male choice of *H. hecale melicerta* and *H. ismenius bouletti* between female wings models of both species. Probabilities of approaching (A) shortly (*localization*) and (B) sustainedly (*hovering*) the *H. ismenius bouletti* models, where 1 means a complete choice of *bouletti* models and 0 a preference for *melicerta* models. A total of 42 *H. h. melicerta* males and 35 *H. i. bouletti* males were tested. Error bars show support limits equivalent to 95% confidence intervals.

### 3.3.5 Wing cues contribute to species recognition by males in the short range

Males approached the homo and heterospecific female wing models indistinctly (*localization* events;  $G = 0.002$ ,  $P = 0.96$ ; Figure 3.7A). In contrast, males of both species tended to hover for more than 3 seconds (*hovering*) over conspecific than heterospecific models ( $G = 21.323$ ,  $P < 0.001$ ; Figure 3.7B). In particular, *H. hecale melicerta* males hovered more readily than *H. i. bouletti* males over conspecific models (see confidence intervals in relation to the line of no-choice at 0.5; Figure 3.7B), showing differences in their general propensity to court.

### 3.3.6 Mate discrimination by *H. hecale* females does not rest on male wing colour cues

Matings to the *sham* versus the *treated* male were equal, suggesting that *H. hecale* females did not discriminate differently-coloured males ( $n_{sham} = 10$ ,  $n_{treated} = 10$ , Figure 3.8A). Females rejected one or the other male equally across all the experiments, but rejection of colour-modified males was more common ( $n_{sham} = 59$ ,  $n_{treated} = 78$ ,  $\chi^2 = 2.635$ ,  $P = 0.10$ ; Figure 3.8A). However, *H. hecale zuleika* females ( $N = 12$ ) rejected treated males more often than *H. hecale melicerta* females ( $n_{sham} = 10$ ,  $n_{treated} = 22$ ,  $\chi^2 = 4.500$ ,  $P = 0.04$ ).

### 3.3.7 *H. hecale melicerta* and *H. ismenius bouletti* differ in their chemical blends

Despite the high variance in the abundance of the peaks in the chemical cocktails both between and within sexes and species, we were able to determine which peaks differed consistently between the analysed categories (see Figures 3.9A, 3.10A and 3.11A). The most important compositional differences were found for claspers (C) and overlapping wings (OW) of males (Figures 3.9A and 3.9B, 3.10A and 3.10B). Non-supervised multivariate analyses allowed



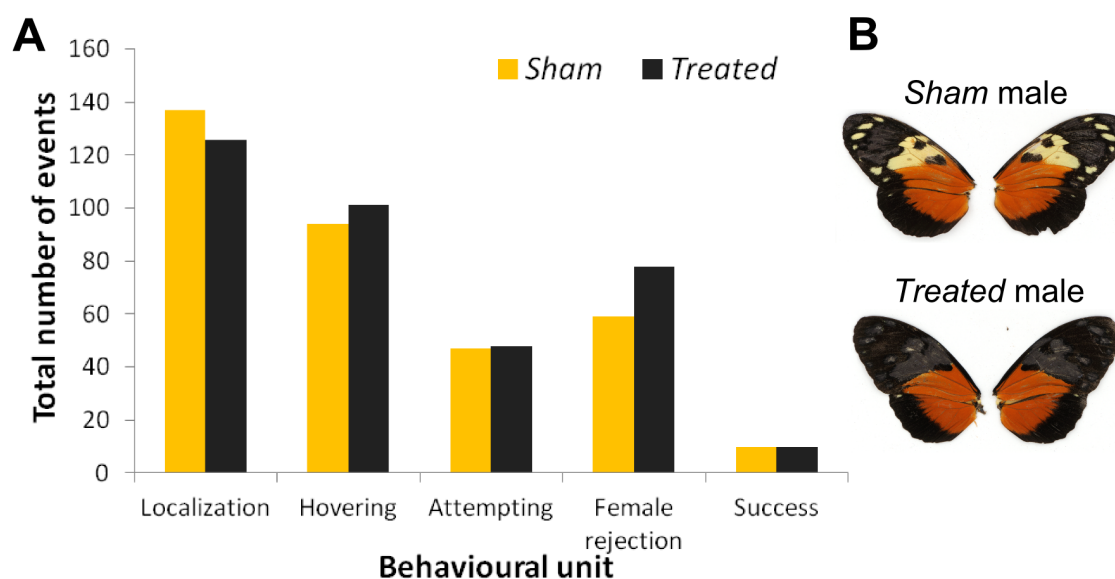


Figure 3.8 – Summary of the behavioural experiments to test for female choice in *H. hecale* based on visual cues. (A) Total events recorded in each of the *main* courtship steps during 3 hours long trials. (B) Each *H. hecale* female could chose between two males of its own species and race. One of them was *treated* by modifying its color pattern (yellow patterns on both sides of the forewings blacked with *Sharpie*). The other male (called *sham*) was black-painted on the black regions of the forewing. Only *H. hecale melicerta* male wings are shown here.

a good discrimination between males of the two species, and between males and females, but could not distinguish between females of the two species (perMANOVA for glands, both sexes together:  $F = 14.4$ ,  $df = 24$ ,  $p = 0.000$ ; perMANOVA for wings, both sexes together:  $F = 7.6$ ,  $df = 17$ ,  $p = 0.000$ ; NMDS plots, Figures 3.9B and 3.10B). The composition of claspers and wings were significantly different between males (perMANOVA for C:  $F = 19.6$ ,  $df = 11$ ,  $p = 0.002$ ; for W:  $F = 18.7$ ,  $df = 10$ ,  $p = 0.003$ ) but not between females (perMANOVA for G:  $F = 1.1$ ,  $df = 12$ ,  $p = 0.322$ ; for W:  $F = 0.7$ ,  $df = 6$ ,  $p = 0.845$ ). Comparing males of the two species, intriguingly, showed that most of the variable compounds were present in *H. h. melicerta* and absent in *H. i. bouletti* tissues, or more concentrated in the former than in the latter (Figures 3.9A and 3.10A, Tables 3.4 and 3.5).

The definitive identification of the compounds is not available yet; therefore, a unique numeric label for each compound is used throughout the text and is included in the first column in Table 3.4. Nevertheless, in the heatmaps for the different tissues (Figures 3.9A, 3.10A and 3.11A), a linear numeration for the different compounds is used, whose correspondence to the unique compound labels is shown in Table 3.4.

In *H. h. melicerta* claspers,  $\beta$ -ocimene (compound number 3, also number 3 in Figure 3.9A; see Table 3.4) dominates the cocktail ( $\sim 46.9\%$  of total), consistently with the findings of Estrada et al. ([Estrada et al., 2011]). However, in contrast with Estrada et al. ([Estrada et al., 2011]), this compound is present in *H. i. bouletti* males, although just as a trace. Contrary to the expectations, we did not find  $\alpha$ -ionone, previously found to be specific to *H. ismenius* claspers ([Estrada et al., 2011]). We ignore whether this is related to the subspecies of *H. ismenius* analysed or to differences in the conditions used among studies. Only one compound of low concentration is exclusive to *H. i. bouletti* males (compound number 35, but number 33 in Figure 3.9A; see Table 3.4) and 3 compounds are more concentrated in *bouletti*'s claspers than

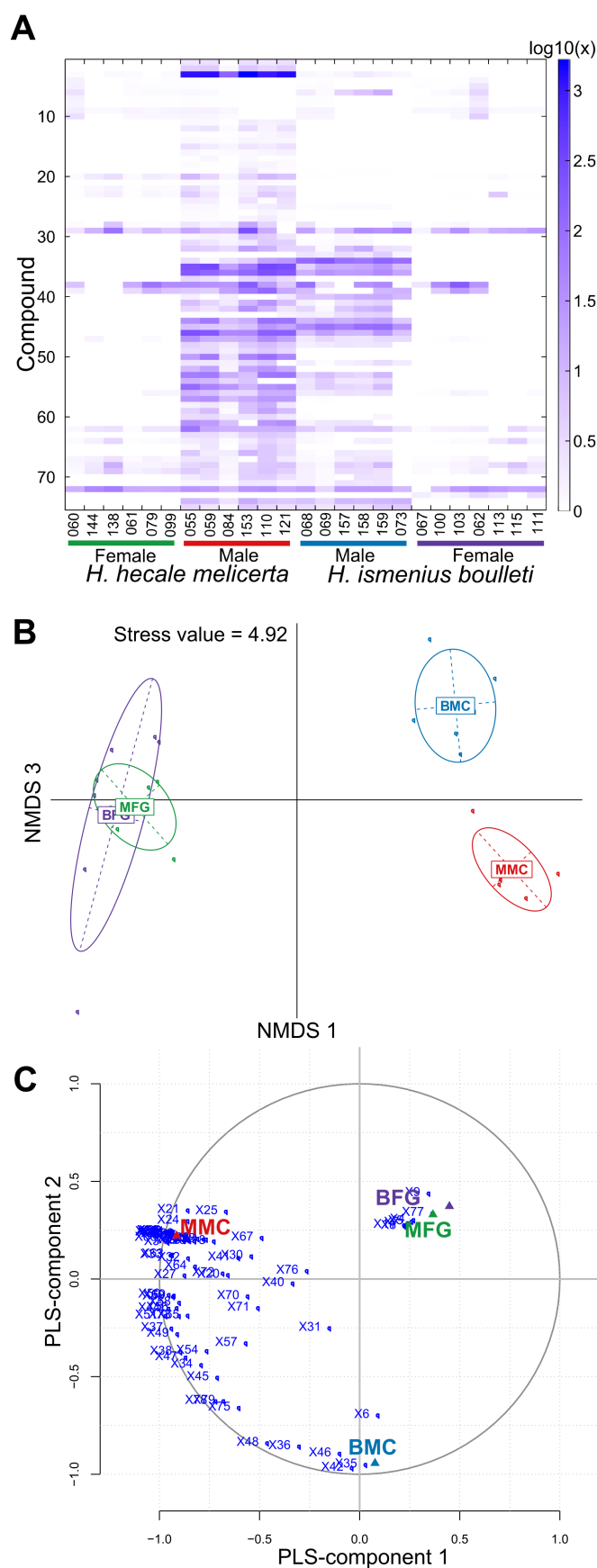


Figure 3.9 – Analysis of the chemical composition of the abdominal glands (C and G) of *H. hecale melicerta* and *H. ismenius bouletti*. (A) Heatmap showing the log-transformed  $[\log_{10}(x+1)]$  relative concentration of each compound (horizontal bands) for males and females of both species (columns). Single individuals are labelled under each column. (B) Non-metric multidimensional scaling (NMDS, stress equal to 4.92) ordination of the chemical cocktails of *H. h. melicerta* males (MMC) and females (MFG), and *H. i. bouletti* males (BMC) and females (BFG). Each dot represents a sample of any of these categories. The discrimination of the groups is significant (perMANOVA:  $F = 14.4$ ,  $df = 24$ ,  $p = 0.000$ ). (C) Partial least squares discriminant analysis (PLS-DA) based on comparisons among chemical blends composition. A correlation circle is plotted between the compounds (variables) (blue dots and labels) and the components. The variables being indicative of the groups are strongly correlated with the groups being separated (close to the concentric circle of radius 1). To look for a correspondence between the labels of the compounds in A and C, see Table 3.4.

in *melicerta*'s claspers (peaks 36, 42 and 48, corresponding to 34, 40 and 45, respectively, in Figure 3.9A; see Table 3.4). 42 compounds differ significantly in their concentration between the male genitalia of the two species, most being exclusive or more abundant in *melicerta* males (Table 3.4). Most of these variables are located close to the concentric circle of radius 1 in the correlation circle shown in Figure 3.9C; thus, they show a strong correlation with the axis separating *melicerta* from *boulleti*. Moreover, the *indval* index pinpointed 44 compounds (largely the same than the 42 mentioned above) significantly associated with either species, but especially with *melicerta* males (Table 3.4). A total of 29 compounds were found as male-specific (Table 3.4).

Abdominal glands (AG) of females showed only a few small peaks (Figure 3.9A and Table 3.4). Only two low concentration compounds differed between species, one being present only in *boulleti* (compound number 77, but 73 in Figure 3.9A; see Table 3.4). This compound is also exclusive to the cuticle of *boulleti* females (Table 3.6, but is also present on female wings of both species (Table 3.5). The second compound is exclusive to *H. h. melicerta* female glands (and male claspers of the same species) (compound number 24, but 22 in Figure 3.9A; see Table 3.4). Nevertheless, neither compound is significantly different between species (Wilcoxon rank-sum test) or indicative of species status (*indval* index; Table 3.5), possibly due to their low general abundance.

In wing extracts, two compounds dominate the cocktails from males *H. h. melicerta* and are absent or only present as traces in males *H. i. boulleti* (Figure 3.10A and Table 3.5). One is the alkane heneicosane ( $\sim 57.1\%$  of total; number 38, but 15 in Figure 3.10A) and the other is presumed to be a ketone, 2-pentadecanone ( $\sim 7.87\%$  of total; number 22, but 10 in Figure 3.10A). Additionally, seven other compounds are also found in higher abundance in *melicerta* males (Table 3.5) and contribute to the separation of the *melicerta* and *boulleti* groups (Figure 3.10B and 3.10C). The *indval* index identified six compounds as indicative of *melicerta* males (Table 3.5). Only two wing compounds are male-specific and one low-concentration compound is female-specific (see Table 3.5). Female wing extracts showed fewer compounds of small abundance, which were highly conserved between species. Only one compound was present at very low concentration in *melicerta* and absent in *boulleti* wings, but, possibly due to its low abundance, this difference is not significant (compound number 79, but 33 in Figure 3.10A; see Table 3.5).

Cuticle extracts (B) have fewer compounds than the other tissues, and are highly conserved between species and sexes (Figure 3.11A; Table 3.6). Although the NMDS plots indicate that *melicerta* males segregate from the rest of the categories (Figure 3.11B), this clustering is not significant (perMANOVA for all categories together:  $F = 1.6$ ,  $df = 18$ ,  $p = 0.132$ ; perMANOVA for males only:  $F = 1.9$ ,  $df = 8$ ,  $p = 0.140$ ; perMANOVA for females only:  $F = 2.3$ ,  $df = 9$ ,  $p = 0.008$ ).

### 3.4 Discussion

#### 3.4.1 The closely-related species *Heliconius hecale* and *H. ismenius* show strong genome-wide differentiation

Here we have used whole-genome sequencing data to examine gene flow and differentiation between two species, *H. hecale* and *H. ismenius*, belonging to the silvaniform radiation of *Heliconius*, and to compare to other *Heliconius* clades. Estimates of genome-wide divergence statistics show weak evidence of interspecific hybridisation. A genomic admixture analysis using *Structure* showed complete genetic differentiation between the two silvaniform species.

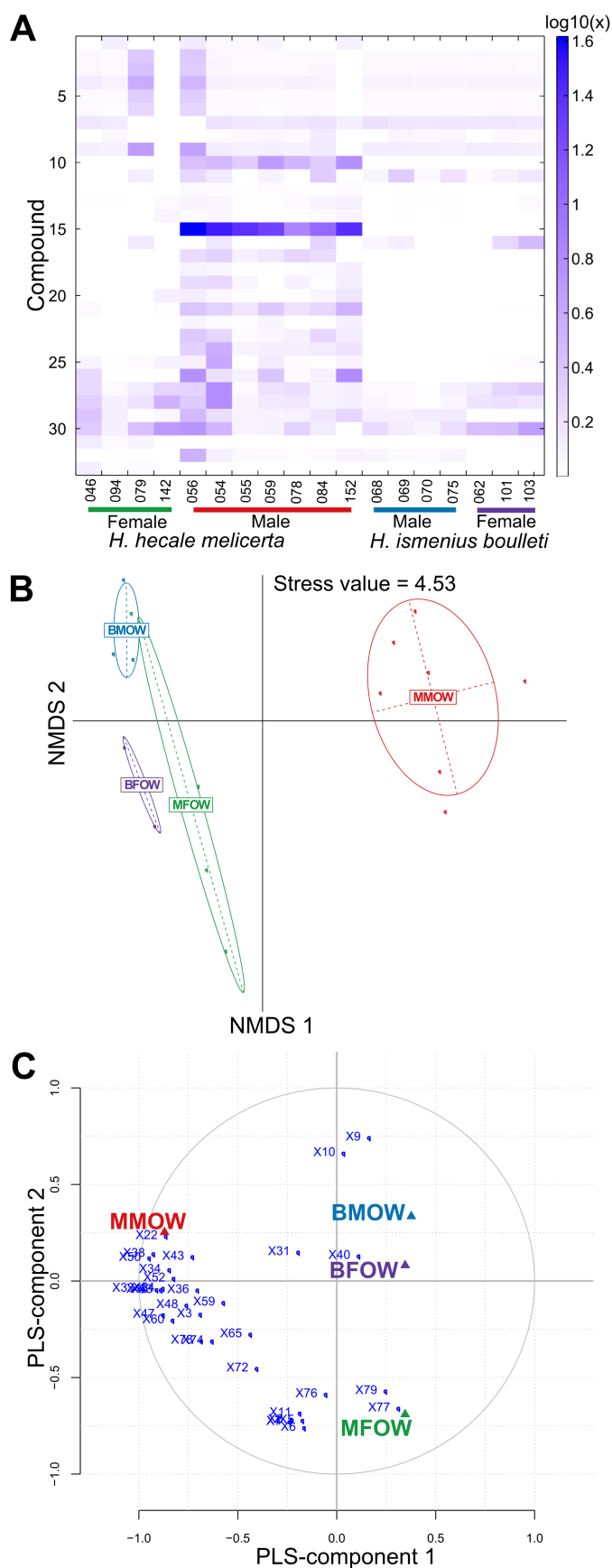
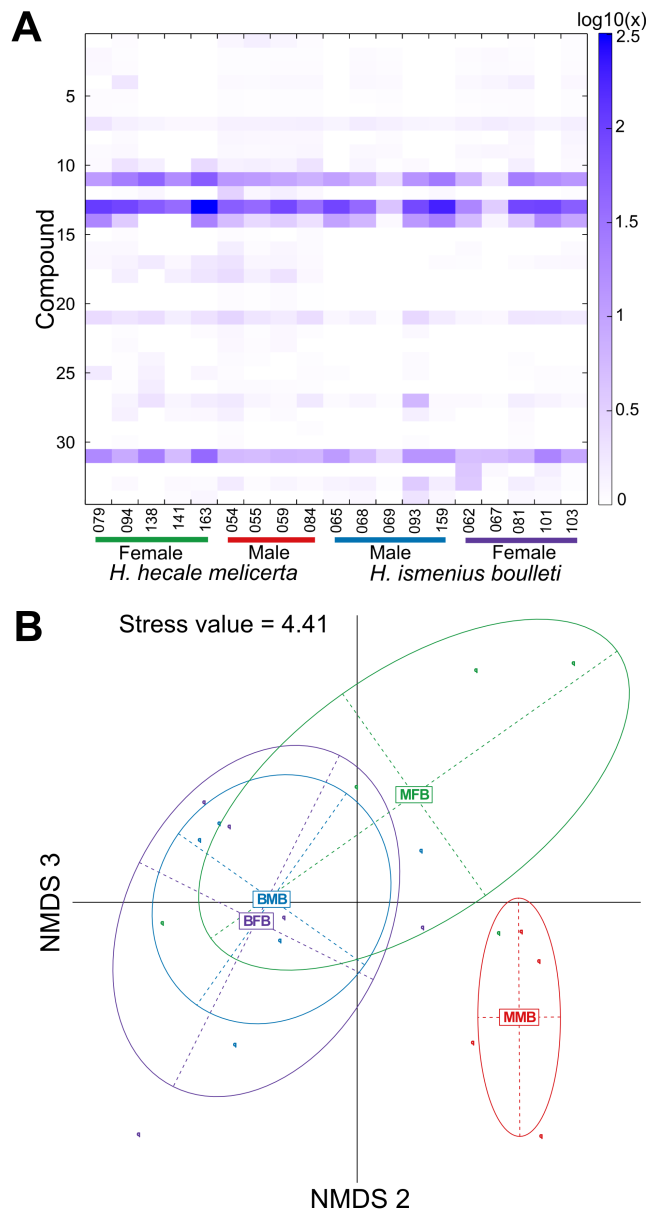


Figure 3.10 – Analysis of the chemical composition of the wings (OW) of *H. hecale melicerta* and *H. ismenius bouletti*. (A) Heatmap showing the log-transformed  $[\log_{10}(x + 1)]$  relative concentration of each compound (horizontal bands) for males and females of both species (columns). Single individuals are labelled under each column. (B) Non-metric multidimensional scaling (NMDS, stress equal to 4.53) ordination of the chemical cocktails of *H. h. melicerta* males (MMOW) and females (MFOW), and *H. i. bouletti* males (BMOW) and females (BFOW). Each dot represents a sample of any of these categories. The discrimination of the groups is significant (perMANOVA:  $F = 7.6$ ,  $df = 17$ ,  $p = 0.000$ ). (C) Partial least squares discriminant analysis (PLS-DA) based on comparisons among chemical blends composition. A correlation circle is plotted between the compounds (variables) (blue dots and labels) and the components. The variables being indicative of the groups are strongly correlated with the groups being separated (close to the concentric circle of radius 1). To look for a correspondence between the labels of the compounds in A and C, see Table 3.5.

Figure 3.11 – Analysis of the chemical composition of the cuticle in the basis of the wings (B) of *H. hecale melicerta* and *H. ismenius bouletti*. (A) Heatmap showing the log-transformed  $[\log_{10}(x+1)]$  relative concentration of each compound (horizontal bands) for males and females of both species (columns). Single individuals are labelled under each column. (B) Non-metric multidimensional scaling (NMDS, stress equal to 4.41) ordination of the chemical cocktails of *H. h. melicerta* males (MMOW) and females (MFOW), and *H. i. bouletti* males (BMOW) and females (BFOW). Each dot represents a sample of any of these categories. The discrimination of the groups is not significant (perMANOVA:  $F = 1.6$ ,  $df = 18$ ,  $p = 0.132$ ). To look for a correspondence between the labels of the compounds in A and the right label of the compounds, see Table 3.6.



Although this analysis might perform better on hyper variable markers like microsatellites, which allow the detection of fine genetic structure variations, a study based on genome-wide RAD-sequencing data found some extent of genetic admixture between sympatric populations of *H. melpomene* and *H. timareta*, and *H. melpomene* and *H. cydno*, even on a limited number of samples ([Nadeau et al., 2013]). This suggests that our analysis should be suitable to detect important levels of gene flow when they exist, which is apparently not the case in the system we investigate. The high genome-wide differentiation between the sympatric species studied here is more important than between coexisting species belonging to the melpomene clade ([Martin et al., 2013]; [Kronforst et al., 2013]), but similar to the estimates between allopatric species in that clade ([Martin et al., 2013]). These high levels of divergence were interpreted as an absence of detectable gene flow between allopatric species *H. cydno* and *H. timareta* in another study ([Nadeau et al., 2013]). This value of *Fst* might be inflated to some extent by the low nucleotide diversity of our study species. However, the estimation of a measure of absolute divergence ( $d_{XY}$ , which is independent of genetic variability) on a subsample of the genome yielded a value almost twice as that estimated between coexisting *H. melpomene* and *H. cydno* in Costa Rica ([Kronforst et al., 2013]). Therefore, any used genetic differentiation statistic indicates stronger genome-wide divergence between the sympatric species *H. hecale* and *H. ismenius* than between other coexisting closely-related *Heliconius* species explored so far, suggesting that strong barriers to gene flow exist between these silvaniform species, which trigger strong genomic structure.

### 3.4.2 Shared wing patterns do not necessarily increase gene flow

Genomic sequencing now allows estimating the admixed portion of the genome and the patterns of gene flow between pairs of diverging taxa. Studying three species in the melpomene clade, [Martin et al., 2013] revealed increased shared variation in sympatry, with up to 40% of the genome being admixed between coexisting species. Here, we also used the so-called ABBA-BABA test ([Green et al., 2010]; [Durand et al., 2011]; [Martin et al., 2013]) to look for evidence of gene exchange between co-mimics *H. hecale melicerta* and *H. ismenius bouletti* relative to the admixture between the non-mimetic pair *H. h. zuleika* and *H. i. clarescens*. By contrast to the melpomene clade ([Martin et al., 2013]), our comparisons showed no excess of ABBA sites relative to BABA sites in any comparison, using either *Patterson's D* or the *f* statistic, suggesting weak and undetected or absent allele sharing in sympatry. These results, as well as the equality of the absolute divergence ( $d_{XY}$ ) in the two parts of our study system, appear to reject the hypothesis of an increased rate of hybridisation and resulting gene flow when species share the same mimicry ring.

Nonetheless, the *Fst* distribution was shifted towards lower values for the pair of mimetic races compared to the non-mimetic pair. Although this trend is expected if species exchange more genes in the area where they are mimetic, *Fst* is not a straightforward indicator of gene flow as it is sensitive to demographic differences and overall levels of polymorphism. For instance, the *Fst* statistic depends upon the allele frequency at a locus ([Jakobsson et al., 2013]). *H. hecale zuleika* shows a higher value of *Tajima's D* than *H. hecale melicerta*, although both values are close to zero. We ignore so far whether a biological or a technical factor could be causing both the slight difference in *Tajima's D* observed between the *H. hecale* races and the shifted estimates of interspecific *Fst* values.

### 3.4.3 Adaptive introgression seems to have occurred at some point of species divergence

Altogether, our analyses do not argue for a significant effect of mimicry on allele sharing at the genome level, although adaptive introgression of some of the mimicry alleles themselves do show the importance of occasional hybridisation in the phenotypic evolution and ecology of those species. Namely, punctual events of introgressive hybridisation seem to have taken place at some point between diverging *H. hecale* and *H. ismenius* lineages. One particular genomic region, the “*optix* scaffold”, shows a striking excess of ABBA over BABA sites between the mimetic races. Interestingly, this region is close to the *cis*-regulatory region of the wing patterning gene *optix* which controls a variety of red elements on the wings of *H. melpomene*, *H. erato* and *H. cydno* ([Reed et al., 2011]; [Supple et al., 2013]; [Martin et al., 2014]). This region was recently shown to control shape variation of the black marginal band on the hindwing of both *H. hecale* and *H. ismenius* ([Huber et al., 2015]). In particular, mimetic races *H. hecale melicerta* and *H. ismenius bouletti* bear a recessive allele for this locus, associated with a wide black marginal band surrounded by orange. Since shared derived mutations are enriched among this pair of races in sympatry relative to the comparison of non-coexisting races of the two species, similar alleles in sympatric races may reveal adaptive introgression between them at the *optix* locus. Furthermore, we found a decrease in the estimate of absolute divergence,  $d_{XY}$ , associated to the ABBA to BABA excess around *optix*. Such a decrease has been related to introgression between taxa with shared ancestry, and not to ancestral structure ([Smith and Kronforst, 2013]). Nonetheless, Martin et al. ([Martin et al., 2014]) found that a reduction in  $d_{XY}$  at Patterson’s *D* and *f* statistics outliers is not enough evidence of introgression, given that these outliers actually tend to cluster in regions of low  $d_{XY}$ . Overall, however, our results seem to be consistent with the findings of adaptive introgression in other pairs of *Heliconius* species for this colour-patterning region ([Pardo-Diaz et al., 2012]; [Heliconius Genome Consortium, 2012]; [Smith and Kronforst, 2013]), although with distinct alleles and distinct mimicry communities. Those results argue for a pervasive role of introgression as a powerful mechanism to achieve mimicry perfection in pairs of closely related species.

Other major effect wing patterning loci known to segregate in *H. hecale* and *H. ismenius* do not show evidence for introgression neither here nor in other species studied so far. In particular, we did not observe any evidence of allele sharing in the genomic scaffold containing colour gene *WntA* ([Martin et al., 2012]), which controls a phenotypic element that is similar between the mimetic pair and also among the non-mimetic pair ([Huber et al., 2015]). We wonder whether the evolutionary mechanisms which underlie phenotypic convergence at different loci reflect differences in selection pressures on pattern elements themselves or intrinsic genetic differences that make some regions of the genome more prone than others to be transferred across the species boundaries and selected to fixation in distinct lineages of the clade. Equally, it is unclear at this stage whether the phenotypic convergence at wing colour elements controlled by loci other than *optix* may be due to parallel evolution (*c.f.* [Stern, 2013]) leading to similar allelic effects in response to mimicry selection and/or to ancestral structure. These two scenarios are not mutually exclusive, and thus it is possible that a combination of both has taken place. Questions about the origin of wing pattern similarities between *H. h. melicerta* and *H. i. bouletti* extend over other similarly-looking races of these species (e.g. *H. hecale australis*, *H. ismenius metaphorus* and *H. i. tilleti*) and other silvaniform species (e.g. *H. ethilla ethilla* and *H. e. claudia*) displaying a leapfrog distribution across the Neo-tropics.

Overall, our results contrast with the findings of pervasive gene flow between taxa within the *H. melpomene* and *H. cydno* lineages ([Bull et al., 2006]; [Kronforst et al., 2006b]; [Martin

et al., 2013]), which nevertheless show distinct ecologies (microhabitat preference, larval host plant use, mimicry shift, assortative mating and partial hybrid sterility ([Jiggins et al., 2001]; [Naisbit et al., 2002])). Although natural hybridisation events usually occur between closely-related species in the genus, these can also happen between non-sister species for more than 3 My after speciation (according to molecular clock dating data; [Mallet et al., 2007])). Species *H. hecale* and *H. ismenius* have diverged over the last  $\sim 3$  My, which is  $\sim 1$  My longer than *H. melpomene* and *H. cydno* ([Kozak et al., 2015])). Hence, we report a case where 3 My seem to have been sufficient to erase most of the signal of admixture found in other comparisons. The occasional hybridisation that may occur between taxa at increasing phylogenetic distances could have genome-wide consequences only during the first 2 My which quickly disappear thereafter, presumably due to stochastic processes such as genetic drift, but also to the build-up and strengthening of barriers to gene flow. The accumulation of these barriers may be associated with an acceleration of the genome-wide divergence accumulation towards later stages of the speciation process ([Kronforst et al., 2013])). Although the relative importance of pre and post-zygotic isolation during the investigated speciation event remains unclear, current hybrids between *H. hecale melicerta* and *H. ismenius bouletti* are potentially viable. Fully-viable first generation hybrids between these taxa produced by hand-pairing looked very similar to *H. hecale melicerta* (non published data) suggesting they might enjoy mimicry protection and good adult survival. Although the fertility of *H. hecale*  $\times$  *H. ismenius* hybrids was not directly assessed, *H. cydno*  $\times$  *H. ismenius* crosses, which are more distantly related, do produce backcrosses in captivity ([Mallet et al., 2007])), suggesting that a similar situation could occur in our study species. In contrast, very strong pre-zygotic isolating mechanisms may exist which impede hybridisation and gene flow between *H. hecale* and *H. ismenius*. Our field observations during sampling indicate that these species have distinct preferences on microhabitat (data not shown), which may be somehow linked to larval host plant specialisation as well, consistent with previous work on ecological speciation occurring in sympatric host races of several herbivore species of insects ([Wood and Keese, 1990]; [Via et al., 2000]; [Emelianov et al., 2001]; [Thomas et al., 2003])). Namely, caterpillars of *H. hecale* feed on *Passiflora vitifolia*, whereas the ones of *H. ismenius* feed on *P. quadrangularis* and *P. ambigua*. Moreover, as we show here, signals putatively involved in species recognition, other than wing colour patterns, also seem to have diverged between these species (see below).

#### 3.4.4 A highly heterogeneous landscape of differentiation

A genome-wide sliding windows analysis of *Fst* between *H. hecale* and *H. ismenius* showed that genomic differentiation was heterogeneous, with some peaks of elevated divergence arising from an overall highly differentiated and variable genomic background. Such heterogeneity has been explained by certain models of speciation with gene flow which are based on the idea that the portions of the genome that confer reproductive isolation between diverging taxa exceed neutral background expectations ([Wu, 2001]; [Feder et al., 2012])). In contrast, the majority of the genome is believed to be affected by genetic drift and migration, and is exchanged freely between hybridising populations. However, *Fst* outliers seen late in speciation, like between our study species, do not necessarily correspond to the genomic regions influenced by divergent selection during early phases of ecological speciation ([Via, 2009])). Instead, these peaks could likely be the result of genetic drift. Genetic drift may have a stronger influence on genomic structure in cases of weak effective population size, which is presumably the case here, in particular in *H. ismenius*, given the extremely low nucleotide diversity of this species. Interestingly, however, the peaks of differentiation we found were in the same genomic regions



in the two parts of the range, supporting the idea that some of these regions may indeed underlie variation of traits involved in adaptation and/or reproductive isolation between the two species (irrespective of their wing patterns), and thus implicated in divergence and speciation (so-called “speciation genes”; [Wu, 2001]). Inflated  $F_{st}$  values can result from a low within-species genetic diversity ([Charlesworth, 1998]; [Cruickshank and Hahn, 2014]). Here, we indeed observed inversed patterns of genetic differentiation and nucleotide diversity, and, notably, the peak of  $F_{st}$  occurring close to *optix* may be attributed to the observed low local nucleotide diversity, since absolute divergence ( $d_{XY}$ ) is decreased at this locus. We ignore whether low diversity in punctual genomic regions respond to low mutation rates, to purifying selective forces, or to other causes. However, loci involved in speciation are expected to be submitted to disruptive then purifying selection with the fixation of species-specific alleles corresponding to fitness optima. Therefore, although misleading high observed differentiation values could result from low nucleotide diversity alone, this low diversity could be the signal of purifying selection on important loci for speciation, which may be associated to functional differences between the species, like “barrier loci” implicated in early stages of speciation, and whose role in divergence is strengthened as the speciation process proceeds.

Nucleotide diversity troughs do not only correspond to colour loci, and the heterogeneity in divergence levels might reflect ecological adaptations other than selection for mimicry. Belonging to the same mimicry ring does not hamper the maintenance of strong barriers to gene flow, which could cause the putative signals of purifying and disruptive selection seen. Such reproductive isolation factors could result from divergent ecological selection like microhabitat preference or larval host specialisation, and from divergent mate choice, among others. Although we suspect distinct ecological adaptations of these species based on preliminary observations (e.g. microhabitat preference and host plant specialisation), rigorous data will be needed to test the hypothesis of microhabitat adaptation and to understand its footsteps in the genome. Here, we provide insights into species-specific chemical recognition signals for assortative mating which may play an outstanding role in speciation.

### 3.4.5 Species recognition seems to take place at short range in males and females

To explore species recognition between *H. hecale melicerta* and *H. ismenius bouletti*, we started by describing courtship behaviour in the former species. Very few studies have characterised the courting behaviour in *Heliconius* and related genera, but the species appear to display conserved sexual behaviour ([Crane, 1957]). We characterised the courtship in *H. hecale* as a conserved sequence of five *main* (conspicuous) events and several *minor* (short) events, with the occasional addition of *atypical* behaviours which happen sporadically.

On the basis of the good understanding of the courtship sequence, we explored the nature and action range of the signals implicated in species recognition, in order to give insights into some of the barriers to hybridisation which maintain two mimetic species separate. We observed that males of both species are highly attracted to females and female wing models of the mimetic species, showing that male-operated heterospecific confusion can happen at a long range, which is consistent with the results of Estrada and Jiggins ([Estrada and Jiggins, 2008]) for *H. melpomene* and *H. erato*. Our results also confirm that visual signals are important signals eliciting male courtship in *Heliconius*. By contrast, *H. hecale* females do not seem to care very much about conspecific male coloration. Nevertheless, although females did not preferentially mate with the wing colour controls over the colour-modified males, they showed a trend to reject colour-modified males more often. Therefore, visual cues may be important at least to some

extent for female choice in this species, but other species-specific signals in the close range may override the effect of the differences in colour and turn the females undiscerning when exposed to homospecific males. A similar switch between signals (from visual to chemical cues) has been reported in Vane-Wright and Boppre ([Vane-Wright and Boppre, 1993]) for males of *Argynnis paphia*, in response to the release of a female pheromone. Although such a switch needs to be tested between distinctly-coloured races of *Heliconius*, our results give insights into the importance of short-range signals other than colour for female choice in *Heliconius* butterflies.

Contrary to what we found in *H. hecale*, male wing coloration has been reported as crucial for female choice in several species of butterflies, both for intra and interspecific recognition (e.g. [Robertson and Monteiro, 2005]). Importantly, Boyko ([Boyko, 2005]) found females to choose on the basis of wing colour in *H. charitonia*, which is to our knowledge, the only available data on this issue in the genus *Heliconius*. It has been proposed that recently emerged females are unable to reject males and that assortative mating is likely to be operated mainly by male choice ([Jiggins et al., 2001]). This is the reason why for the female choice experiments we used more than one day old females. However, contrary to the expectations of no-female choice in newly emerged females, we observed that during interspecific encounters, hours-old females drastically rejected the few mating attempts by heterospecific males. This indicates that even very young females exert a mate choice that has been mostly overlooked in the *Heliconius* literature. A determinant role of females opposed to male confusion in species recognition was also found in the close to identical species *Leptidea sinapis* and *Leptidea reali* ([Friberg et al., 2007]). Females (not males) are the choosing sex in all but sex-role-reversed species or species in which both sexes make substantial investments in the offspring. *Heliconius* males make little investment in their offspring, besides the transfer of a spermatophore during copulation. Overall, these observations suggest that species discrimination takes place in the short range for both sexes, with males and females decreasing and refusing, respectively, the number of heterospecific short range courtship events. Also, experiments using female wing models showed that males pursue into more advanced steps of courtship (into a closer contact with the female wings) preferentially on their own species models, a choice that is more accentuated in *H. hecale* than in *H. ismenius* males. These results indicate that female wings carry at least some species-related information that is used by males in the short range for species recognition. Several factors influence species recognition in other butterflies and might contribute to species discrimination in the close range between the mimetic species studied here. These encompass species-dependent courtship behaviours, wing, body or appendage morphology, wing motion, disparity in colour or UV reflectance, minute differences in colour pattern, and more importantly, chemical signals ([Rutowski, 1977a]; [Rutowski, 1977b]; [Vane-Wright and Boppre, 1993]; [Fordyce et al., 2002]). Among these, the latter have been shown to be more important than the rest in several species in view of their high evolvability ([Schulz et al., 1993]; [Nieberding et al., 2008]). Therefore, we focused on evaluating whether these cues differed between the studied species.

### 3.4.6 Chemical signals differ between the mimetic species, especially between males

To test the hypothesis that chemical differences between *H. h. melicerta* and *H. i. bouletti* exist both in males and in females, we compared chemical blends from distinct tissues. We found that while males have rich and complex chemical cocktails that strongly differ between species in the relative abundance of the compounds (significant group separation in the multivariate space), females of both species have chemical blends which are both poor in compounds and

similar in their composition. Although the identification of most compounds is still unachieved, the general description of the extracts from abdominal glands and wings puts in evidence a rich battery of compounds being apparently specific to males and particularly to *H. hecale melicerta* males. Few compounds are restricted to *H. i. bouletti* males or more concentrated in their tissues than in *H. h. melicerta*, but we ignore whether the scarcity of species-specific signals in *H. i. bouletti* males is biologically relevant or if some compounds may not be detected by our assay. Independently of this, the difference between the males of the two species is notable, particularly concerning the chemical blends from wings and claspers. The high number of species-specific compounds contrasts with the pair of closely related co-mimetic species *H. melpomene amaryllis* and *H. timareta thelxinoe* ([Mérot et al., 2015]). The latter are somewhat less divergent and have been found to hybridise and to exchange genes ([Heliconius Genome Consortium, 2012]). This may suggest that the chemical differentiation cumulates along evolutionary time between diverging *Heliconius* lineages.

Cuticular extracts from male and female bodies of both species were similar, suggesting a conserved chemical constitution of the exoskeleton of those parts which are presumably not directly involved in the production or release of signals for intra and interspecific chemical communication. This allowed us to use body extracts as a negative control, which contrasted with the high differences in chemical blends of wings and abdominal glands. Male androconia include modified scales in the region where wings overlap ([Emsley, 1963]; [Brown, 1981]), and are clearly exposed, presumably facilitating the diffusion chemical signals, when the male hovers close to the female, or when the male lands close to the female and attempts mating. This provides evidence for the implication of wing androconia in the production and/or release of chemical signals.

Regarding abdominal glands, *H. melpomene* males contain an important amount of  $\beta$ -ocimene in their claspers acting as an antiaphrodisiac transferred to females during copulation ([Schulz et al., 2008]). In other species, other compounds are found in important amounts ([Estrada et al., 2011]). If major and minor compounds produced in the claspers vary among males of different species, they can potentially be used in species discrimination. We indeed found strong differences between the claspers' composition of the two species. We found a huge peak of  $\beta$ -ocimene in the *H. hecale* glands, contrasting with the trace amounts found in *H. ismenius*. Several other compounds also showed qualitative and significant quantitative differences between species. Such variation could be indicative of male species and status during courtship, as males typically alight with claspers wide open, presumably releasing chemical cues.

Although female species-specific compounds are largely described from moths ([Arn et al., 1992]; [Lofstedt, 1993]; [Millar, 2000]; [Jurenka et al., 2003]), they have barely been explored in butterflies ([Dapporto, 2007]; [Ômura and Honda, 2011]; [Heuskin et al., 2014]). In the butterfly *Argynnis paphia*, female pheromones are necessary to elicit male courtship at short range ([Magnus, 1958]). Here, three compounds found at very low concentration in female wings and female abdominal glands varied between *H. h. melicerta* and *H. i. bouletti*, but did not differ significantly among them. Whether these compounds play a role in male behaviour is unknown. However, preliminary evidence suggests they could have a role in attracting males. Indeed, abdominal gland extracts of 4 females poured onto the wings of a mature male rendered him very attractive to the other males, and elicited active courtship lasting several minutes (data not shown). We therefore hypothesized the presence of female cues in the abdominal glands of females, whose specificity is unknown. The small chemical differences we found might therefore contribute to a female signal together with other short distance cues such as semiochemicals produced elsewhere or other emission patterns (e.g. temporal patterns). Chemical differences

in the extracts from female wings are depreciable. Thus, minor wing colour pattern differences between species might contribute to male discrimination happening in the short range, in a similar way than in *Lycaeides* butterflies ([Fordyce et al., 2002]). Actually, in spite of the minute-wise mimetic convergence between *H. hecale melicerta* and *H. ismenius bouletti*, minor differences exist in the shape and position of the forewing yellow elements.

Overall, our study suggests some candidate chemicals which might act as important pheromones used in species recognition. The large chemical differences seen in males suggest that species recognition might operate largely through female choice based on chemical signals. Consistently, several studies have reported chemical-mediated species recognition for butterfly females (e.g. [Vane-Wright and Boppre, 1993]; [Costanzo and Monteiro, 2007]). The important difference in the chemical constitution between males and females gives a hint of their implication for sex recognition too.

### 3.4.7 Compound volatility may not be a requirement for efficiency on species recognition

Among the three major compounds found in the claspers and wings of *H. h. melicerta* males and virtually absent in *H. i. bouletti* males, two are highly volatile ( $\beta$ -ocimene, 2-pentadecanone), which is consistent with an air-borne pheromone function. A third compound (heneicosane) has also been considered a volatile substance ([Ômura and Honda, 2011]), but it is an alkane of 21 carbons without additional functional groups and an elevated retention time, suggesting a rather low volatility. Nevertheless, we were able to detect it in a solid phase microextraction (SPME) experiment in which the males did not get into contact with the filter (results not shown). Most other compounds of minor concentration in the chemical cocktail seem to be non-volatile based on their retention times and mass spectra. Highly volatile pheromones travelling long distances are well known from moths ([Arn et al., 1992]), but aliphatic hydrocarbons may also be implicated in semiochemical communication ([Roelofs and Cardé, 1971]). Butterflies on the other hand do not exert long range chemical recognition and courtship signals happen at a short range. Yet direct contact between sensorial organs (antennae and legs with wings and head) occurs during courtship in *Heliconius* butterflies (this and previous studies: [Crane, 1955]; [Mega and Araújo, 2010]; [Klein and Araújo, 2010]), so some compounds like cuticular hydrocarbons could act as pheromones of contact, as in other insects including Lepidoptera ([Millar, 2000]; [Jurenka et al., 2003]; [Ferveur, 2005]; [Hay-Roe et al., 2007]; [Dapporto, 2007]; [Heuskin et al., 2014]). Males also hover within a few centimeters of the female's head, and could mechanically disperse compounds of low volatility. A similar fast wing motion known as wing fanning has been documented in males of the moth *Bombyx mori* to enhance female volatile pheromone interception by antennae ([Loudon and Koehl, 2000]).

Our data suggest that chemical signatures may play an important role in keeping species boundaries in *Heliconius* butterflies, and represent one of the several barriers to interspecific hybridisation. However, we are still lacking the functional link between species-specific chemical cocktails and mate choice. Therefore, future work will be needed to assess and confirm the identity of the chemical compounds, and perform functional assays to test their implication in species recognition. Although factors influencing mate choice are crucial in the process of speciation with gene flow, microhabitat-related and other ecological adaptations are also highly important and probably contributed to the divergence of *H. hecale* and *H. ismenius* and to their advanced stage of speciation where interspecific gene flow is not really detectable with a genomics approach. Our work allows having a view on the genomic structure at a late stage of the speciation continuum and making comparisons throughout distinct *Heliconius*

clades, contributing to answer whether speciation rates may differ across these clades. In that sense, analysing gene flow between pairs of silvaniform species whose divergence time is alike to the one within the *melpomene* clade will be highly useful. Our results ultimately contribute to a better understanding of the process of speciation with gene flow and the forces leading diversification in the genus *Heliconius*.

## 3.5 Supplementary information

### 3.5.1 Supplementary tables

Table 3.2 – Sampling localities of *Heliconius hecale* and *H. ismenius specimens* (this study) and of some specimens from the *H. melpomene* clade used as outgroup (Martin et al. 2013). Specimens in **bold script** (3 per race) constitute the restricted dataset used for most analyses.

ID	Code for sequencing and analysis	Species	Race	Gender	Country	Locality	GPS coordinates	Mean read coverage
MJ12-2110	STRI_WOM_5560	<b>H. hecale</b>	<b>melicerta</b>	<b>Male</b>	<b>Panama</b>	<b>Apiario, Darién</b>	N8.18647, W77.48550	<b>23.9</b>
MJ12-2111	STRI_WOM_5561	<i>H. hecale</i>	<i>melicerta</i>	Female	Panama	IFAD, Darién	N8.30255, W77.81658	17.2
MJ12-2118	STRI_WOM_5568	<i>H. hecale</i>	<i>melicerta</i>	Male	Panama	Canglón, Darién	N8.28008, W77.80957	23.7
MJ12-2122	STRI_WOM_5572	<b>H. hecale</b>	<b>melicerta</b>	<b>Male</b>	<b>Panama</b>	<b>Santa Rosa, Darién</b>	<i>NR</i>	<b>29.8</b>
MJ12-2138	STRI_WOM_5588	<b>H. hecale</b>	<b>melicerta</b>	<b>Male</b>	<b>Panama</b>	<b>Canglón, Darién</b>	N8.28008, W77.80957	<b>24.0</b>
MJ12-2062	STRI_WOM_5519	<i>H. hecale</i>	<i>zuleika</i>	Male	Panama	Meseta de Chorchá, Chiriquí	N8.41608, W82.21675	19.1
MJ12-2065	STRI_WOM_5522	<i>H. hecale</i>	<i>zuleika</i>	Male	Panama	Meseta de Chorchá, Chiriquí	N8.41608, W82.21675	19.0
MJ12-2068	STRI_WOM_5525	<b>H. hecale</b>	<b>zuleika</b>	<b>Male</b>	<b>Panama</b>	<b>Meseta de Chorchá, Chiriquí</b>	N8.41608, W82.21675	<b>34.4</b>
MJ12-2070	STRI_WOM_5527	<b>H. hecale</b>	<b>zuleika</b>	<b>Male</b>	<b>Panama</b>	<b>Meseta de Chorchá, Chiriquí</b>	N8.41608, W82.21675	<b>27.3</b>
MJ12-2071	STRI_WOM_5528	<b>H. hecale</b>	<b>zuleika</b>	<b>Male</b>	<b>Panama</b>	<b>Meseta de Chorchá, Chiriquí</b>	N8.41608, W82.21675	<b>26.0</b>
MJ11-2141	STRI_WOM_5711	<b>H. ismenius</b>	<b>boulleti</b>	<b>Male</b>	<b>Panama</b>	<b>Santa Librada</b>	N8.16783, W77.48589	<b>33.8</b>
18154	STRI_WOM_5712	<i>H. ismenius</i>	<i>boulleti</i>	Male	Panama	Santa Librada	N8.16783, W77.48589	18.4
18117	STRI_WOM_5713	<b>H. ismenius</b>	<b>boulleti</b>	<b>Male</b>	<b>Panama</b>	<b>Santa Librada</b>	N8.16783, W77.48589	<b>23.1</b>
18101	STRI_WOM_5714	<i>H. ismenius</i>	<i>boulleti</i>	Male	Panama	Santa Librada	N8.16783, W77.48589	16.8
18100	STRI_WOM_5715	<b>H. ismenius</b>	<b>boulleti</b>	<b>Male</b>	<b>Panama</b>	<b>Santa Librada</b>	N8.16783, W77.48589	<b>26.6</b>
NA	STRI_WOM_5428	<b>H. ismenius</b>	<b>clarescens</b>	<b>Male</b>	<b>Panama</b>	<b>Meseta de Chorchá, Chiriquí</b>	N8.41608, W82.21675	<b>21.9</b>
MJ12-2074	STRI_WOM_5531	<b>H. ismenius</b>	<b>clarescens</b>	<b>Male</b>	<b>Panama</b>	<b>Meseta de Chorchá, Chiriquí</b>	N8.41608, W82.21675	<b>24.9</b>
MJ12-2214	STRI_WOM_5664	<b>H. ismenius</b>	<b>clarescens</b>	<b>Male</b>	<b>Panama</b>	<b>Meseta de Chorchá, Chiriquí</b>	N8.41608, W82.21675	<b>33.0</b>
553	chi553	<i>H. cydno</i>	<i>chioneus</i>	Male	Panama	NR	N9.1714, W79.7573	35.8
1	meP1	<i>H. melpomene</i>	<i>melpomene</i>	NR	Panama	NR	N8.6136, W78.1398	23.0
2071	ro2071	<i>H. melpomene</i>	<i>rosina</i>	Male	Panama	NR	N9.1206, W79.6969	36.7
13435	meF13435	<i>H. melpomene</i>	<i>melpomene</i>	Male	French Guiana	NR	N4.9151, W52.3755	35.8
09-108	ag108	<i>H. melpomene</i>	<i>aglaope</i>	Male	Peru	NR	S5.9103, W76.2258	36.6
06-216	am216	<i>H. melpomene</i>	<i>amaryllis</i>	Male	Peru	NR	S5.6756, W77.6747	32.6
09-86	tip86	<i>H. timareta</i>	<i>thelxinoe</i>	Male	Peru	NR	S6.4550, W76.2983	45.6

NR: Not reported  
NA: Not assigned

Table 3.3 – Ethogram of the male-female interactions during courtship and mating in *Heliconius hecale*.

Behavioral unit	Alternative names	Some other butterfly species for which this behavior has been reported	Description
Male-related			
Main and minor units			
Localization (A)	Inspection (B), nudging (E)	<i>H. melpomene</i> (G), <i>H. erato</i> (B, E), <i>Dryas iulia</i> (C), <i>Bicyclus anynana</i> (A), several other species in which males fly almost constantly in search of females (patrolling species, see F).	The male approaches the alighted female
Flickering (A)	Hovering (B, C), fluttering of wings (D), fanning (E)	<i>H. erato</i> (B, D, E), <i>Dryas iulia</i> (C)	The male flights sustainedly over the alighted female (mostly over her head) moving his wings very fast. The forewings and hindwings are very distant from each other, the androconia are exposed. The male can come really close to the female during this step, almost touching her sometimes
Alighting	Parallel alighting (C)		Male alights very closely besides the female, always keeping forewings and hindwings separated
Slow flapping (B)	Wing clapping (C)		Standing beside the female, the male opens and closes the wings slowly. Generally linked to androconia exposition and orientation behaviors
Claspers opening			The male opens his claspers during slow flapping and orientation
Androconia exposition (B)	Silvery friction surfaces exposition (D)		The male increases the distance between the forewings and hindwings, exposing the silverish overlapping region between them (androconia)
Attempting (A)	Abdomen bending (B)		Alighted, next to the female, the male bends his abdomen towards hers. His head can be very close to the female's one

Atypical units	
Alighting on wings (B, C)	<i>H. erato phyllis</i> (B), <i>Dryas iulia</i> (C)
Close contact while alighting	The male lands on the female wings briefly or sustainedly
Orientation	The male lands on the basal part of the wings and the thorax of the female, touching her legs with his legs The male changes the angle between his body and the female's one, flapping slowly his wings and exposing androconia. At the end of this step, the male generally recovers a parallel position in relation to the female
Facing	<i>H. erato hydara</i> (D) The male faces the female, so that both heads are very close to each other. The antennae seemingly make contact
Touching with proboscis (B, C)	<i>H. erato phyllis</i> (B), <i>Dryas iulia</i> (C), <i>H. erato hydara</i> (D) The male extends his proboscis and touches different parts of the female with it
Female-related	
Female rejection	<i>H. melpomene</i> (G), <i>H. erato phyllis</i> (B), <i>Dryas iulia</i> (C), <i>Agraulis vanilla</i> (H), several species in the <i>Pieridae</i> family (this behavior has been called the <i>Pieridae</i> rejection posture; for review see F) Mate-refusal posture (J), Posterior vibration (B), abdomen raising and/or wing pressure (C) The female raises the abdomen to place it above the level of the wings and extrudes the abdominal glands. Simultaneously, her wings are entirely opened and pressed against the ground
Female acceptance	<i>H. erato phyllis</i> (B), <i>Dryas iulia</i> (C), <i>Pieris napi</i> (I), <i>Eurema lisa</i> (K), <i>Colias philodice</i> (L) Wing shutting (C), contraction(B) The female shuts her wings, descends her abdomen under the hindwings line and remains motionless



Male and female-related		
Flight pursuit (B)	Flight (E)	<i>H. erato phyllis</i> (B), several species in the subfamily <i>Heliconiinae</i> (E)
Genital contact (A)		
		The male chases the female while both are flying
		The male claspers come into contact with the tip of the abdomen of the female, which may precede or not mating
Success	Copulation (A, D)	The male grabs the tip of the female's abdomen with his claspers, and they stay attached for a long time (between 15 minutes and many hours)
Rotation	Coupled rotation (C)	After copulation the male turns the angle of its body, pointing with its head on the opposite direction than the female's one

The references from which the information is taken are shown in brackets.

A: Nieberding et al., 2008;  
B: Mega and Araújo, 2010;  
C: Klein and Araújo, 2010;  
D: Crane, 1955; E: Crane, 1957;  
F: Scott, 1972;  
G: Schulz et al., 2008;  
H: Rutowski and Schaefer, 1984;  
I: Forsberg and Wiklund, 1989;  
J: Obara, 1964; K: Rutowski, 1977;  
L: Silberglied and Taylor, 1978.

Table 3.4 – Compositional similarities and differences between the chemical blends extracted from abdominal glands of males and females of *H. hecale melicerta* and *H. ismenius bouletti*.

No of CP	No of CP (heatmap) <sup>b</sup>	RT <sup>c</sup>	RI <sup>d</sup>	Relative concentration (average)				Relative concentration (standard deviation) <sup>e</sup>				Kruskal-Wallis H	p-value	Pairwise Wilcoxon rank-sum (p-value)			
				MFG	MMC	BMC	BFG	MFG	MMC	BMC	BFG			MFG-MMC	MFG-BFG	MMC-BMC	BMC-BFG
1	1	4.631	991.500	0.000	<b>0.337</b>	0.000	0.000	0.000	0.154	0.000	0.000	23.425	<b>0.000</b>	0.008	NA	<b>0.008</b>	NA
2	2	5.373	1036.551	0.000	<b>0.594</b>	0.000	0.000	0.000	0.000	0.251	0.000	23.425	<b>0.000</b>	0.008	NA	<b>0.008</b>	NA
3	3	5.546	1046.972	0.045	<b>2.932</b>	0.150	0.003	0.038	0.397	0.077	0.009	21.386	<b>0.000</b>	0.030	0.250	<b>0.013</b>	<b>0.012</b>
4	4	5.98	1071.830	0.051	0.015	0.011	0.036	0.075	0.016	0.017	0.066	2.704	0.440	1.000	1.000	1.000	1.000
5	5	6.053	1076.355	0.060	0.014	0.013	0.070	0.087	0.017	0.021	0.147	2.607	0.456	1.000	1.000	1.000	1.000
6	6	6.178	1083.948	0.100	0.037	0.483	0.110	0.158	0.049	0.449	0.202	5.000	0.172	1.000	1.000	0.438	0.546
7	7	6.276	1089.916	0.045	0.013	0.011	0.052	0.072	0.016	0.017	0.112	0.855	0.836	1.000	1.000	1.000	1.000
8	8	6.318	1092.574	0.035	0.015	0.011	0.043	0.059	0.016	0.017	0.094	0.522	0.914	1.000	1.000	1.000	1.000
9	9	6.443	1100.104	0.146	0.084	0.058	0.152	0.056	0.096	0.090	0.123	2.984	0.394	1.000	1.000	1.000	1.000
11	10	6.631	1111.285	0.130	0.053	0.044	0.112	0.168	0.061	0.069	0.250	1.375	0.711	1.000	1.000	1.000	1.000
12	11	6.93	1129.681	0.000	<b>0.152</b>	0.000	0.000	0.000	0.105	0.000	0.000	23.425	<b>0.000</b>	<b>0.008</b>	NA	<b>0.008</b>	NA
13	12	9.678	1290.294	0.000	<b>0.301</b>	0.000	0.000	0.000	0.160	0.000	0.000	23.425	<b>0.000</b>	<b>0.008</b>	NA	<b>0.008</b>	NA
14	13	13.102	1504.195	0.000	<b>0.077</b>	0.000	0.000	0.000	0.072	0.000	0.000	23.425	<b>0.000</b>	<b>0.008</b>	NA	<b>0.008</b>	NA
15	14	14.232	1580.650	0.000	<b>0.092</b>	0.000	0.000	0.000	0.039	0.000	0.000	23.425	<b>0.000</b>	<b>0.008</b>	NA	<b>0.008</b>	NA
16	15	14.331	1587.415	0.000	<b>0.315</b>	0.000	0.000	0.000	0.117	0.000	0.000	23.425	<b>0.000</b>	<b>0.008</b>	NA	<b>0.008</b>	NA
17	16	14.421	1594.757	0.000	<b>0.107</b>	0.000	0.000	0.000	0.055	0.000	0.000	23.425	<b>0.000</b>	<b>0.008</b>	NA	<b>0.008</b>	NA
18	17	15.878	1696.233	0.000	0.048	0.000	0.000	0.000	0.044	0.000	0.000	18.701	<b>0.000</b>	<b>0.029</b>	NA	<b>0.029</b>	NA
19	18	16.319	1728.924	0.000	<b>0.078</b>	0.000	0.000	0.000	0.060	0.000	0.000	23.425	<b>0.000</b>	<b>0.008</b>	NA	<b>0.008</b>	NA
20	19	17.468	1815.608	0.000	<b>0.087</b>	<b>0.014</b>	0.000	0.000	0.114	0.012	0.000	20.693	<b>0.000</b>	0.014	NA	0.124	<b>0.028</b>
21	20	17.728	1835.985	0.223	<b>0.675</b>	0.004	0.023	0.144	0.313	0.007	0.029	19.733	<b>0.000</b>	<b>0.156</b>	<b>0.028</b>	<b>0.026</b>	0.397
23	21	17.865	1846.980	0.000	<b>0.068</b>	0.000	0.000	0.000	0.048	0.000	0.000	23.425	<b>0.000</b>	<b>0.008</b>	NA	<b>0.008</b>	NA
24	22	18.041	1861.098	<b>0.075</b>	<b>0.322</b>	0.000	0.000	0.057	0.194	0.000	0.000	22.343	<b>0.000</b>	<b>0.152</b>	<b>0.007</b>	<b>0.014</b>	NA
25	23	18.269	1878.493	0.100	<b>0.434</b>	0.000	0.111	0.081	0.264	0.000	0.275	16.402	<b>0.001</b>	<b>0.156</b>	0.315	<b>0.017</b>	0.269
26	24	18.396	1888.549	0.000	<b>0.232</b>	0.000	0.000	0.000	0.172	0.000	0.000	23.425	<b>0.000</b>	<b>0.008</b>	NA	<b>0.008</b>	NA
27	25	18.542	1901.268	0.000	<b>0.166</b>	<b>0.039</b>	0.000	0.000	0.122	0.026	0.000	20.553	<b>0.000</b>	0.014	NA	0.130	<b>0.028</b>
28	26	18.658	1909.367	0.000	<b>0.056</b>	0.000	0.000	0.000	0.051	0.000	0.000	18.701	<b>0.000</b>	<b>0.029</b>	NA	<b>0.029</b>	NA

29	27	18.746	1916.763	0.000	0.101	0.000	0.000	0.000	0.000	0.092	0.000	0.000	23.425	0.000	0.008	NA	0.008	NA
30	28	18.987	1936.409	0.168	0.434	0.117	0.032	0.308	0.442	0.184	0.051	6.387	0.094	1.000	0.976	1.000	1.000	1.000
31	29	19.291	1960.686	0.939	1.083	1.246	0.997	0.551	0.777	0.514	0.448	2.030	0.566	1.000	1.000	1.000	1.000	1.000
32	30	19.677	1993.060	0.000	0.379	0.042	0.000	0.000	0.293	0.048	0.000	20.263	0.000	0.014	NA	0.041	0.092	0.092
33	31	19.756	2000.446	0.000	0.445	0.056	0.000	0.000	0.208	0.059	0.000	20.639	0.000	0.014	NA	0.025	0.092	0.092
34	32	19.988	2021.239	0.000	0.799	0.589	0.000	0.000	0.322	0.388	0.000	21.281	0.000	0.014	NA	1.000	0.007	0.007
35	33	20.522	2069.144	0.000	0.000	0.391	0.000	0.000	0.000	0.190	0.000	23.425	0.000	NA	NA	0.008	0.004	0.004
36	34	20.6	2077.553	0.000	0.731	2.049	0.000	0.000	0.509	0.128	0.000	23.249	0.000	0.014	NA	0.011	0.007	0.007
37	35	20.745	2088.240	0.000	2.154	1.075	0.000	0.000	0.479	0.277	0.000	22.998	0.000	0.014	NA	0.022	0.007	0.007
38	36	20.94	2106.828	0.000	2.162	1.355	0.000	0.000	0.201	0.172	0.000	23.249	0.000	0.014	NA	0.011	0.007	0.007
39	37	21.007	2113.055	0.000	0.497	0.000	0.000	0.000	0.205	0.000	0.000	23.425	0.000	0.008	NA	0.008	NA	NA
40	38	21.342	2143.741	0.913	1.480	0.936	0.808	0.794	0.479	0.871	0.939	1.673	0.643	1.000	1.000	1.000	1.000	1.000
41	39	21.585	2166.646	0.434	1.297	0.510	0.436	0.429	0.568	0.532	0.579	6.915	0.075	0.182	1.000	0.390	1.000	1.000
42	40	21.723	2178.246	0.000	0.135	0.971	0.000	0.000	0.151	0.163	0.000	20.128	0.000	0.370	NA	0.024	0.007	0.007
44	41	22.133	2216.663	0.000	0.846	0.301	0.000	0.000	0.397	0.262	0.000	20.553	0.000	0.014	NA	0.130	0.028	0.028
45	42	22.289	2230.462	0.000	0.740	0.583	0.000	0.000	0.451	0.458	0.000	18.539	0.000	0.014	NA	1.000	0.028	0.028
46	43	22.605	2259.405	0.000	0.049	0.355	0.000	0.000	0.069	0.253	0.000	15.911	0.001	0.142	NA	0.219	0.028	0.028
47	44	22.677	2266.154	0.000	1.551	1.086	0.000	0.000	0.446	0.189	0.000	21.646	0.000	0.014	NA	0.660	0.007	0.007
48	45	22.807	2278.554	0.000	0.964	1.820	0.000	0.000	0.300	0.209	0.000	23.249	0.000	0.014	NA	0.011	0.007	0.007
49	46	22.984	2292.869	0.000	2.052	1.082	0.000	0.000	0.323	0.241	0.000	23.249	0.000	0.014	NA	0.011	0.007	0.007
50	47	23.104	2305.527	0.039	1.286	0.441	0.018	0.067	0.229	0.105	0.029	20.661	0.000	0.026	1.000	0.013	0.018	0.018
51	48	23.222	2316.317	0.000	0.638	0.267	0.000	0.000	0.205	0.143	0.000	20.818	0.000	0.014	NA	0.076	0.028	0.028
52	49	23.464	2339.395	0.000	0.512	0.000	0.000	0.000	0.169	0.000	0.000	23.425	0.000	0.008	NA	0.008	NA	NA
53	50	23.928	2381.952	0.000	1.495	0.000	0.000	0.000	0.572	0.000	0.000	23.425	0.000	0.008	NA	0.008	NA	NA
54	51	24.133	2401.147	0.000	0.377	0.256	0.000	0.000	0.277	0.149	0.000	18.356	0.000	0.048	NA	1.000	0.007	0.007
55	52	24.202	2407.591	0.000	1.034	0.000	0.000	0.000	0.461	0.000	0.000	23.425	0.000	0.008	NA	0.008	NA	NA
56	53	24.769	2464.018	0.000	1.571	0.605	0.000	0.000	0.455	0.345	0.000	21.097	0.000	0.014	NA	0.043	0.028	0.028
57	54	24.849	2472.672	0.000	0.566	0.371	0.000	0.000	0.473	0.225	0.000	13.774	0.003	0.142	NA	1.000	0.028	0.028
58	55	25.032	2490.739	0.000	1.547	0.504	0.000	0.000	0.334	0.305	0.000	21.702	0.000	0.014	NA	0.011	0.028	0.028
59	56	25.114	2500.194	0.048	0.873	0.315	0.062	0.076	0.283	0.169	0.062	17.493	0.001	0.026	1.000	0.013	0.224	0.224
60	57	25.438	2533.940	0.000	1.309	0.011	0.000	0.000	0.225	0.017	0.000	20.278	0.000	0.014	NA	0.022	0.699	0.699
61	58	25.606	2551.156	0.000	0.490	0.000	0.000	0.000	0.397	0.000	0.000	23.425	0.000	0.008	NA	0.008	NA	NA
62	59	25.901	2582.668	0.000	0.771	0.000	0.000	0.000	0.409	0.000	0.000	23.425	0.000	0.008	NA	0.008	NA	NA
63	60	26.369	2631.234	0.000	0.418	0.034	0.000	0.000	0.119	0.068	0.000	20.278	0.000	0.014	NA	0.022	0.699	0.699
64	61	26.705	2666.405	0.000	1.123	0.211	0.000	0.000	0.605	0.178	0.000	15.911	0.001	0.048	NA	0.219	0.092	0.092

65	62	26.986	2695.486	0.283	<b>1.247</b>	0.641	0.232	0.165	0.330	0.188	0.171	18.879	<b>0.000</b>	<b>0.013</b>	1.000	<b>0.013</b>	<b>0.028</b>
66	63	27.272	2726.195	0.000	<b>0.541</b>	<b>0.027</b>	0.000	0.000	0.175	0.043	0.000	20.278	<b>0.000</b>	<b>0.014</b>	NA	<b>0.022</b>	0.699
67	64	27.492	2748.846	0.173	0.333	0.103	0.184	0.197	0.189	0.111	0.241	5.419	0.144	1.000	1.000	0.120	1.000
68	65	27.724	2774.265	0.000	<b>0.359</b>	0.000	0.000	0.000	0.275	0.000	0.000	18.701	<b>0.000</b>	<b>0.029</b>	NA	<b>0.029</b>	NA
69	66	27.879	2789.869	0.000	<b>0.592</b>	<b>0.164</b>	0.000	0.000	0.296	0.147	0.000	20.068	<b>0.000</b>	<b>0.014</b>	NA	0.325	<b>0.028</b>
70	67	27.954	2797.936	0.191	0.315	0.183	0.074	0.125	0.250	0.159	0.079	5.518	0.138	1.000	0.321	1.000	1.000
71	68	28.587	2864.101	0.335	0.590	0.416	0.308	0.386	0.254	0.369	0.265	2.984	0.394	1.000	1.000	1.000	1.000
72	69	28.72	2879.061	0.349	0.577	0.277	0.180	0.256	0.225	0.217	0.177	6.974	0.073	1.000	1.000	0.390	1.000
73	70	29.196	2928.189	0.000	<b>0.429</b>	<b>0.180</b>	0.000	0.000	0.224	0.190	0.000	18.914	<b>0.000</b>	<b>0.014</b>	NA	0.225	0.092
75	71	29.988	3011.496	0.000	<b>0.177</b>	<b>0.222</b>	0.000	0.000	0.118	0.164	0.000	18.338	<b>0.000</b>	<b>0.014</b>	NA	1.000	<b>0.028</b>
76	72	30.417	3056.154	1.336	1.326	1.161	1.091	0.296	0.212	0.331	0.526	1.060	0.787	1.000	1.000	1.000	1.000
77	73	30.728	3088.143	0.000	0.000	0.000	<b>0.273</b>	0.000	0.000	0.000	0.195	19.022	<b>0.000</b>	NA	<b>0.020</b>	NA	<b>0.020</b>
78	74	30.775	3093.912	0.000	<b>0.837</b>	<b>0.891</b>	0.000	0.000	0.258	0.201	0.000	21.045	<b>0.000</b>	<b>0.014</b>	NA	1.000	<b>0.007</b>
79	75	30.998	3117.691	0.000	<b>0.446</b>	<b>0.517</b>	0.000	0.000	0.198	0.148	0.000	20.995	<b>0.000</b>	<b>0.014</b>	NA	1.000	<b>0.007</b>

<sup>a</sup> Unique numeric label for each compound. <sup>b</sup> Label of a compound in the respective heatmap. <sup>c</sup> Retention time. <sup>d</sup> Retention index. <sup>e</sup> Values in **bold** script indicate a significant association between a given compound and the category, following the indval index with  $p < 0.001$ . CP : Compound. HM: Heatmap. mfg: Glands from *H. h. melicerta* females. mmc: Claspers from *H. h. melicerta* males. bmc: Claspers from *H. i. bouletti* males. bfg: Glands from *H. i. bouletti* females.

Code of cell colours:

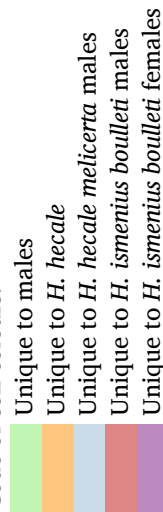


Table 3.5 – Compositional similarities and differences between the chemical blends extracted from wings of males and females of *H. hecale melicerta* and *H. ismerius bouletii*.

No of Cpa	No of CP (HM) <sup>b</sup>	R <sup>Tc</sup>	R <sup>Id</sup>	Relative concentration (average) <sup>e</sup>				Relative concentration (standard deviation)				Kruskal- Wallis H	p-value	Pairwise Wilcoxon rank-sum (p-value)			
				MFOW	MMOW	BMOW	BFOW	MFOW	MMOW	BMOW	BFOW			MFOW- MMOW	MFOW- BFOW	MMOW- BMOW	BMOW- BFOW
3	1	5.546	1046.972	0.029	0.057	0.000	0.000	0.057	0.054	0.000	0.000	6.764	0.080	1.000	1.000	0.251	NA
4	2	5.98	1071.830	0.114	0.081	0.037	0.040	0.179	0.132	0.003	0.005	0.400	0.940	1.000	1.000	1.000	1.000
5	3	6.053	1076.355	0.129	0.082	0.047	0.042	0.202	0.119	0.005	0.006	0.825	0.843	1.000	1.000	1.000	1.000
6	4	6.178	1083.948	0.201	0.129	0.073	0.078	0.263	0.166	0.008	0.001	0.746	0.862	1.000	1.000	1.000	1.000
7	5	6.276	1089.916	0.103	0.072	0.032	0.033	0.160	0.119	0.004	0.003	0.423	0.935	1.000	1.000	1.000	1.000
8	6	6.318	1092.574	0.085	0.061	0.027	0.036	0.134	0.101	0.003	0.012	1.070	0.784	1.000	1.000	1.000	1.000
9	7	6.443	1100.104	0.075	0.142	0.179	0.161	0.087	0.070	0.020	0.008	6.081	0.108	1.000	0.650	1.000	1.000
10	8	6.508	1103.779	0.013	0.043	0.044	0.028	0.025	0.050	0.030	0.024	3.269	0.352	1.000	1.000	1.000	1.000
11	9	6.631	1111.285	0.222	0.175	0.116	0.127	0.310	0.213	0.011	0.014	1.429	0.699	1.000	1.000	1.000	1.000
22	10	17.813	1842.891	0.000	0.505	0.007	0.000	0.000	0.163	0.005	0.000	14.976	0.002	0.045	NA	0.030	0.593
31	11	19.291	1960.686	0.031	0.120	0.166	0.094	0.063	0.142	0.143	0.066	3.441	0.328	1.000	1.000	1.000	1.000
33	12	19.756	2000.446	0.000	0.027	0.000	0.000	0.000	0.013	0.000	0.000	15.729	0.001	0.027	NA	0.027	NA
34	13	19.988	2021.239	0.014	0.053	0.007	0.012	0.013	0.035	0.005	0.021	8.320	0.040	0.349	1.000	0.073	1.000
36	14	20.6	2077.553	0.017	0.032	0.005	0.000	0.035	0.024	0.005	0.000	6.902	0.075	1.000	1.000	0.268	1.000
38	15	20.94	2106.828	0.000	1.295	0.000	0.000	0.000	0.267	0.000	0.000	15.729	0.001	0.027	NA	0.027	NA
40	16	21.342	2143.741	0.039	0.068	0.045	0.245	0.077	0.122	0.058	0.236	2.552	0.466	1.000	1.000	1.000	1.000
43	17	22.016	2204.892	0.000	0.192	0.000	0.000	0.000	0.113	0.000	0.000	12.673	0.005	0.068	NA	0.068	NA
45	18	22.289	2230.462	0.000	0.045	0.000	0.000	0.000	0.039	0.000	0.000	15.750	0.001	0.027	NA	0.027	NA
47	19	22.677	2266.154	0.000	0.112	0.000	0.000	0.000	0.104	0.000	0.000	12.673	0.005	0.068	NA	0.068	NA
48	20	22.807	2278.554	0.029	0.067	0.000	0.000	0.058	0.062	0.000	0.000	11.836	0.008	0.525	1.000	0.045	NA
50	21	23.104	2305.527	0.012	0.324	0.000	0.006	0.014	0.123	0.000	0.006	13.844	0.003	0.063	1.000	0.054	0.736
52	22	23.464	2339.395	0.000	0.040	0.000	0.000	0.000	0.038	0.000	0.000	12.673	0.005	0.068	NA	0.068	NA
53	23	23.928	2381.952	0.000	0.178	0.000	0.000	0.000	0.134	0.000	0.000	15.729	0.001	0.027	NA	0.027	NA
54	24	24.133	2401.147	0.000	0.216	0.000	0.000	0.000	0.177	0.000	0.000	15.729	0.001	0.027	NA	0.027	NA
59	25	25.114	2500.194	0.054	0.137	0.000	0.017	0.037	0.209	0.000	0.017	9.816	0.020	1.000	1.000	0.054	0.736
60	26	25.438	2533.940	0.068	0.341	0.000	0.000	0.127	0.323	0.000	0.000	12.576	0.006	0.360	1.000	0.045	NA
65	27	26.986	2695.486	0.184	0.245	0.046	0.257	0.116	0.256	0.037	0.098	7.302	0.063	1.000	1.000	0.255	0.343

72	28	28.72	2879.061	0.296	0.244	0.041	0.196	0.181	0.256	0.020	0.022	9.797	<b>0.020</b>	1.000	1.000	<b>0.036</b>	0.343
74	29	29.894	3000.839	0.143	0.207	0.127	0.007	0.184	0.144	0.020	0.006	6.828	0.078	1.000	1.000	1.000	0.343
76	30	30.417	3056.154	0.396	0.264	0.170	0.519	0.242	0.266	0.075	0.129	4.604	0.203	1.000	1.000	1.000	0.343
77	31	30.728	3088.143	<b>0.063</b>	0.000	0.000	<b>0.008</b>	0.060	0.000	0.000	0.013	9.910	<b>0.019</b>	0.081	0.995	NA	1.000
78	32	30.775	3093.912	0.000	<b>0.125</b>	<b>0.048</b>	0.000	0.000	0.188	0.021	0.000	13.138	<b>0.004</b>	<b>0.045</b>	NA	1.000	0.218
79	33	30.998	3117.691	<b>0.044</b>	0.000	0.000	0.000	0.055	0.000	0.000	0.000	11.784	<b>0.008</b>	<b>0.049</b>	0.356	NA	NA

<sup>a</sup> Unique numeric label for each compound. <sup>b</sup> Label of a compound in the respective heatmap. <sup>c</sup> Retention time. <sup>d</sup> Retention index. <sup>e</sup> Values in **bold** script indicate a significant association between a given compound and the category, following the indval index with  $p < 0.001$ . CP: Compound. HM: Heatmap. mfow: Wings from *H. h. melicerta* females. mmow: Wings from *H. h. melicerta* males. bmow: Wings from *H. i. bouletti* males. bfow: Wings from *H. i. bouletti* females.

Code of cell colours:







	Unique to males
	Unique to females
	Unique to <i>H. hecale</i>
	Unique to <i>H. hecale melicerta</i> males
	Unique to <i>H. hecale melicerta</i> females
	Unique to <i>H. ismenius bouletti</i> females




Table 3.6 – Compositional similarities and differences between the chemical blends extracted from the cuticle near the insertion of the wings of males and females of *H. hecale melicerta* and *H ismenitus bouletti*.

No of CP <sup>a</sup>	No of CP (HM) <sup>b</sup>	Relative concentration (average) <sup>c</sup>				Relative concentration (standard deviation)				Kruskal- Wallis H	p-value	Pairwise Wilcoxon rank-sum (p-value)			
		MFB	MMB	BMB	BFB	MFB	MMB	BMB	BFB			MFB- MMB	MFB- BFB	MMB- BMB	BMB- BFB
3	1	5.546	1046.972	0.009	0.100	0.000	0.011	0.019	0.054	0.000	0.015	0.178	1.000	0.065	1.000
4	2	5.98	1071.830	0.022	0.030	0.016	0.032	0.035	0.005	0.022	0.020	1.000	1.000	1.000	1.000
5	3	6.053	1076.355	0.022	0.037	0.019	0.033	0.033	0.006	0.027	0.019	1.000	1.000	1.000	1.000
6	4	6.178	1083.948	0.054	0.060	0.025	0.083	0.120	0.009	0.035	0.063	0.956	1.000	1.000	0.636
7	5	6.276	1089.916	0.015	0.029	0.010	0.027	0.021	0.004	0.014	0.016	1.000	1.000	0.266	0.959
8	6	6.318	1092.574	0.013	0.025	0.005	0.024	0.018	0.005	0.011	0.014	1.000	1.000	0.178	0.448
9	7	6.443	1100.104	0.144	0.155	0.141	0.165	0.090	0.009	0.043	0.033	1.000	1.000	1.000	1.000
10	8	6.508	1103.779	0.010	0.081	0.009	0.042	0.022	0.006	0.019	0.050	0.090	1.000	0.090	1.000
11	9	6.631	1111.285	0.025	0.070	0.026	0.063	0.034	0.013	0.036	0.035	0.632	0.402	1.000	0.402
30	10	18.987	1936.409	0.188	0.191	0.037	0.053	0.166	0.076	0.042	0.068	1.000	1.000	0.117	1.000
31	11	19.291	1960.686	1.413	0.999	0.946	0.961	0.280	0.119	0.397	0.443	0.381	0.905	1.000	1.000
38	12	20.94	2106.828	0.024	0.206	0.018	0.005	0.040	0.186	0.028	0.011	0.272	1.000	0.108	1.000
40	13	21.342	2143.741	1.968	1.702	1.681	1.498	0.352	0.177	0.631	0.599	1.000	1.000	1.000	1.000
41	14	21.585	2166.646	0.636	0.496	0.859	0.728	0.661	0.168	0.496	0.417	1.000	1.000	1.000	1.000
49	15	22.984	2292.869	0.007	0.027	0.000	0.000	0.016	0.022	0.000	0.000	0.888	1.000	0.208	NaN
50	16	23.104	2305.527	0.027	0.091	0.000	0.020	0.037	0.067	0.000	0.028	1.000	1.000	0.065	1.000
59	17	25.114	2500.194	0.151	0.135	0.008	0.043	0.081	0.112	0.018	0.033	1.000	0.190	0.090	0.713
60	18	25.438	2533.940	0.128	0.232	0.000	0.000	0.125	0.132	0.000	0.000	1.000	0.360	0.054	NaN
63	19	26.369	2631.234	0.000	0.030	0.000	0.000	0.000	0.016	0.000	0.000	0.032	NaN	0.032	NaN
64	20	26.705	2666.405	0.000	0.030	0.000	0.000	0.000	0.016	0.000	0.000	0.032	NaN	0.032	NaN
65	21	26.986	2695.486	0.309	0.329	0.183	0.184	0.080	0.100	0.147	0.089	1.000	0.333	0.667	1.000
66	22	27.272	2726.195	0.015	0.056	0.014	0.000	0.021	0.031	0.031	0.000	0.632	1.000	0.330	1.000
67	23	27.492	2748.846	0.011	0.049	0.000	0.000	0.024	0.037	0.000	0.000	0.700	1.000	0.208	NaN
69	24	27.879	2789.869	0.024	0.027	0.011	0.005	0.044	0.021	0.024	0.011	1.000	1.000	1.000	1.000
70	25	27.954	2797.936	0.096	0.018	0.017	0.018	0.096	0.021	0.024	0.018	1.000	1.000	1.000	1.000
71	26	28.587	2864.101	0.045	0.043	0.000	0.012	0.086	0.013	0.000	0.017	1.000	1.000	0.062	1.000

72	27	28.72	2879.061	0.147	0.158	0.201	0.093	0.115	0.090	0.318	0.062	1.770	0.621	1.000	1.000	1.000
73	28	29.196	2928.189	0.058	0.057	0.042	0.011	0.062	0.071	0.072	0.025	2.611	0.456	1.000	0.948	1.000
74	29	29.894	3000.839	0.000	0.015	0.000	0.008	0.000	0.017	0.000	0.017	4.719	0.194	0.656	1.000	0.656
75	30	29.988	3011.496	0.006	0.010	0.007	0.007	0.013	0.019	0.015	0.015	0.202	0.977	1.000	1.000	1.000
76	31	30.417	3056.154	1.174	0.757	0.894	0.890	0.353	0.035	0.332	0.260	3.919	0.270	0.667	1.000	1.000
77	32	30.728	3088.143	0.000	0.000	0.000	<b>0.175</b>	0.000	0.000	0.000	0.249	13.208	<b>0.004</b>	NaN	0.076	NaN
78	33	30.775	3093.912	0.033	0.104	0.212	0.182	0.031	0.078	0.199	0.227	4.494	0.213	1.000	1.000	1.000
79	34	30.998	3117.691	0.026	0.000	0.103	0.024	0.041	0.000	0.116	0.034	6.449	0.092	1.000	1.000	0.636

<sup>a</sup> Unique numeric label for each compound. <sup>b</sup> Label of a compound in the respective heatmap. <sup>c</sup> Retention time. <sup>d</sup> Retention index. <sup>e</sup> Values in **bold** script indicate a significant association between a given compound and the category, following the indval index with  $p < 0.001$ . CP: Compound. HM: Heatmap. MFB: Cuticule from *H. h. melicerta females*. MMB: Cuticule from *H. h. melicerta males*. BMB: Cuticule from *H. i. bouletti males*. BFB: Cuticule from *H. i. bouletti females*.

Code of cell colours:

	Unique to <i>H. hecale</i>
	Unique to <i>H. hecale melicerta</i> males
	Unique to <i>H. ismenius bouletti</i> females



## 3.5.2 Supplementary figures

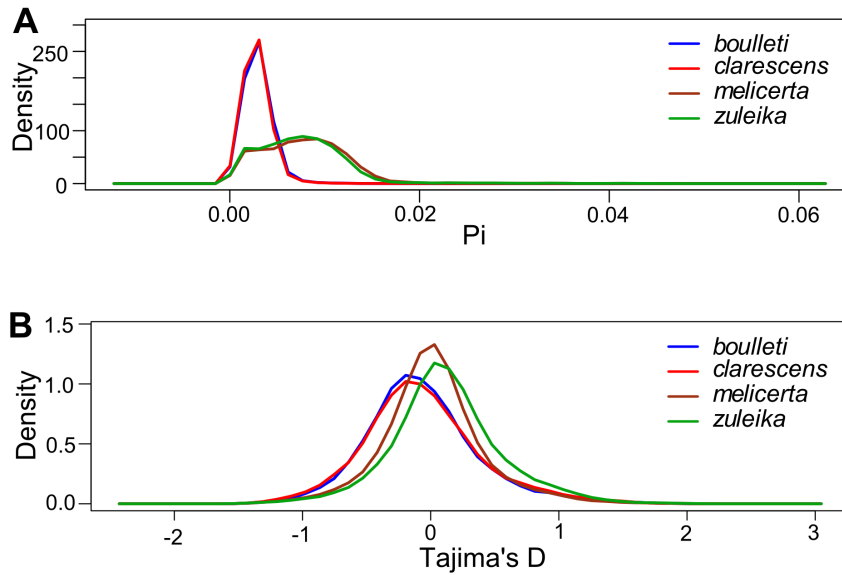


Figure 3.12 – Density plots of nucleotide diversity ( $\pi$  or  $\pi$ ) and Tajima's D values for non-overlapping 25-kb windows across the whole genome of *H. hecale* and *H. ismenius*. *bouletti*: *H. ismenius bouletti*; *clarescens*: *H. ismenius clarescens*; *melicerta*: *H. hecale melicerta*; *zuleika*: *H. h. zuleika*.

## Chapter 4

# Concluding remarks

### 4.1 It's all about selection

I have investigated the genetic and ecological architecture of phenotypic diversification in butterflies, contrasting instances of intraspecific phenotypic divergence to instances of both interspecific phenotypic convergence and divergence. I have shown that genetic architecture of wing coloration is at the same time conserved and flexible, and this versatility may be the reason for its ready response to selection.

By mapping wing colour loci in the silvaniform species *Heliconius hecale* and *H. ismenius* and placing my results on a comparative framework, I confirmed theoretical findings suggesting that the genetic architecture of a trait is an evolvable rather than a constant feature ([Lande, 1980]; [Barton, 1995]; [Coyne and Orr, 1998]; [Carter et al., 2005]; [Hansen, 2006]). In particular, I found that gene effect sizes, and not only gene frequencies, are evolutionary variables ([Hansen, 2006]). Moreover, I suggested that the evolution of the genetic architecture of adaptive traits is associated with the regime of selection acting on those traits. More precisely, I found that a multilocus architecture like the one existing in the two species studied here and in most other *Heliconius* species is associated with directional selection for local mimicry, leading to local monomorphism, which seems to be the ancestral state. In contrast, the single-locus (*supergene*) architecture restricted to *H. numata*, is associated with balancing selection keeping several morphs locally and is apparently a derived feature.

These results allow giving insights into the evolution of supergenes ([Nabours, 1933]; [Ford, 1966]; [Turner, 1977]). Rather than involving the sequestering and tight-linking of genetic regions brought together from distinct parts of the genome by translocations ([Nabours, 1933];

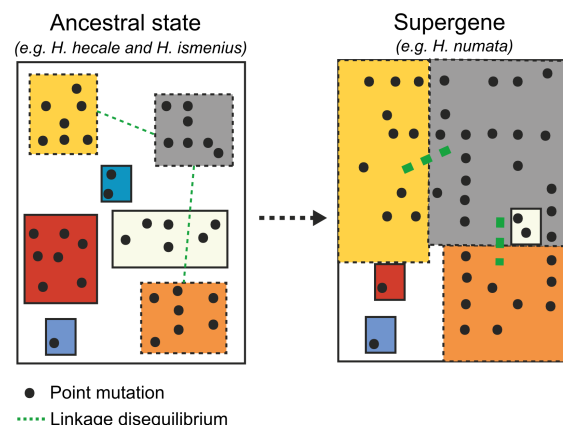


Figure 4.1 – **Hypothetical representation of the evolution of supergene P from an ancestral multilocus architecture.** Boxes of different colours correspond to different wing colour loci, whose phenotypic effect is reflected in their size relative to the whole rectangle area. The effect size of the supergene is enhanced presumably due to an increase in the strength of linkage (green dashed line) of coexisting genetic regions on a chromosome and to the accumulation of punctual mutations (black dots) lying on these regions.

[Ford, 1966]), I found evidence for the “sieve” hypothesis ([Turner, 1977]), according to which the recruitment of loci that are already closely linked and able to produce together major phenotypic effects gives rise to supergenes. Within the region homologous to the supergene *P* in *H. numata*, I found more than one locus being spatially linked and having near-to-equal size effects in *H. hecale* (see cartoon in Figure 4.1). Tight linkage between these co-varying loci could have resulted from the suppression of recombination in *H. numata* thanks to polymorphic inversions ([Joron et al., 2011]). Moreover, the large effect of the supergene *P* (Figure 4.1), could have hypothetically resulted from the aggregation of several independent small effect mutations (black dots in Figure 4.1) on the co-adapted loci rather than by large-effect mutations. This would be in consistence with the observations by Martin and Orgogozo ([Martin and Orgogozo, 2013]) on a large number of documented cases for large-effect variants. In *H. hecale* and *H. ismenius*, I also mapped a few major effect colour loci (although with an individual effect size that strongly contrasts with the one of the supergene *P*; see cartoon on Figure 4.1) whose phenotypic effect may also be potentially fuelled by small mutations and tuned by the small effect of several other loci for mimetic convergence (the two-step model; see Introduction; [Turner, 1985]; [Orr, 1998]).

Understanding whether or not convergent evolution is mirrored by convergence at the genotypic level is one of the main questions in evolutionary genetics. Attempts to answer this question have resulted from evaluating the phenotype-genotype correspondence at functional levels such as the genetic pathway, the gene and the nucleotide levels (Figure 8; e.g. ([Protas et al., 2006]; [Hoekstra, 2006]; [Arendt and Reznick, 2008]; [Manceau et al., 2010])). Here, I found that phenotypic convergence can be due to strikingly distinct genetic architectures in terms of the gene effect sizes, which has not been, to my knowledge, explicitly reported in literature. Namely, the multilocus and the supergene architectures are both able to control mimetic and non-mimetic but similar wing patterns, which highlights the power of selection for mimetic resemblance. This observation supports the idea that convergent evolution results from the unfettered ability of natural selection in solving similar ecological problems from different starting points (c.f. the adaptationist perspective on the evolutionary process; [Losos, 2011])). Pigmentation-related traits have been proposed to be more prone to result in convergent evolution via different genetic basis, given their several different underlying genetic pathways ([Hoekstra, 2006]; [Manceau et al., 2010]; [Kronforst et al., 2012]), and this has been observed in an array of instances ([Hoekstra, 2006]; [Manceau et al., 2010])). Furthermore, I also found a lack of genetic architecture-to-phenotype determinism in the fact that tiger-patterns like those displayed by *H. hecale* and *H. ismenius* are controlled by the exact same toolkit of wing colour loci used by contrastingly distinct coloured species in the genus *Heliconius* ([Kronforst et al., 2006a]; [Baxter et al., 2008b]; [Papa et al., 2008])). Importantly, this same toolkit can underlie mimetic phenotypes as well, even between very distantly-related species such as *H. melpomene* and *H. erato*. Thus, the flexibility in the phenotypic outcomes of the same loci is here shaped by strong selective constraints for mimetic resemblance among species forming the mimicry rings. Such versatility and evolvability in the genotype-to-phenotype map account for both phenotypic diversification and mimetic convergence in the genus *Heliconius*.

Mimetic convergence cannot be explained by chance. Instead, strong selective pressures are imposed by similar environmental conditions (i.e. mimicry rings), like other cases of phenotypic convergence ([Losos, 1992]; [Colosimo et al., 2005]; [Rogers and Bernatchez, 2007]; [Soria-Carrasco et al., 2014])). I have found evidence of selection acting on wing coloration at early stages of divergence, by using cline theory with population genetics, and by studying genome-wide patterns of divergence. Disruptive selection determining local adaptation to distinct mimetic environments on either side of the hybrid zone between *H. hecale melicerta* and *H.*

*hecale zuleika* is reflected in the clinal allele distribution at the colour loci, which contrasts with the absence of genetic structure found elsewhere on the genome. Selective pressures are of intermediate strength, as reflected by a hybrid zone of intermediate width relative to other interracial hybrid zones in the genus ([Mallet, 1986]; [Blum, 2002]; [Arias et al., 2012]). However, moderate selection is still able to keep colour patterning-related phenotypic distinctiveness in spite of an enormous level of gene flow among the two races. The complexity of tiger patterns and the highly speciose and variable tiger-patterned mimetic environments may determine disruptive selection of intermediate strength on hybrid butterflies at either side of this hybrid zone. I also found that the strength of selection on the distinct wing pattern elements differs, which suggests that all wing colour elements are not equal in the eyes of predators, and perhaps neither in the eyes of mates.

Recently, genome-wide molecular surveys have helped in pinpointing genomic regions under divergent selection, and presumed to be involved in speciation (e.g. [Jones et al., 2012]; [Lawniczak et al., 2010]; [Nadeau et al., 2012]; [Ellegren et al., 2012]; [Kronforst et al., 2013]). Such regions, which have been called genomic islands of divergence may be attributed to the direct action of divergent selection, but may also result from features of genome architecture, such as local rates of recombination or local nucleotide diversity ([Carneiro et al., 2010]; [Ellegren et al., 2012]; [Cruickshank and Hahn, 2014]). To really understand the processes underlying heterogeneous landscapes, the statistical exploration of the data coupled to a better knowledge on the differentiated ecological traits and their genetic basis is needed. Combining genomic scans with candidate gene approaches may allow distinguishing regions involved in diversification from false positives (e.g. ([Michel et al., 2010])). Here, I adopted such a synergistic approach and found that some islands of divergence are associated with wing colour loci at early stages of differentiation, thus showing that the regions known to underlie ecological differences under strong disruptive selection such as wing patterns appear as outliers of differentiation in the genome, and may therefore resist gene flow. However, I also found peaks of differentiation around wing colour regions between well diverged co-mimetic lineages, although genetic divergence at alleles causing similar phenotypes would not really be expected. I hypothesize that these molecular signals, rather than being the response to diverging selection, are the by-product of purifying selection on alleles for mimicry resemblance, manifested in the low nucleotide diversity around the wing colour loci. I observed a highly heterogeneous differentiation landscape in the interracial comparison, with multiple peaks contrasting with the few peaks of divergence associated to colour loci when comparing other pairs of colour races in *Heliconius* species explored so far. The increase in the number of islands of divergence has been suggested to result from the advancement of speciation towards later stages ([Kronforst et al., 2013]) contrasting with the classical model of speciation with gene flow, where islands are expected to become larger instead of more numerous along this continuum ([Feder et al., 2012]; see Introduction). Instead, the multiple peaks I observed may result from the overall very low nucleotide diversity in *H. hecale*, and interpreting this as evidence that those populations are at a more advanced stage of divergence than other pairs of races studied so far is not straightforward. Because of the statistical issues outlined here, I suggest that a deeper evaluation of this genomic data is lacking to be able to yield conclusions about how genomic differentiation has evolved along the two stages of the speciation continuum studied here. Although genomic surveys of differentiation can be highly useful to detect some of the genomic regions controlling ecologically important traits during speciation with gene flow, the genomic profile of differentiation can be misleading. So please, mind the gap between differentiation outliers and speciation genes!

## 4.2 Powerful, but not almighty

I have explored the process of speciation with gene flow among two pairs of taxa spanning the speciation continuum: between parapatric races of *H. hecale*, and between the two sympatric closely-related species *H. hecale* and *H. ismenius*. By doing so, I integrated the “spyglass” and the “magnifying glass” views on the speciation process (see Introduction or [Via, 2009]). This allows investigating the mechanisms for the evolution of reproductive isolation along the speciation continuum in the genus *Heliconius*.

Ecological speciation has been suggested to be favoured by divergent selection on ecologically important traits, in particular if these traits are also involved in assortative mating ([Bush, 1975]; [Felsenstein, 1981]; [Diehl and Bush, 1989]; [Kirkpatrick and Ravigné, 2002]; [Malausa et al., 2005]). In *Heliconius* butterflies like in other model systems, “magic” traits are able to enhance or maintain both extrinsic post-zygotic and pre-zygotic isolation simultaneously. Several instances exemplifying a similar situation involve divergent lineages at very early stages of speciation (e.g.; [Feder and Filchak, 1999]). A debated issue is whether single adaptive traits are able to drive speciation to completion or if instead, speciation will never proceed if additional features do not diverge in the process. This large and unsolved question applies to the multiple coloured races of *Heliconius* butterflies.

The parapatric races of *Heliconius hecale* are distinguishable merely at the wing colour pattern level, and they do not seem to have evolved other specific ecological adaptations. In particular, no clear differences are known on macrohabitat and microhabitat preference between the *H. hecale* races. Both races were indeed usually collected in similar habitats, with open spaces formed by grasses and bushes, close to the borders of forests, or in forest gaps (personal observation). Furthermore, larvae of the two races feed on the same host-plant *Passiflora vitifolia*. These observations are in accordance to previous work; more precisely, distinct ecological factors (e.g. larval host plant or habitat specialisation) have not been shown to be involved in geographic race formation in the genus ([Mallet and Gilbert, 1995]; [Jiggins et al., 1996]).

I observed that assortative mating based on wing coloration, although partial and asymmetric, exists between the races of *H. hecale* studied here, confirming that colour-based male mate choice is ubiquitous across the genus and evolves early in the speciation continuum. In contrast, at advanced stages of the speciation process, wing coloration does not play an outstanding role in keeping species boundaries. Indeed, species *H. hecale* and *H. ismenius* are mimetic and still maintain their genomic integrity (see below). Here, I put in evidence that the evolution of divergent chemical signals has occurred between these “good” sympatric species that share mimetic patterns, and might contribute to the maintenance of the species boundaries between lineages that would be expected to hybridise. I suggest that species recognition based on pheromones is strongly operated by females, contrary to mate choice based on wing coloration, which is operated mainly by males. Other factors such as host plant specificity and microhabitat preference also seem to have diverged between these species, although rigorous ecological data is lacking.

My results do not contrast with findings on other sub-clades in the genus. Shifts in coloration have been suggested to have been important for speciation ([Mallet et al., 1998a]; [Mallet, 2010]; [Jiggins et al., 2001]; [Chamberlain et al., 2009]). For instance, sister species *H. melpomene* and *H. cydno* were proposed to have speciated in association with a mimicry shift. However, they have also evolved host plant specificity, microhabitat preference and partial hybrid infertility due to genomic incompatibilities ([Jiggins et al., 2001]; [Naisbit et al., 2002]), suggesting that colour pattern divergence may not be a sufficient driving force of speciation in the genus.

My results reinforce the view that wing coloration divergence is central at the beginning of the speciation process but is presumably not enough to achieve speciation. Speciation in the clade might instead be a multidimensional process, with distinct traits acting as reproductive isolation barriers accumulating along the process. In accordance to this idea, there are 10-fold more geographic colour races than species in the genus, suggesting that divergence in wing coloration alone does not always lead to separate species. Moreover, some pairs of geographic colour races within species have deeper genealogies than pairs of species within the genus ([Kozak et al., 2015]), indicating that such races can be maintained at early stages of divergence for long evolutionary periods without speciating. In the same way, the hybrid zone between the races of *H. hecale* studied here could be potentially maintained indefinitely under a stable balance between diverging selection on wing coloration and high genetic admixture if environmental conditions remain stable.

Overall, I put my results in a comparative framework and found that similar mechanisms operate on the diversification of the whole clade. Consistent with previous findings, I observed that intrinsic genomic incompatibilities evolve late along this continuum. Namely, no intrinsic genomic incompatibilities exist between the races of *H. hecale*. More surprisingly, hybrids between sympatric species *H. hecale* and *H. ismenius* are viable. In contrast to these instances of speciation with gene flow, the accumulation of barriers in allopatry seem to follow a very different pattern. For instance, I found that the Haldane rule applies to hybrids between *H. hecale melicerta* and *H. hecale clearei*, which are allopatric races of a same species (results not shown). Namely, hybrid females are less viable than males and totally infertile, whereas males are completely fertile. In contrast, “good” species in sympatry do not show hybrid unviability. This indicates that, as expected, strong barriers to gene flow evolve easily at the intraspecific level when geographic isolation exists.

### 4.3 Pervasive hybridisation is not the rule in the genus

Ongoing gene flow at multiple levels of divergence in the genus *Heliconius* has been found ([Bull et al., 2006]; [Kronforst et al., 2006b]; [Martin et al., 2013]). Here, I did not find molecular signals of introgressive hybridisation among the pair of closely related species *H. hecale* and *H. ismenius*, despite their relatively short divergence time. This divergence time might have been long enough for the evolution of very strong reproductive barriers to gene flow, like the chemical signals mentioned above, which act as pre-zygotic isolation barriers. I thus suggest that hybridisation is not ubiquitous in the genus *Heliconius* (consistent with the results of [Krzysztof Kozak, *pers. comm.*]), although a notable number of species in the genus hybridise, mainly in the *melpomene*, *cydno* and *silvaniforms* sub-clades ([Mallet et al., 2007]), the latter of which contains the species studied here.

## Appendix A

### Résumé français

#### A.1 Introduction

Les êtres vivants présentent des variations de forme et de fonction remarquables, variations naturelles qui sont le fondement de la diversification biologique. L'origine de cette incroyable diversité est l'une question centrale en biologie évolutive. Quelles sont les forces qui agissent sur la diversité de couleurs, de formes, de comportements ou d'autres traits ? Comment l'évolution module-t-elle ces traits en sélectionnant les organismes plus adaptés à l'environnement ? Et enfin, comment le processus d'adaptation à des conditions environnementales particulières peut-il donner lieu aux entités discrètes que l'on nomme espèces ? Les nombreuses données disponibles et la combinaison fructueuse de données empiriques et des modèles permet aujourd'hui d'étudier ces questions de manière intégrative et d'explorer comment les gènes, les organismes et l'environnement interagissent dans le processus d'adaptation et de diversification. Cependant, plus de 150 ans après la publication de *The Origin of Species* ([Darwin, 1859]), de nombreux aspects restent mal compris et les recherches se poursuivent pour comprendre les facteurs et les mécanismes de l'évolution.

Pour résoudre ces questions cruciales en biologie évolutive, les hotspots de biodiversité ont été largement étudiés car ces zones comprennent de larges groupes taxonomiques résultant d'une radiation adaptative. Ils fournissent des exemples de choix pour tester les scénarios de diversification. Les néotropiques sont l'une des régions les plus riches du monde en nombre d'espèces. Ce n'est donc pas une surprise si ces régions ont inspiré de célèbres naturalistes et permis le développement des théories de biologie évolutive. L'une des plus remarquables provient des travaux de Henry Walter Bates et Fritz Müller à la fin du XIX<sup>ème</sup> siècle ([Bates, 1862] ; [Müller, 1879]). En explorant les communautés de papillons néotropicaux, ils ont observé que des groupes d'espèces partageaient une ressemblance frappante de patron de couleur alaire, une ressemblance probablement pas due au hasard. Ces auteurs ont proposé une hypothèse explicative en écho aux théories darwinienne. Ils ont suggéré que les motifs alaires étaient utilisés comme un signal de danger pour les prédateurs (notamment les oiseaux), et avaient une fonction de protection. Le mimétisme est cette ressemblance adaptative entre différentes espèces d'une même région géographique, inclus des espèces comestibles mimant des modèles toxiques (mimétisme batesien) ou plusieurs espèces toxiques partageant un patron commun et bénéficiant mutuellement de l'apprentissage des prédateurs (mimétisme müllerien ; voir la Figure 2 pour une description mutuelle du phénomène). Le mimétisme, observé dans une large gamme de taxon incluant les insectes, les amphibiens, les serpents ou les millipèdes ([Brown, 1981] ; [Pfennig et al., 2001] ; [Marek and Bond, 2009] ; [Alexandrou et al., 2011] ;

[Twomey et al., 2013]), module l'adaptation des lignées à un environnement mimétique, appelé cercle mimétique (Figure 2).

Les papillons du genre *Heliconius* (tribu des Heliconiini) sont un exemple classique de mimétisme müllerien. Ces papillons, non comestibles pour les prédateurs, sont célèbres pour leur grande diversité de patron de couleur alaire qui avertit les prédateurs de leur toxicité. Les multiples races géographiques de la quarantaine d'espèces de ce groupe présentent des patrons variés et des relations mimétiques entre espèces coexistant dans une région géographique donnée. Cette convergence phénotypique de mimétisme est soumise à de très fortes pressions de sélection. Par ailleurs, la couleur des ailes est également impliquée dans la reconnaissance de partenaire, ce qui renforce l'importance de ce trait pour l'isolement reproducteur. Ainsi, les papillons du genre *Heliconius* apparaissent comme un modèle idéal pour l'étude de la spéciation, offrant une opportunité unique pour comprendre comment des processus évolutifs de convergence et de diversification agissent. Ce groupe d'espèces représente un laboratoire naturel dans lequel nous pouvons tester les hypothèses de diversification phénotypique et de spéciation.

En conséquence de la combinaison de la convergence locale pour le mimétisme et de la divergence à une large échelle géographique, les papillons du genre *Heliconius* représentent un éventail de lignées différenciés au niveau phénotypique, et qui couvrent le continuum de spéciation, en partant de populations polymorphiques, passant par des races parapatriques et en arrivant jusqu'à des espèces complètement isolées existant en sympatrie ([Nadeau et al., 2012] ; [Martin et al., 2013] ; [Merrill et al., 2011a]). Ces dernières années, nous sommes entrés dans l'ère de la génomique, qui fournit un éclairage nouveau et pertinent sur de nombreuses questions de biologie évolutive. Les recherches sur la diversification d'*Heliconius* ne font pas exception. Le génome d'*Heliconius melpomene* a été publié il y a trois ans ([Heliconius Genome Consortium, 2012]) et a ouvert de nouvelles portes sur la combinaison des approches traditionnelles et de la génomique pour étudier l'évolution de l'isolement reproducteur au long du continuum de spéciation ([Nadeau et al., 2012], [Martin et al., 2013]). En particulier, cette synopse d'approches a permis de cartographier les gènes de la couleur, explorer leurs mécanismes, les forces de sélection auxquelles ils sont soumis, leur rôle dans la spéciation, leur contribution au mimétisme ou leur implication dans des événements d'introgession adaptative ([Reed et al., 2011] ; [Pardo-Diaz et al., 2012] ; [Supple et al., 2013]). Ma thèse tire parti de cette approche intégrée pour tester l'architecture génétique des patrons de couleur et l'implication de ces traits dans différents stades de diversification, en mettant l'accent sur les espèces du groupe des sylvaniformes.

Les recherches sur *Heliconius* ont, jusqu'à maintenant, été centrées sur deux sous-clades assez éloignés phylogénétiquement, *melpomene* et *erato*, qui correspondent aux deux principales branches du genre (Figure 4 ; voir ([Beltran et al., 2007])). D'ailleurs, le cercle mimétique le mieux connu est celui formé par les deux espèces *H. melpomene* et *H. erato*. Ces deux espèces présentent une radiation parallèle remarquable, ayant chacune un grand nombre de patrons mimétiques correspondant à des races géographiques qui imitent parfaitement l'autre espèce localement (Figure 10). Une autre espèce a attiré l'attention des scientifiques, *H. numata*. Celle-ci appartient au sous-groupe des sylvaniformes (ou clade « *numata* ») et est remarquable par son polymorphisme local, absent ou limité chez les autres espèces, et par le contrôle de la couleur via un supergène. La plupart des cercles mimétiques sylvaniformes incluent d'autres groupes de papillons, principalement de la sous-famille des Danainae ou des Ithomiinae, et ces complexes mimétiques multi-clade sont un composant important de l'écologie des néotropiques. Un patron « tigré » typique est composé d'une mosaïque d'éléments noirs, orange et jaunes/blancs et caractérise la plupart des sylvaniformes (à l'exception de *H. beskei* et *H. elevatus*). Etant donné



la complexité de ces patrons, la classification taxonomique des sylvaniformes basée sur le patron de couleur était considérée comme l'une des plus difficiles chez les lépidoptères ([Brown, 1981]). Cette tâche a été considérablement améliorée par les travaux sur le terrain et en insectarium montrant que des adultes très similaires pouvaient être distingués au stade larvaire ([Brown, 1976]). Aujourd'hui, dix espèces sont décrites chez les sylvaniformes et leur relation phylogénétique sont établies dans la phylogénie la plus récente (Figures 4 et 1.4; voir [Kozak et al., 2015]).

Dans ma thèse, j'étudie deux espèces de sylvaniformes peu connues, *H. ismenius*, l'espèce-sœur de *H. numata*, et *H. hecale*, une espèce moyennement éloignée de *H. ismenius*. *H. hecale* présente une large distribution de l'Amérique centrale à l'Amérique du Sud alors que *H. ismenius* est limité au Nord-ouest de l'Amérique du Sud et à l'Amérique centrale, où les deux espèces sont présentes (Figures 11 et 12). *H. hecale* présente une grande variation phénotypique dans son aire de distribution, avec trente races géographiques de patron alaire distinct (Figure 12) alors que *H. ismenius* comprend moins de dix sous-espèces (Figure 11). Malgré cette grande diversité géographique, les deux espèces présentent un polymorphisme local très limité, avec au maximum deux races coexistant dans une localité donnée (par exemple les formes blanches et jaunes de *H. hecale zuleika* au Nord-ouest du Panama ou *H. ismenius telchinia* au Panama central). Dans certaines zones, les races locales de *H. hecale* et *H. ismenius* appartiennent au même cercle mimétique. En dépit de la ressemblance phénotypique des adultes, ces deux espèces peuvent être facilement séparées sur la base de leur plante-hôte et de la morphologie larvaire. Au Panama par exemple, *H. hecale* dépose généralement ses œufs sur *Passiflora vitifolia* alors qu'*H. ismenius* utilise *Passiflora quadrangularis* ou *P. ambigua*. En insectarium, nous avons également observé que les femelles d'*H. hecale* pondent des œufs séparés à l'extrémité de la plante-hôte alors que les femelles d'*H. ismenius* déposent des groupes d'œufs sur des feuilles matures ou immatures, et les larves sont semi-grégaires (voir Figure 13). La morphologie des chenilles diffère au niveau du nombre et de la taille des lignes longitudinales de points noirs (Figure 13).

Mes travaux de thèse exploitent la synergie entre données génomiques, phénotypiques et écologiques pour répondre aux questions posées par la biologie évolutive. En faisant le lien entre différents niveaux d'organisation biologique, du gène aux communautés, je cherche à comprendre comment la diversification biologique avec flux de gènes et comment les barrières à l'hybridation évoluent et fonctionnent. Cette thèse est composée de trois chapitres, schématisés sur la Figure 15. Je cherche à comprendre les facteurs et les mécanismes de la diversification chez *Heliconius* en m'intéressant au cas d'*H. hecale* et *H. ismenius*, deux espèces aux patrons de couleurs complexes, bien différents des espèces traditionnellement étudiées jusqu'à maintenant.

La première partie de ma thèse (Chapitre 1, en vert sur la Figure 15) s'intéresse aux bases génétiques de la diversification biologique et pose les questions suivantes : Dans quelle mesure la diversité phénotypique adaptative est déterminée par des architectures génétiques similaires ? Des phénotypes convergents reposent-ils sur des architectures génétiques communes ? Quels facteurs influencent l'évolution de l'architecture génétique des traits adaptatifs ? J'y explore le conservatisme d'architecture génétique du patron mimétique dans le clade des sylvaniformes afin d'offrir un aperçu général de l'évolution des architectures génétiques impliquées dans la convergence adaptative et la diversification chez *Heliconius* et de fournir un nouvel angle de compréhension de l'origine du supergène de *H. numata*.

Mes résultats contribuent à définir plus précisément le rôle des traits d'effet majeur, comme les patrons de couleur. Ces traits d'effet écologique/phénotypique majeur peuvent-ils conduire seuls à la spéciation ou la spéciation écologique est-elle plutôt un processus multidimensionnel associés à plusieurs traits et plusieurs facteurs écologiques ? Comment des taxons inter-fertiles restent-ils phénotypiquement distincts ? Quelle est l'importance des différentes barrières impli-

quées dans l'isolement reproducteur selon le degré de divergence des espèces ? Comment la divergence génomique se construit-elle au cours du temps ? Dans le chapitre 2 (en orange sur la Figure 15), j'étudie la zone hybride entre *H. hecale melicerta* et *H. hecale zuleika* au Panama afin de tester comment les pressions de sélection agissant sur la coloration alaire contribuent à maintenir les différences phénotypiques malgré l'existence d'un flux de gènes important à ces premiers stades de divergence. En combinant génétique et génomique des populations, théories des clines et des tests comportementaux sur le choix de partenaire par les mâles, je mesure le flux de gènes entre races à l'échelle du génome, j'estime la force d'isolement reproducteur post-zygotique extrinsèque (imposé par la prédation) et j'évalue la force d'isolement reproducteur pré-zygotique dû aux choix des mâles.

Dans le chapitre 3 (en jaune sur la Figure 15), je teste les rôles relatifs de la couleur et de la communication chimique en tant que barrières au flux de gènes à un stade avancé de spéciation. J'utilise des données de génomique associées à des approches d'écologie chimique et des expériences comportementales afin d'évaluer le flux de gènes à l'échelle du génome entre populations naturelles d'*H. hecale*, de tester l'existence d'hybridation interspécifique, d'explorer les paysages de divergence génomique entre « vraies » espèces et de rechercher si les informations chimiques sont impliquées dans le maintien des barrières interspécifiques. Pour cela, je contraste deux situations, l'une où *H. hecale* et *H. ismenius* partage le même patron de couleur (l'est du Panama) et l'autre où les races géographiques de ces deux espèces ont des patrons différents (l'ouest du Panama).

## A.2 Chapitre 1. Du conservatisme et de la nouveauté dans l'architecture génétique de l'adaptation chez les papillons du genre *Heliconius*

Bárbara Huber, Annabel Whibley, Yann Le Poul, Nicolas Navarro, Arnaud Martin, Simon Baxter, Abhijeet Shah, Benoît Gilles, Thierry Wirth, W. Owen McMillan, Mathieu Joron

Arriver à faire le lien entre la diversification phénotypique et l'architecture génétique des traits adaptatifs permet de mieux comprendre les processus de diversification et de spéciation. Pourtant, en contraste avec la grande disponibilité de données écologiques montrant l'importance de la sélection comme une force importante de la divergence des populations et de la spéciation, des données génétiques sous-jacentes à ces processus restent peu abondantes ([Linneen et al., 2009] ; [Butlin et al., 2012]). En particulier, comment la divergence phénotypique et écologique ont lieu au niveau génétique est une question qui n'est pas encore résolue. Les traits écologiquement importants sont souvent complexes et sont contrôlés par des groupes de gènes et ses interactions, ce qui constitue l'architecture génétique de ces traits. Le terme architecture génétique comprend le nombre et la position génomique des gènes qui contribuent à un trait, la taille de ses effets phénotypiques, l'interaction de ses allèles (i.e. dominance), la façon dont les gènes interagissent pour moduler l'effet d'autres gènes (i.e. épistasie et additivité), la contribution d'un gène à des traits différents (i.e. pléiotropie) et les interactions des gènes avec l'environnement (i.e. épigénétique).

L'architecture génétique détermine le potentiel d'un trait à varier et en conséquence, à évoluer (voir [Hansen, 2006]). Donc, la compréhension de l'évolvabilité d'un trait peut être seulement achevée dans un cadre architectural. Pourtant, une brèche existe entre la riche théorie mathématique et la croissante masse empirique concernant la base génétique de l'adaptation ([Orr, 2005]). En particulier, l'évaluation de comment l'architecture génétique des traits adaptatifs influence la diversification de ces traits a été ralentie par le manque de données empiriques.

Au long de la dernière décennie, l'accès aux données génomiques d'haut débit ont fourni une puissance accrue pour l'identification des cibles de la sélection, c'est-à-dire des loci, des gènes et des variantes génétiques qui contrôlent des phénotypes adaptatifs ([Stinchcombe and Hoekstra, 2007]). Ces outils ont élargi les frontières au-delà des espèces modèles et ont permis, en particulier, de mieux comprendre l'évolution convergente (c.f. [Arendt and Reznick, 2008]), où le même phénotype évolue au sein de deux ou plus lignées de façon indépendante, couramment en réponse à des défis environnementaux similaires. Des systèmes naturels montrant de la convergence phénotypique constituent des cadres robustes pour évaluer si cette convergence est l'issue du recrutement des mêmes ou des différents gènes ou mécanismes génétiques, permettant ainsi une meilleure compréhension de la base moléculaire de l'évolution adaptative ([Stern, 2013]).

Le mimétisme est un exemple exceptionnel d'évolution convergente. Ici, nous avons étudié l'évolution de l'architecture génétique sous-jacente à la diversité des motifs alaires chez *Heliconius*, un genre de papillons qui a subi une radiation adaptative pour le mimétisme basé sur les couleurs des ailes. Les patrons alaires des *Heliconius*, qui sont aposématifs et qui avertissent les prédateurs sur la toxicité de ces papillons, se distinguent pas seulement par sa diversité frappante, mais aussi parce que sa complexité est à une base génétique plutôt simple. Des grands efforts ont été faits pour cartographier les gènes responsables du mimétisme dans ce clade (voir Box 6), puisque cela peut révéler quel est le matériel brut pour l'adaptation et comment la sélection agit sur ce matériel. Dans la plupart des espèces d'*Heliconius* étudiées jusqu'au présent, quelques peu de loci d'effet majeur et d'hérédité mendélienne contrôlent des éléments de la couleur alaire séparés; des variantes de ces éléments apparaissant ensemble sur l'aile générant un patron mimétique donné. Ce groupe de gènes, appelé le toolbox ou toolkit de la coloration alaire, est distribué au long de 4 des 21 chromosomes qui caractérisent le genre ([Joron et al., 2006]; [Kronforst et al., 2006a]; [Papa et al., 2008]; [Kronforst et al., 2006a]). Pourtant, des QTLs d'effet mineur ont aussi été identifiés (voir Table 1.1 pour un résumé). À ce jour, le gros de notre connaissance de l'architecture de la variation des patrons de couleur chez *Heliconius* provient de l'étude de quelques espèces montrant des éléments rouges, blancs et jaunes dans un contexte noir ([Jiggins and McMillan, 1997]; [Jiggins et al., 2005b]; [Kronforst et al., 2006a]; [Reed et al., 2011]; [Nadeau et al., 2014]). En revanche, la base génétique des patrons dits « tigrés » des sylvaniformes a juste été caractérisée dans l'espèce *H. numata* ([Joron et al., 2006]). Comparé à l'architecture à plusieurs loci décrite plus haut, cette espèce montre une architecture de la couleur très différente. Notamment, un supergène, c'est-à-dire un ensemble de plusieurs gènes fortement liés, monopolise le contrôle de la variation de la couleur alaire de cette espèce.

Les régimes de sélection pour le mimétisme sont largement déterminés par la distribution et l'abondance des différents signaux d'avertissement montrés par les communautés locales de proies. La plupart des espèces d'*Heliconius*, y incluses les espèces ayant des patrons tigrés, montrent des races géographiques différenciées au niveau des couleurs alaires, en réponse à la sélection directionnelle favorisant seulement un patron protégé dans une localité ([Brown, 1981]). En revanche, *H. numata* montre un polymorphisme local très riche, et toutes les populations de cette espèce abritent des formes distinctes qui imitent des multiples espèces ayant des patrons tigrés ([Brown and Benson, 1974]). Ce polymorphisme est probablement maintenu grâce à la sélection balancée agissant au niveau local. L'hétérogénéité dans le régime de sélection façonnant les patrons alaires chez *Heliconius* permet d'étudier la relation entre les régimes de sélection et l'évolution des architectures génétiques distinctes sous-jacentes aux traits adaptatifs complexes.

Ici, nous nous sommes concentrés sur le clade des sylvaniformes et avons examiné si l'architecture génétique des motifs colorés des ailes est-elle associée avec la variation phénotypique

elle-même ou avec le régime de sélection qui modèle cette variation. Pour achever ce but, nous avons analysé l'hérédité de la couleur alaire chez deux espèces de sylvaniformes, notamment *H. hecale* et *H. ismenius*. Nous avons combiné la cartographie génétique classique complémentée par du séquençage haut débit avec des analyses multivariées de génétique quantitative et du génotypage pour des gènes candidats. Cette combinaison d'approches nous a permis d'identifier les régions génomiques qui contrôlent la variation de coloration alaire chez ces deux espèces et pour explorer l'évolution des architectures génétiques dans un cadre comparatif plus large.

Nous avons commencé par faire trois types de rétrocroisements intraspécifiques entre des races géographiques de *H. hecale* et *H. ismenius* différentes pour les motifs colorés alaires (Figure 1.1). Nous avons obtenu des familles nombreuses surtout pour les croisements de *H. hecale*, mais pas tellement pour ceux concernant des races panamiennes de *H. ismenius*. Nous avons en suite génotypé une partie de ces familles avec du séquençage de l'ADN associé à des sites de restriction (*RAD sequencing*), et avons analysé ces données massives avec des logiciels de bioinformatique divers. Nous avons ainsi pu reconstruire des chartes chromosomiques denses pour les trois types de croisements faits.

Nous avons ensuite analysé la ségrégation des allèles pour les différents loci d'hérédité mendélienne contrôlant des éléments de la couleur alaire et les avons cartographiés par cartographie de liaison. Nous avons également utilisé des analyses morphométriques multivariées profitant d'une description de l'espace morphométrique complet grâce à un outil développé au sein de notre équipe (Colour Pattern Modelling tool or CPM; [Le Poul et al., 2014]; voir Figure 1.3). Cette description ne repose pas sur la détection subjective des éléments variantes et permet de visualiser des heatmaps d'association entre le génotype et le phénotype pour chaque locus de caractères quantitatifs (QTL) séparément. En mettant en œuvre cet ensemble d'analyses, nous avons caractérisé l'architecture génétique de la variation mimétique chez les deux espèces explorées (Figure 1.4). Nous avons trouvé une architecture à plusieurs loci de la couleur non liés pour le contrôle des variations de couleur chez ces espèces, indiquant que l'architecture à un seul gène (le supergène) est limitée à *H. numata* et est probablement un état dérivé. L'architecture à plusieurs loci, en revanche, est répandue dans le genre et paraît être ancestral.

Les régions de la couleur que nous avons cartographiés sont homologues aux composants génétiques de la toolbox retrouvé ailleurs dans le genre (Figure 1.4). Cela indique que des architectures génétiques conservées peuvent être à l'origine des phénotypes très différents. Aussi, des architectures génétiques exceptionnellement distinctes (par rapport à l'effet individuel des différents loci) peuvent être à la base des phénotypes très similaires et même mimétiques. En conséquence, nous concluons que l'évolution de l'architecture génétique des traits adaptatifs répond aux régimes de sélection qui façonnent la variation de ces traits, mais n'est pas lié intrinsèquement à la variation phénotypique elle-même.

### A.3 Chapitre 2. La zone d'hybridation entre deux races de *Heliconius* *hecale* : une exploration du rôle de la couleur alaire dans des stades précoces de diversification

Bárbara Huber, Annabel Whibley, Benjamin Rice, Francisco Delgado, Luis Murillo, Adriana Tapia, Thierry Wirth, W. Owen McMillan, Violaine Llaurens, Mathieu Joron

Les zones d'hybridation offrent la possibilité d'étudier la spéciation et d'explorer l'évolution des barrières d'isolement reproducteur entre des lignées divergentes mais qui s'hybrident encore. Comment ces barrières peuvent-elles évoluer au sein d'une population essentiellement continue et comment arrivent-elles à maintenir les limites entre des lignées divergées ? Comment

les différents facteurs prézygotiques et postzygotiques interagissent-ils pour promouvoir la diversification ? Et de façon plus générale, quels sont les mécanismes qui permettent que certains traits adaptatifs mènent à la radiation d'un clade ? Ce sont des questions qui ont été largement étudiées, mais qui restent encore sans réponse claire ([Coyne and Orr, 2004]). Même si les mécanismes évolutifs qui maintiennent et favorisent la diversification vis-à-vis du flux de gènes varient selon le taxon étudié, la sélection naturelle et la sélection sexuelle semblent jouer un rôle central dans ce processus ([Coyne and Orr, 1998] ; [Schemske, 2000] ; [Schluter, 2000] ; [Malausa et al., 2005] ; [Via, 2009]).

Les zones hybrides illustrent différents stades d'isolement reproducteur ([Barton and Hewitt, 1989] ; [Szymura and Barton, 1991] ; [Harrison, 1993] ; [Veen et al., 2001] ; leur étude a contribué à la description de la spéciation comme étant un processus graduel et cumulatif ([Jiggins and Mallet, 2000]). La description des différents stades du continuum de spéciation au sein d'un clade permet d'explorer comment les différents facteurs écologiques et génétiques s'accumulent et contribuent à l'isolement reproducteur entre des populations ([Hendry et al., 2009] ; [Powell et al., 2013]). Ici, nous nous sommes intéressés aux papillons néotropicaux du genre *Heliconius*, qui illustrent des stades variés de ce continuum et qui ont subi une radiation adaptative des couleurs des ailes qui sont aposématiques et agissent comme des signaux d'avertissement de la toxicité de ces papillons ([Brown, 1981] ; [Mallet, 1993]). Plusieurs espèces d'*Heliconius* et d'autres espèces toxiques partagent localement les mêmes patrons de coloration alaire (i. e. mimétisme müllerien), ce qui décroît le risque de prédation et donc est à la base de forces sélectives fortes sur la couleur alaire au niveau géographique locale.

La plupart des espèces d'*Heliconius* montrent une mosaïque de races parapatriques à motifs alaires différents, chacune d'elles participant à des communautés locales de papillons toxiques et mimétiques, ce que l'on appelle des cercles mimétiques. Des zones d'hybridation existent entre ces races et leur largeur est surtout dépendante de la sélection disruptive agissant contre les papillons montrant des patrons de couleur intermédiaires entre les types parentaux, qui sont eux reconnus par les prédateurs ([Mallet, 1986] ; [Jiggins et al., 1997] ; [Blum, 2008] ; [Arias et al., 2012] ; voir Table 2.4). Les couleurs des ailes ne sont pas seulement soumises à la sélection disruptive forte due à la sélection contre les individus hybrides (en étant des barrières d'isolement postzygotique ; [Merrill et al., 2012]), mais elles sont aussi essentielles comme des signaux impliqués dans le comportement de cour (en étant des barrières pré-zygotiques au flux de gènes ; [Jiggins et al., 2001] ; [Merrill et al., 2011a]). En conséquence, les motifs de coloration des ailes sont considérés comme des traits *magiques* (d'après [Servedio et al., 2011]) pouvant faciliter la spéciation écologique ([Gavrilets, 2004] ; [Jiggins et al., 2005a], [Jiggins, 2008]).

Pourtant, même si les patrons de coloration sont un facteur important dans la diversification du clade *Heliconius* ([Mallet et al., 1998a]), il reste à déterminer si la sélection diversifiante agissant sur ces traits est suffisante pour mener à l'achèvement du processus de la spéciation. L'étude des zones d'hybridation interraciales chez les papillons du genre *Heliconius* permet de rechercher comment les barrières au flux de gènes s'accumulent au long de la spéciation, comment la sélection naturelle disruptive et la sélection sexuelle sur la coloration alaire agissent ensemble et comment ces forces sélectives contrebalancent le flux de gènes pendant des stades précoces de diversification.

Ici, nous avons exploré la zone d'hybridation entre les races panamiennes *H. hecale melicerta* et *H. hecale zuleika* qui ont des motifs colorés différents. L'espèce *H. hecale* fait parti du clade des sylvaniformes et, à notre connaissance, aucune zone d'hybridation impliquant des sylvaniformes n'a été étudiée jusqu'à présent. Etant donnée que les sylvaniformes appartiennent à des cercles mimétiques plus complexes et plus riches en espèces que d'autres espèces dans le genre mieux connues (par exemple *H. melpomene* et *H. erato*), l'étude de la zone d'hybridation entre des races

de *H. hecale* permet d'explorer dans un cadre comparatif à quel point la complexité des patrons de coloration et la composition des communautés mimétiques déterminent la structure et la dynamique des zones d'hybridation dans le genre *Heliconius*.

Nous avons profité du fait que l'architecture génétique des motifs colorés chez *H. hecale* est connue (voir Figure 2.2 ; [Huber et al., 2015]). Nous avons mis en œuvre des approches de génétique et génomique des populations, utilisé la théorie des clines et l'analyse du comportement de choix de partenaire, pour caractériser le stade de divergence entre les races de *H. hecale* et le placer dans un contexte comparatif dans le continuum de spéciation dans le genre *Heliconius*. En cherchant à avoir un regard global de cette zone hybride, nous avons exploré comment la sélection contribue à la maintenance de la divergence phénotypique en présence du flux de gènes dans des stades précoces de divergence.

Nous avons commencé par évaluer le niveau de flux de gènes entre les races parapatriques de *H. hecale* et n'avons retrouvé aucune différenciation génétique entre des populations quasiment phénotypiquement pures situées aux extrêmes de la zone hybride (Figure 2.1), que ce soit en utilisant des marqueurs hypervariables du type microsatellites ou en analysant des marqueurs à haute densité provenant du re-séquençage de génomes entiers (Figure 2.3 et Figure 2.4A). Le haut niveau de flux de gènes retrouvé sur l'ensemble du génome contraste avec des pics de différenciation ponctuels parsemés le long du génome (Figure 2.4A). Nous avons exploré si les régions génomiques contenant les gènes de la couleur étaient plus divergentes que le reste du génome, ce qui serait le résultat de pressions sélectives disruptives pour l'adaptation aux milieux mimétiques locaux. Nous avons trouvé que quelques uns de ces pics de différenciation sont effectivement associés à des gènes de la couleur des ailes (Figure 2.4B). Notamment, le pic de divergence le plus haut est retrouvé autour de la région cis-régulatrice du gène *optix*, qui contrôle des motifs alaires rouges chez plusieurs espèces d'*Heliconius* ([Reed et al., 2011] ; [Martin et al., 2014] ; [Reed et al., 2011] ; [Supple et al., 2013]), et aussi chez *H. hecale* ([Huber et al., 2015]). Un autre pic de différenciation est situé sur la région régulatrice putative du gène *WntA*, qui contrôle des éléments mélaniques chez *H. hecale* et d'autres espèces ([Martin et al., 2012] ; [Gallant et al., 2014] ; [Martin and Reed, 2014] ; [Huber et al., 2015]). . Ensuite, nous avons exploré la force de la sélection naturelle agissant sur les loci de la couleur sur le terrain. Notamment, nous avons analysé la zone de transition entre les races panamiennes d' *H. hecale*, en traçant des clines phénotypiques pour chacun des trois loci qui définissent la différence de coloration alaire entre ces deux races (Figure 2.5). Cela nous a permis d'évaluer la force de la sélection disruptive pour le mimétisme, qui est inversement proportionnelle à la largeur du cline et qui maintient la divergence phénotypique entre les deux races en face d'un flux de gènes très répandu. Pour cela, nous avons bénéficié de la connaissance de la ségrégation des allèles pour chaque locus et d'un échantillonnage de 333 spécimens provenant de sept populations le long d'un axe ouest-est couvrant 800 km depuis San José (Costa Rica) jusqu'à Darién (Panama) (Figure 2.1). Nous avons retrouvé que cette zone hybride a une largeur de 92 km (Figure 2.5A), c'est à dire qu'elle a une dimension moyenne quand elle est comparée à d'autres zones hybrides étudiées dans le genre ([Mallet et al., 1998a]). Cela peut-être interprété comme étant le résultat de forces sélectives plutôt faibles qui sont possiblement dues à la complexité des patrons alaires tigrés et des communautés mimétiques auxquels ces papillons appartiennent (Figure 2.7). Nous avons fait une comparaison des clines et avons trouvé une différence significative pour la largeur de ces clines (Table 2.1 et Figure 2.5B). Nous suggérons que cette différence peut être expliquée par des pressions sélectives d'intensité variable entre les différents éléments composant les patrons de couleur alaire. Cette variation pourrait refléter des variations dans les communautés mimétiques et dans la capacité d'apprentissage des prédateurs.

Finalement, nous avons testé si les mâles des deux races d'*H. hecale* exercent un choix de

partenaire sélectif qui pourrait fortifier ou maintenir le niveau de divergence entre ces races. Nous avons trouvé que les mâles d'une des deux races (*H. h. melicerta*) préfèrent faire la cour aux femelles de sa propre race (mais cette préférence est faible ; voir Figure 2.6). Par contre, les mâles de l'autre race (à savoir *H. h. zuleika*) ne montrent pas de préférence (Figure 2.6). Nous faisons l'hypothèse que ce choix asymétrique basé sur les patrons de couleur, bien que faible, pourrait être l'un des facteurs en jeu qui mènerait à la mobilité de cette zone d'hybridation, qui semble être en train d'être déplacée vers l'ouest ([Mallet, 1993] ; [Rosser et al., 2012]).

Dans l'ensemble, la distribution clinale des allèles des gènes de la couleur et l'évidence d'un flux de gène restreint au niveau de ces gènes, contrastent avec l'absence de structure génétique dans l'ensemble du génome. Cela confirme que ces gènes contrôlent des traits directement impliqués dans l'adaptation locale, ce qui est à la fois concordant avec les signaux moléculaires liés à ces gènes qui ont été détectés dans cette étude et dans des travaux antérieurs ([Nadeau et al., 2012] ; [Supple et al., 2013] ; [Martin et al., 2013] ; [Kronforst et al., 2013]). Les pressions sélectives détectées pourraient pourtant ne pas être assez fortes pour mener ces races parapatriques à des stades plus avancés de divergence. Cette zone hybride pourrait se retrouver a contrario dans un équilibre stable de migration-sélection.

#### A.4 Chapitre 3. Des espèces apparentées et sympatriques avec le même déguisement : un aperçu sur la reconnaissance de l'espèce entre *Heliconius hecale* et *H. ismenius*

Bárbara Huber, Annabel Whibley, Catalina Estrada, Darha Solano, Robert Orpet, W. Owen McMillan, Brigitte Frérot, Mathieu Joron

L'introduction d'un concept d'espèce plus relaxé que le concept biologique d'espèce de Mayr (e.g. [Mallet, 1995]) permet de voir la spéciation comme étant un processus où la divergence s'accumule de différentes façons, dont la divergence neutre, l'adaptation locale et la coévolution, mais où l'hybridation et l'introggression ont lieu de façon presque inévitable dans la plupart des événements de spéciation, à exception de la spéciation allopatrique stricte ([Coyne and Orr, 2004]). Malgré la perméabilité de la limite entre espèces, celles-ci constituent des entités discrètes qui peuvent être distinguées d'un point de vue morphologique, génétique, écologique ou comportemental (e.g. [Noor et al., 2000] ; [Saint-Laurent et al., 2003] ; [Kraus et al., 2012]). Un consensus existe actuellement sur l'importance de l'hybridation dans la diversification ([Stebbins, 1959] ; [Arnold, 1992] ; [Dowling et al., 1997] ; [Seehausen, 2004] ; [Barton, 2013] ; [Abbott et al., 2013]). Pourtant, les façons dont les différences entre des populations s'accumulent et dont la limite entre des espèces peut être maintenue malgré l'effet homogénéisateur du flux génique restent mal comprises. Quels traits agissent comme des barrières réelles ou potentielles contre le flux de gènes entre des lignées coexistantes ? Quelle est l'importance relative de certaines barrières par rapport à d'autres ? L'abord de ces questions plus spécifiques contribue à la compréhension de la diversification biologique.

En particulier, la caractérisation du flux génique entre des lignées à différents stades du continuum de spéciation est importante pour comprendre l'histoire évolutive de l'isolement reproducteur (e.g. [Hendry et al., 2009] ; [Powell et al., 2013]). Les papillons du genre *Heliconius* constituent une collection de lignées étalées au long du continuum de spéciation commençant par des populations polymorphiques et finissant avec des espèces sympatriques totalement isolées. L'accouplement sélectif basé sur les patrons de coloration augmente au long de ce continuum ([Merrill et al., 2011a]). De plus, des sondages génomiques de différenciation entre des espèces et races différentes d'*Heliconius* ont fourni un aperçu du niveau de flux génique, en

plus du taux et de la dynamique de la divergence génomique au long du continuum de spéciation ([Nadeau et al., 2012] ; [Hill et al., 2013] ; [Martin et al., 2013] ; [Kronforst et al., 2013]). Par exemple, [Martin et al., 2013] ont trouvé que le flux génique a persisté au long des ~2 millions d'années de divergence entre les espèces sympatriques *Heliconius melpomene* et *H. cydno*, qui restent néanmoins de « bonnes » espèces, montrant des écologies distinctes et des fortes barrières d'isolement reproducteur au niveau prézygotique comme au niveau postzygotique ([Jiggins et al., 2001] ; [Naisbit et al., 2002] ; [Merrill et al., 2012]). Ces résultats confirment le flux génique répandu entre des espèces divergentes, malgré la rareté des hybrides rencontrés dans la nature ([Gilbert, 2003] ; [Bull et al., 2006] ; [Kronforst et al., 2006b] ; [Mallet et al., 2007] ; [Mallet, 2009] ; [Pardo-Diaz et al., 2012] ; [Heliconius Genome Consortium, 2012] ; [Martin et al., 2013]). Les papillons du genre *Heliconius* sont ainsi un modèle idéal pour étudier comment la divergence a lieu en présence du flux génique.

Les motifs colorés alaires chez *Heliconius* sont importants pour l'isolement sexuel opéré par le mâle, mais sont aussi impliqués dans la sélection contre les hybrides qui ne sont pas mimétiques ; c'est pour cela que ces patrons de couleur ont été couramment considérés comme étant des « traits magiques » agissant contre l'hybridation et facilitant la spéciation chez ce groupe de papillons. Au contraire, la part du choix de la femelle basé sur la coloration alaire n'a pas été testée (à l'exception de [Boyko, 2005]). En outre, le rôle des signaux chimiques dans l'isolement reproducteur prézygotique pouvant promouvoir la diversification et la spéciation chez *Heliconius* n'avait pas été exploré jusqu'à récemment ([Mérot et al., 2015] ; [Estrada et al., 2011]), malgré son importance chez plusieurs espèces dans l'ordre Lepidoptera, où ces signaux sont impliqués dans la communication intraspécifique et interspécifique, ainsi que dans la sélection sexuelle ([Vane-Wright and Boppre, 1993] ; [Millar, 2000] ; [Jurenka et al., 2003] ; [Schulz et al., 2008] ; [Nieberding et al., 2008]).

Ici, nous avons examiné si l'introgression est un point commun à tout le clade des *Heliconius*. Nous nous sommes concentrés sur les espèces proches *Heliconius hecale* et *H. ismenius*, qui ont divergé il y a environ 3 millions d'années ([Kozak et al., 2015]) et qui se chevauchent géographiquement dans le nord-ouest de l'Amérique du Sud et dans le sud de l'Amérique Centrale. Cette étude a permis d'évaluer le niveau de flux génique à un stade plus avancé de la spéciation que ce qui avait été étudié précédemment. En particulier, nous basant sur des données génomiques dérivées du re-séquençage de génomes entiers, nous avons complété des études précédentes dans ce domaine, lesquelles ont impliqué des stades précoces de divergence, pour explorer comment l'isolement reproducteur et ses traces au niveau génomique s'accumulent au long du processus de la spéciation. L'estimation des statistiques de divergence génomique a montré une faible évidence d'introgression interspécifique entre *Heliconius hecale* et *H. ismenius*. Notamment, une analyse de mélange génétique avec le programme *Structure* a montré une différenciation génétique complète entre les deux espèces de sylvaniformes (Figure 3.2A).

Aussi, nous avons trouvé des valeurs de  $F_{st}$  et  $d_{XY}$  (deux mesures distinctes de différenciation génétique) très élevées par rapport à d'autres comparaisons dans le genre. Nos résultats contrastent ainsi avec la découverte d'un flux génique fort et généralisé entre des taxons appartenant aux sous-clades *melpomene* et *cydno* ([Bull et al., 2006] ; [Kronforst et al., 2006b] ; [Martin et al., 2013]).

En jouant un rôle dans la survie et dans le choix de partenaire, des décalages mimétiques favorisent la spéciation ([Jiggins et al., 2001] ; [Mallet, 2010]). En fait, des espèces sœurs participent souvent à des cercles mimétiques différents (mais regardez [Brower, 1996] ; [Giraldo et al., 2008] ; [Mérot et al., 2013]). Pourtant, les espèces proches *H. hecale* et *H. ismenius* partagent elles aussi les mêmes patrons de coloration dans des régions particulières de sa distribution géographique. Dans l'est du Panama, et dans l'ouest de la Colombie et du Venezuela,



*H. h. melicerta* et *H. i. bouletti* sont extrêmement similaires et rejoignent le même cercle mimétique avec plusieurs autres espèces. En revanche, dans l'ouest du Panama, les deux espèces (*H. h. zuleika* et *H. i. clarescens*) ont des patrons distincts, chacune appartenant à un cercle mimétique différent (Figure 3.1). Ici, nous avons utilisé des approches génomiques, comportementales et chimiques pour tester l'hypothèse que la limite entre les espèces est plus perméable au flux génique quand les deux espèces partagent les mêmes motifs colorés. Aussi, nous avons examiné quels signaux de reconnaissance pourraient être en train d'agir pour maintenir ces espèces séparées.

Nous avons mis en œuvre un test pertinent, notamment une analyse nommée ABBA-BABA test ([Green et al., 2010] ; [Durand et al., 2011] ; voir Figure 3.4A) basé ici sur le génotypage de l'ensemble du génome, et n'avons pas retrouvé d'évidence d'un flux génique plus important entre la paire mimétique *H. h. melicerta*/*H. i. bouletti* qu'entre la paire non mimétique *H. h. zuleika*/*H. i. clarescens*. Notamment, les statistiques *Patterson's D* et *f* n'ont pas montré un excès des sites du type ABBA en relation à des sites du type BABA dans aucune des deux configurations testées. Nous avons cependant prouvé que des événements ponctuels d'introgession ont eu lieu à un moment donné de l'histoire de divergence entre les lignées *H. hecale* et *H. ismenius*. En regardant la distribution des deux types de sites polymorphiques tout au long du génome, des biais locaux ont été détectés. Spécifiquement, en testant le partage d'allèles entre la paire mimétique de races (voir la configuration dans le haut de la Figure 3.4A) nous avons observé un excès frappant de sites ABBA en relation à des sites BABA, lié au gène de la couleur *optix* (Figure 3.4B and Figure 3.4C), et également associé à une baisse dans la valeur de  $d_{XY}$  (Figure 3.4C)). Cela indique que l'introgession adaptative joue un rôle clé dans le contexte de l'évolution du mimétisme, tout en soulignant l'importance des événements occasionnels d'hybridation pour l'évolution phénotypique et pour l'écologie des sylvaniformes, qui ne conduisent néanmoins pas à des frontières plus perméables entre ces espèces.

Globalement, nos analyses ne plaident pas en faveur d'un effet significatif du mimétisme sur le taux de partage d'allèles au niveau du génome. Cette observation, en plus des expériences de comportement menées, suggère que des signaux différents de la couleur alaire déterminent un isolement sexuel fort entre *H. hecale* et *H. ismenius*. Pour explorer la nature et la portée de ces signaux, nous avons commencé par examiner le comportement de cour de l'espèce *H. hecale*, et l'avons décrit comme étant composé par cinq étapes principales (Figure 3.5). Nous avons ensuite conduit des expériences visant à explorer le comportement sexuel entre la paire mimétique *H. h. melicerta* et *H. i. bouletti*, et avons observé que les mâles des deux espèces sont fortement attirés par les femelles et par les modèles alaires de l'autre espèce (Figure 3.6), ce qui montre que la confusion hétérospécifique opérée par le mâle peut avoir lieu dans la longue portée, en accord avec les résultats de [Estrada and Jiggins, 2008] sur *H. melpomene* et *H. erato*. Nous avons cependant observé que la reconnaissance de l'espèce s'inscrit dans la courte portée pour les mâles aussi bien que pour les femelles, suggérant que des signaux chimiques pourraient être différenciés entre *H. hecale* et *H. ismenius*.

Nous avons donc testé si des différences existaient entre ces espèces au niveau des bouquets de composés chimiques produits par trois différents tissus potentiellement impliqués dans la production de phéromones, notamment la zone de chevauchement entre les ailes, les glandes abdominales et le tégument à la base des ailes. Nous avons procédé à l'extraction de ces composés avec de l'hexane, et nous avons passé les extraits par chromatographie en phase gazeuse couplée à la spectrométrie de masse (GC-MS). Nous avons ainsi retrouvé des différences chimiques importantes entre les espèces étudiées (voir Figures 3.9, 3.10 et 3.11). Nous faisons l'hypothèse que quelques uns des composés différenciés pourraient servir de médiateurs pour le choix de partenaire et la reconnaissance de l'espèce. De façon plus spécifique, nous avons trouvé

que ce sont les mâles qui montrent les cocktails chimiques les plus complexes et les plus différents entre espèces un niveau de l'abondance des composés, ce qui peut-être observé comme une séparation significative dans l'espace multivarié (voir Figures 3.9B et 3.10B). Cette forte différence est évidente au niveau des ailes et des claspers des mâles mais pas au niveau de la cuticule (Figure 3.11). Au contraire, les femelles présentent des mélanges pauvres en composés et plutôt similaires entre espèces. La grande différence compositionnelle observée chez les mâles suggère que la reconnaissance de l'espèce est opérée de façon importante par le choix des femelles basé sur des signaux chimiques. Dans un clade où le choix de partenaire est largement provoqué par la divergence des patrons de coloration alaire, nos résultats montrent que cette divergence toute seule n'est pas suffisante pour mener à la spéciation, et que l'accumulation d'autres barrières au flux génique peut être importante. Nos données montrent que des signatures chimiques peuvent être décisives pour la maintenance des limites entre espèces d'*Heliconius*, et qu'elles constituent l'une des barrières contre l'hybridation interspécifique. Pourtant, il nous manque encore le lien fonctionnel entre les cocktails chimiques spécifiques aux espèces et le choix de partenaire. Donc, l'identification de ces composés est nécessaire pour pouvoir évaluer son vrai rôle dans la reconnaissance de l'espèce. Aussi, bien que les facteurs impliqués dans le choix de partenaire soient essentiels à la spéciation en présence de flux de gènes, des adaptations écologiques, y compris celles liées au microhabitat, ont aussi leur importance et ont sûrement contribué à la divergence entre *H. hecale* et *H. ismenius* et à l'atteinte d'un stade de spéciation avancé où le flux génique interspécifique ne peut pas être détecté en utilisant une approche génomique.

## A.5 Conclusions

### A.5.1 C'est la sélection qui a le dernier mot

J'ai étudié l'architecture génétique et écologique de la diversification phénotypique des papillons, en contrastant des cas de divergence phénotypique au sein d'une espèce avec des cas de convergence et de divergence phénotypique impliquant des différentes espèces. J'ai montré que l'architecture génétique de la coloration de l'aile est à la fois conservée et flexible, et je suggère que cette polyvalence pourrait être la raison de sa réponse à la sélection.

En cartographiant des loci contrôlant la couleur des ailes chez les espèces de sylvaniformes *H. hecale* et *H. ismenius*, et en plaçant mes résultats dans un cadre comparatif, j'ai confirmé des attendus théoriques suggérant que l'architecture génétique d'un trait évolue et qu'elle n'est pas une caractéristique constante ([Lande, 1980] ; [Barton, 1995] ; [Coyne and Orr, 1998] ; [Carter et al., 2005] ; [Hansen, 2006]). En particulier, j'ai trouvé que la taille de l'effet des gènes, et non seulement les fréquences des gènes, sont des variables évolutives ([Hansen, 2006]). En outre, j'ai suggéré que l'évolution de l'architecture génétique des caractères adaptatifs est associée avec le régime de sélection agissant sur ces traits. Plus précisément, j'ai trouvé qu'une architecture à plusieurs gènes comme celle existant dans les deux espèces étudiées ici ainsi que dans la plupart des autres espèces d'*Heliconius*, semble être l'état ancestral. Cette architecture est associée à la sélection directionnelle pour le mimétisme local et conduit à un monomorphisme local. En revanche, l'architecture à un seul locus (c'est-à-dire un supergène), est restreinte à *H. numata* et est apparemment l'état dérivé. Le supergène est associée à une sélection équilibrante, permettant la coexistence de plusieurs formes colorées localement.

Ces résultats permettent d'avoir un aperçu de l'évolution des supergènes ([Nabours, 1933] ; [Ford, 1966] ; [Turner, 1977]). Plutôt que de dériver de la liaison serrée des régions génétiques réunies à partir de parties distinctes du génome par des translocations ([Nabours, 1933] ; [Ford, 1966]) j'ai trouvé des preuves de l'hypothèse nommée the sieve hypothesis ([Turner, 1977]),

selon laquelle ce qui donne lieu à des supergènes c'est le recrutement de loci qui sont déjà étroitement liés et capables de produire ensemble des effets phénotypiques majeurs. Dans la région homologue au supergène *P* chez *H. numata*, j'ai trouvé des loci étant spatialement liés et ayant des effets phénotypiques presque égaux chez *H. hecale* (Figure 4.1). Un lien étroit entre ces loci pourrait avoir résulté de la suppression de la recombinaison chez *H. numata* grâce à des inversions chromosomiques ([Joron et al., 2011]). De plus, le grand effet du supergène *P* (Figure 4.1), pourrait hypothétiquement être dû à l'agrégation de plusieurs mutations à effet faible (voir les points sur la Figure 4.1) sur les loci co-adaptés plutôt que par des mutations à grand effet. Ce serait cohérent avec les observations de [Martin and Orgogozo, 2013] sur un grand nombre de cas documentés pour des variantes à grand-effet. Chez *H. hecale* et *H. ismenius* j'ai également cartographié des loci de couleur à grand-effet (ayant pourtant un effet individuel qui contraste fortement avec celui du supergène *P*; voir Figure 4.1) dont l'effet phénotypique peut aussi être potentiellement alimenté par de petites mutations et réglé par l'effet faible de plusieurs autres loci pour la convergence mimétique (le *two-step model*, voir Introduction; [Turner, 1985]; [Orr, 1998]).

Comprendre si l'évolution convergente est reflétée par la convergence au niveau du génotype est l'une des principales questions en génétique évolutive. Les tentatives pour répondre à cette question ont permis d'évaluer la correspondance du phénotype au génotype à des niveaux fonctionnels tels que la voie génétique, le gène et le nucléotide (Figure 8; e.g. ([Protas et al., 2006]; [Hoekstra, 2006]; [Arendt and Reznick, 2008]; [Manceau et al., 2010])). Ici j'ai trouvé que la convergence phénotypique peut être due à des architectures génétiques étonnamment distinctes en termes de la taille de l'effet de gènes, ce qui n'a pas été, à ma connaissance, explicitement rapporté dans la littérature. Notamment, les architectures à plusieurs loci et à un seul supergène sont à la fois capables de contrôler des patrons alaires mimétiques et non-mimétiques mais similaires, ce qui met en évidence la puissance de la sélection pour la ressemblance mimétique. Cette observation soutient l'idée que l'évolution convergente découle de la capacité de la sélection naturelle à résoudre des problèmes écologiques similaires à partir de points de départ différents ([Losos, 2011]). Les caractères liés à la pigmentation semblent plus enclins à entraîner l'évolution convergente à partir d'une base génétique différente, compte tenu des plusieurs voies génétiques sous-jacentes ([Hoekstra, 2006]; [Manceau et al., 2010]; [Kronforst et al., 2012]) ce qui a été observé dans une variété de cas ([Hoekstra, 2006]; [Manceau et al., 2010]). En outre, j'ai aussi trouvé un manque de déterminisme allant de l'architecture génétique au phénotype dans le fait que des patrons tigrés comme ceux affichées par *H. hecale* et *H. ismenius* soient contrôlées par la même boîte à outils de la couleur utilisée par des espèces du genre *Heliconius* aux couleurs fortement distincts ([Kronforst et al., 2006a]; [Baxter et al., 2008b]; [Papa et al., 2008]). Cette même boîte à outils est sous-jacente à des phénotypes mimétiques même entre espèces éloignées telles que *H. melpomene* et *H. erato*. Ainsi, la flexibilité dans le résultat phénotypique des mêmes loci est ici marquée par des fortes contraintes sélectives pour la ressemblance mimétique entre les espèces appartenant à un même cercle mimétique. Cette polyvalence et évolutivité de l'architecture génétique cause à la fois la diversité phénotypique et la convergence mimétique observées dans le genre *Heliconius*.

La convergence mimétique ne peut pas être expliquée par le hasard. En revanche, des conditions environnementales similaires (i.e. des cercles mimétiques) imposent de fortes pressions sélectives, de façon similaire à ce qui a lieu dans d'autres cas de convergence phénotypique ([Losos, 1992]; [Colosimo et al., 2005]; [Rogers and Bernatchez, 2007]; [Soria-Carrasco et al., 2014]). J'ai constaté de la sélection agissant sur la coloration alaire à une étape précoce de la divergence, en utilisant la théorie des clines couplée à la génétique des populations, bien que en étudiant le profil de divergence au long du génome. De la sélection disruptive déterminant une

adaptation locale à des environnements mimétiques distincts sur chaque côté de la zone d'hybridation entre *H. hecale melicerta* et *H. hecale zuleika* est reflétée dans la distribution clinale des allèles au loci de la couleur. Ceci contraste avec l'absence de structure génétique trouvée ailleurs sur le génome. Ces pressions sélectives sont de force intermédiaire, comme le montre la largeur intermédiaire de la zone d'hybridation en relation à d'autres zones hybrides interraciales dans le genre ([Mallet, 1986] ; [Blum, 2002] ; [Arias et al., 2012]). Cependant, cette sélection de force modérée est encore capable de garder l'individualité phénotypique, c'est-à-dire des patrons de coloration différenciés, malgré un niveau considérable de flux génique entre les deux races. La complexité des patrons tigrés et les environnements mimétiques riches en espèces et fortement variables pourraient déterminer une sélection disruptive de force intermédiaire sur des papillons hybrides sur chaque côté de cette zone hybride. J'ai aussi trouvé que la force de la sélection sur les différents éléments du patron diffère, ce qui suggère que les éléments de coloration de l'aile ne sont pas égaux aux yeux des prédateurs, et peut-être pas non plus aux yeux des partenaires.

Récemment, des études moléculaires à l'échelle du génome ont aidé à localiser des régions génomiques sous sélection divergente et potentiellement impliquées dans la spéciation (e.g. [Jones et al., 2012] ; [Lawniczak et al., 2010] ; [Nadeau et al., 2012] ; [Ellegren et al., 2012] ; [Kronforst et al., 2013]). De telles régions, qui ont été appelées des îlots de divergence, peuvent être attribués à l'action directe de la sélection divergente, mais peuvent aussi résulter des particularités de l'architecture du génome, comme des taux locaux de recombinaison ou de la diversité nucléotidique locale ([Carneiro et al., 2010] ; [Ellegren et al., 2012] ; [Cruickshank and Hahn, 2014]). Pour comprendre vraiment les raisons à la base des paysages hétérogènes, nous avons besoin d'une exploration statistique des données associée à une meilleure connaissance des traits écologiques différenciés et de leur base génétique. La combinaison des scans génomiques avec des approches basées sur des gènes candidats peut permettre de distinguer entre des régions impliquées dans la diversification et des faux positifs (e.g. ([Michel et al., 2010])). Ici, j'ai adopté cette synergie d'approches et j'ai trouvé que quelques îlots de divergence sont associés avec des loci de la couleur à des étapes précoces de différenciation. Cela montre que les régions connues comme étant sous-jacentes à des différences écologiques sous forte sélection disruptive (tels que les patrons de couleur) apparaissent comme des valeurs aberrantes de différenciation génomique, et pourraient donc résister au flux génique. Pourtant, j'ai aussi trouvé des pics de différenciation autour des régions de la couleur de l'aile entre des lignées co-mimétiques bien divergées, bien que l'on ne s'attendrait pas vraiment à trouver de divergence génétique au sein des allèles causant des phénotypes similaires. Je fais l'hypothèse que ces signaux moléculaires, plutôt que d'être une réponse à la sélection divergente, sont le résultat d'une sélection purifiante sur des allèles impliqués dans la ressemblance mimétique, ce qui est reflété dans le peu de diversité nucléotidique autour des loci de la couleur de l'aile. J'ai observé un paysage de différenciation fortement hétérogène dans la comparaison interraciale, avec de multiples pics. Ceci contraste avec le peu de pics de divergence associés aux loci de la couleur qui ont été retrouvés dans des comparaisons interraciales des espèces explorées jusqu'à maintenant. Il a été suggéré que l'augmentation du nombre d'îlots de divergence serait due à la progression de la spéciation vers des étapes avancées ([Kronforst et al., 2013]), ce qui contraste avec le modèle classique de spéciation avec flux génique, selon lequel on s'attend à voir grandir les îlots plutôt que de les voir augmenter en nombre au long de ce continuum ([Feder et al., 2012] ; voir Introduction). Au contraire, les multiples pics que j'ai observés pourraient résulter de la diversité nucléotidique globalement très basse chez *H. hecale*. Donc, interpréter ces multiples pics comme une évidence que ces populations se trouvent à une étape de divergence plus avancée que d'autres paires étudiées jusqu'ici n'est pas évident. Basée sur les ambiguïtés décrites ici, je suggère qu'une évaluation plus profonde de ces données génomiques nous manque pour pouvoir affirmer des conclusions

sur la manière dont la différenciation génomique a évolué au cours des deux étapes de continuum de spéciation étudiées ici. Bien que des études génomiques de différenciation puissent être très utiles pour détecter quelques-unes des régions génomiques contrôlant des caractères écologiquement importants pendant la spéciation avec flux génique, le profil génomique de différenciation peut être prêt à confusion. C'est pourquoi il faut faire attention au décalage entre les valeurs aberrantes de différenciation et les gènes de la spéciation.

### A.5.2 Puissant, mais pas tout-puissant

J'ai exploré le processus de spéciation avec flux génique entre deux paires de taxons au long du continuum de spéciation : entre des races parapatriques de *H. hecale*, et entre les deux espèces proches sympatriques *H. hecale* et *H. ismenius*. En faisant cela, j'ai intégré deux vues distinctes sur le processus de spéciation nommées spyglass et magnifying glass (voir Introduction ou [Via, 2009]). Ceci a contribué à rechercher les mécanismes de l'évolution de l'isolement reproducteur au long du continuum de spéciation dans le genre *Heliconius*.

Il a été proposé que la spéciation écologique est favorisée par la sélection divergente sur des traits écologiquement importants, en particulier si ces traits sont aussi impliqués dans l'accouplement sélectif ([Bush, 1975]; [Felsenstein, 1981]; [Diehl and Bush, 1989]; [Kirkpatrick and Ravigné, 2002]; [Malausa et al., 2005]). Chez les papillons du genre *Heliconius* comme dans d'autres systèmes modèles, des traits dits « magiques » sont capables de fortifier ou de maintenir à la fois un isolement extrinsèque postzygotique et prézygotique. Plusieurs cas présentant une situation similaire implique des lignées divergentes dans des stades précoces du processus de spéciation (e.g.; [Feder and Filchak, 1999]). Pourtant, on ne sait pas de manière sûre si des traits adaptatifs importants sont capables seuls d'achever la spéciation ou au contraire, si la spéciation ne pourrait jamais aboutir si des traits additionnels ne divergent pas dans le processus. Cette question ouverte s'applique aux multiples races colorées des papillons *Heliconius*.

On peut distinguer les races parapatriques de *Heliconius hecale* seulement au niveau de la couleur alaire. Elles ne semblent pas avoir évolué d'autres adaptations écologiques spécifiques. En particulier, aucune différence au niveau de la préférence du macrohabitat ou du microhabitat semble exister entre ces races. Les deux races ont été récoltées en général dans des habitats similaires, dans des espaces ouverts d'herbes et de maquis, près des bordures de forêts, ou dans des clairières (observation personnelle). De plus, les larves des deux races se nourrissent de la même plante hôte *Passiflora vitifolia*. En effet, on n'a pas observé que des facteurs écologiques distincts (par exemple la plante hôte de la larve ou la spécialisation de l'habitat) aient été impliqués dans la formation des races géographiques dans le genre ([Mallet and Gilbert, 1995]; [Jiggins et al., 1996]).

J'ai observé que de la préférence de partenaire sélective basé sur la coloration de l'aile, bien que partielle et asymétrique, existe parmi les races de *H. hecale* étudiées ici. Cela confirme que le choix du partenaire basé sur la coloration et opéré par le mâle est répandu dans le genre et qu'il évolue tôt dans le continuum de spéciation. Au contraire, à des étapes plus avancées du processus de spéciation, la coloration de l'aile ne joue pas de rôle prépondérant pour garder les frontières de l'espèce. En effet, les espèces *H. hecale* et *H. ismenius* sont mimétiques et elles gardent encore leur intégrité génomique (voir plus loin). Ici, j'ai mis en évidence que l'évolution de signaux chimiques divergents s'est produite entre ces « bonnes » espèces sympatriques qui partagent des patrons mimétiques. Ces signaux pourraient contribuer à la conservation des frontières de l'espèce entre des lignées dont on se serait attendu qu'ils s'hybrident. Je suggère que cette reconnaissance de l'espèce basée sur des phéromones est opérée par les femelles, contrairement

au choix du partenaire basé sur la coloration de l'aile, qui est opéré principalement par les mâles. D'autres facteurs tels que la spécificité de la plante hôte et la préférence du microhabitat semblent aussi avoir divergé entre ces espèces, bien que des données écologiques rigoureuses manquent à ce stade.

Mes résultats ne contrastent pas avec les constatations faites dans d'autres sous-clades dans le genre. Il a été suggéré que des changements dans la coloration ont dû être importants pour la spéciation ([Mallet et al., 1998a] ; [Mallet, 2010] ; [Jiggins et al., 2001] ; [Chamberlain et al., 2009]). Par exemple, les espèces sœurs *H. melpomene* et *H. cydno* ont été suggérées d'avoir spéciées en association avec un changement mimétique. Pourtant, la spécificité de la plante hôte a aussi évolué, tout comme la préférence du microhabitat et l'infertilité hybride partielle en raison des incompatibilités génomiques ([Jiggins et al., 2001] ; [Naisbit et al., 2002]). Cela suggère que la divergence des patrons de coloration pourrait ne pas être une force suffisante pour conduire à l'achèvement de la spéciation dans le genre.

Mes résultats renforcent l'opinion que la divergence de la coloration alaire est centrale au début du processus de spéciation, mais n'est pas suffisante pour achever la spéciation. La spéciation dans le clade est probablement, au contraire, un processus multidimensionnel, avec divers traits opérant comme des barrières reproductives d'isolement accumulées durant le processus. En accord avec cette idée, il y a 10 fois plus de races géographiques à des couleurs diverses que d'espèces dans le genre, ce qui suggère que la divergence dans la coloration de l'aile seule ne conduit pas toujours à la spéciation. De plus, quelques paires de races au sein d'une espèce ont des généalogies plus profondes que des paires d'espèces dans le genre ([Kozak et al., 2015]), ce qui indique que ces races peuvent restées dans des stades de divergence précoces pendant une longue période évolutive sans spéciation. De la même façon, la zone hybride entre les races de *H. hecale* étudiées ici pourrait potentiellement être conservée indéfiniment dans un équilibre stable entre la sélection divergente agissant sur la coloration de l'aile et un haut mélange génétique, si les conditions environnementales restent stables.

Dans l'ensemble, j'ai mis mes résultats dans un cadre comparatif et j'ai trouvé que des mécanismes similaires opèrent sur la diversification de tout le clade. En accord avec des résultats précédents, j'ai observé que des incompatibilités génomiques intrinsèques évoluent dans des stades avancés du continuum. Notamment, aucune incompatibilité génomique intrinsèque n'existe entre les races de *H. hecale*. Encore plus surprenant, des hybrides entre les espèces sympatriques *H. hecale* et *H. ismenius* sont viables. Contrastant avec ces cas de spéciation avec flux génique, l'accumulation de barrières en allopatrie semble suivre une modalité très différente. Par exemple, j'ai trouvé que la règle d'Haldane s'applique à des hybrides entre *H. hecale melicerta* et *H. hecale clearei*, qui sont des races allopatriques d'une même espèce (résultats non montrés). A savoir, des femelles hybrides sont moins viables que des mâles et totalement infertiles, alors que les mâles sont complètement fertiles. Au contraire, les hybrides entre les « bonnes » espèces en sympatrie étudiées ici ne montrent pas d'inviableté. Ceci indique que, comme on s'attendrait, de fortes barrières contre le flux génique évoluent facilement au niveau intraspécifique quand des conditions d'isolement géographique existent.

### A.5.3 Une hybridation généralisée n'est pas la règle dans le genre

Le flux génique à de multiples niveaux de divergence dans le genre *Heliconius* a été trouvé ([Bull et al., 2006] ; [Kronforst et al., 2006b] ; [Martin et al., 2013]). Ici, je n'ai pas observé de signaux moléculaires d'introgession dans la paire d'espèces proches *H. hecale* et *H. ismenius*, malgré le fait qu'elles aient divergé depuis peu de temps. Leur temps de divergence peut avoir été assez long pour produire de très fortes barrières reproductives contre le flux génique, comme

par exemple les signaux chimiques mentionnés ci-dessus, qui semblent agir comme des barrières d'isolement pré-zygotique. C'est pourquoi je suggère que l'hybridation n'est pas généralisée dans le genre *Heliconius* (en accord avec les résultats de Krzysztof Kozak, communication personnelle), bien qu'un nombre important d'espèces dans le genre s'hybrident, principalement des espèces des sous-clades *melpomene*, *cydno* et des sylvaniformes ([Mallet et al., 2007]), la dernière desquels comprend les espèces étudiées ici.

# List of Tables

1.1	Summary table of major and minor effect wing colour pattern loci mapped in the genus <i>Heliconius</i> . . . . .	73
1.2	Sampling localities for the specimens of <i>Heliconius hecale</i> and <i>H. ismenius</i> used in the crosses . . . . .	78
1.3	List of markers used to fine-map colour loci in <i>H. hecale</i> and <i>H. ismenius</i> . . . . .	79
1.4	Allelic segregation of colour pattern loci in the mapping families of <i>H. hecale</i> and <i>H. ismenius</i> . . . . .	82
1.6	Summary of RAD library sequencing and mapping statistics. . . . .	84
1.7	Number of per-individual 100-bp Illumina reads in Library 1 (Brood 122: <i>H. hecale melicerta</i> × <i>H. h. zuleika</i> ) . . . . .	85
1.8	Number of per-individual 100-bp Illumina reads in Library 2 (Brood 112: <i>H. hecale melicerta</i> × <i>H. h. clearei</i> ) . . . . .	87
1.9	Number of per-individual 100-bp Illumina reads in Library 3 (Broods 96 and 101 mainly: <i>H. ismenius bouletti</i> × <i>H. h. telchinia</i> ) . . . . .	89
1.10	Summary of RAD library marker types and segregation patterns . . . . .	91
1.11	Summary of major and minor (suggestive) QTLs mapped in <i>H. hecale</i> and <i>H. ismenius</i> . . . . .	92
2.1	Maximum likelihood parameter estimates for the phenotypic clines of the three colour loci determining most of wing pattern differences between <i>H. hecale zuleika</i> and <i>H. h. melicerta</i> . . . . .	114
2.2	Characterisation of microsatellite markers used to explore the genetic structure between <i>H. hecale zuleika</i> and <i>H. hecale melicerta</i> . . . . .	124
2.3	Number of courtship events directed toward <i>H. hecale melicerta</i> ( <i>melicerta</i> ) and <i>H. hecale zuleika</i> ( <i>zuleika</i> ) female wing models by males of both species ( $N = 31$ <i>melicerta</i> and $N = 32$ <i>zuleika</i> ) . . . . .	125
2.4	Summary of the best documented interracial and interspecific hybrid zones in <i>Heliconius</i> butterflies . . . . .	126
3.1	Genome-wide statistics derived from ABBA-BABA tests . . . . .	142
3.2	Sampling localities of <i>Heliconius hecale</i> and <i>H. ismenius specimens</i> and of some specimens from the <i>H. melpomene</i> clade used as outgroup . . . . .	159
3.3	Ethogram of the male-female interactions during courtship and mating in <i>Heliconius hecale</i> . . . . .	160
3.4	Compositional similarities and differences between the chemical blends extracted from abdominal glands of males and females of <i>H. hecale melicerta</i> and <i>H. ismenius bouletti</i> . . . . .	163
3.5	Compositional similarities and differences between the chemical blends extracted from wings of males and females of <i>H. hecale melicerta</i> and <i>H. ismenius bouletti</i> . . . . .	166



3.6	Compositional similarities and differences between the chemical blends extracted from the cuticula near the insertion of the wings of males and females of <i>H. hecale melicerta</i> and <i>H. ismenius bouletti</i> . . . . .	168
-----	---	-----

## List of Figures

1	Schematic representation of Müllerian mimicry . . . . .	14
2	Examples of mimicry rings formed by several toxic neo-tropical butterfly species .	15
3	The continuous nature of divergence during speciation . . . . .	16
4	Phylogeny of the tribe Heliconiini . . . . .	24
5	Speciation continuum in <i>Heliconius</i> butterflies . . . . .	27
6	The allopatry/sympatry spectrum, using spatial definitions . . . . .	29
7	Distinctive plumage ornamentation among duck species despite high levels of interspecific hybridisation . . . . .	35
8	Molecular patterns at multiple levels underlying phenotypic convergence . . . . .	43
9	Phylogenetic relationships and candidate genes for pelvic reduction in threespine and ninespine sticklebacks . . . . .	46
10	The parallel radiation of mimetic colour-pattern complexes in <i>Heliconius erato</i> and <i>H. melpomene</i> . . . . .	48
11	Distribution of <i>Heliconius ismenius</i> geographic races . . . . .	49
12	Distribution of <i>Heliconius hecale</i> geographic races . . . . .	50
13	Stages of life cycle of <i>Heliconius ismenius</i> and <i>H. hecale</i> . . . . .	51
14	Local polymorphism in <i>H. numata</i> is controlled by a supergene . . . . .	52
15	Schema summarising this study . . . . .	53
1.1	Summary of crosses performed in <i>Heliconius hecale</i> and <i>H. ismenius</i> . . . . .	58
1.2	Fine mapping of wing patterning loci in <i>H. hecale</i> and <i>H. ismenius</i> . . . . .	61
1.3	Phenotypic effect of Mendelian wing-patterning loci and major identified QTLs .	64
1.4	Conservatism and novelty in the genetic architecture underlying the diversity of <i>Heliconius</i> wing patterns . . . . .	68
1.5	Number of retained positions after filtering . . . . .	95
1.6	Distribution of sequencing depth . . . . .	96
1.7	Linkage map of <i>Heliconius hecale</i> derived from the <i>H. hecale melicerta</i> × <i>H. hecale zuleika</i> cross . . . . .	97
1.8	Linkage map of <i>Heliconius hecale</i> derived from the <i>H. hecale melicerta</i> × <i>H. hecale clearei</i> cross . . . . .	98
1.9	Linkage map of <i>Heliconius ismenius</i> derived from the <i>H. ismenius bouletti</i> × <i>H. ismenius telchinia</i> cross . . . . .	99
1.10	Genomic position for each colour locus on its linkage group . . . . .	100
1.11	Detail of QTL mapping around the highest peaks of LOD score in each cross . . .	101
1.12	Genomic position and phenotypic effect of suggestive QTLs . . . . .	103
2.1	Sampling sites and studied populations on the analysed transect . . . . .	108

2.2	Sample of pure and hybrid phenotypes at the color genes accross Panama . . . .	109
2.3	Factorial Correspondence Analysis (FCA) of the microsatellite data . . . . .	112
2.4	Genome-wide genetic differentiation ( <i>Fst</i> ) between <i>H. hecale zuleika</i> and <i>H. h. melicerta</i> . . . . .	113
2.5	Phenotypic clines for the three wing colour loci and distribution of genotypic classes along the studied transect . . . . .	115
2.6	Male choice of <i>H. hecale melicerta</i> and <i>H. hecale zuleika</i> between female wing models of both races . . . . .	117
2.7	Partial representation of tiger-patterned mimicry rings in Panama . . . . .	120
2.8	Genome-wide differentiation ( <i>Fst</i> ) between <i>H. hecale zuleika</i> and <i>H. h. melicerta</i> and nucleotide diversity of both races . . . . .	128
3.1	Distribution and sampling localities of the subspecies of <i>Heliconius hecale</i> and <i>H. ismenius</i> involved in this study . . . . .	133
3.2	Results of the genomic analyses used to test for gene flow between species <i>H. hecale</i> and <i>H. ismenius</i> . . . . .	140
3.3	Genome-wide genetic differentiation ( <i>Fst</i> ) between <i>H. hecale</i> and <i>H. ismenius</i> and nucleotide diversity of the mimetic sympatric races . . . . .	141
3.4	Four clade ABBA-BABA test of introgression . . . . .	143
3.5	Synthetic diagram of the <i>main</i> steps of courting and mating behaviour of <i>Heliconius hecale</i> . . . . .	143
3.6	Comparison of interspecific and intraspecific male-female encounters . . . . .	144
3.7	Male choice of <i>H. hecale melicerta</i> and <i>H. ismenius bouletti</i> between female wings models of both species . . . . .	145
3.8	Summary of the behavioural experiments to test for female choice in <i>H. hecale</i> based on visual cues. . . . .	146
3.9	Analysis of the chemical composition of the abdominal glands of <i>H. hecale melicerta</i> and <i>H. ismenius bouletti</i> . . . . .	147
3.10	Analysis of the chemical composition of the wings of <i>H. hecale melicerta</i> and <i>H. ismenius bouletti</i> . . . . .	149
3.11	Analysis of the chemical composition of the cuticule in the basis of the wings of <i>H. hecale melicerta</i> and <i>H. ismenius bouletti</i> . . . . .	150
4.1	Hypothetical representation of the evolution of supergene P from an ancestral multilocus architecture . . . . .	171

# Bibliography

- [Abbott et al., 2013] Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., Jones, J., Keller, B., Marczewski, T., Mallet, J., Martinez-Rodriguez, P., Möst, M., Mullen, S., Nichols, R., Nolte, A. W., Parisod, C., Pfennig, K., Rice, A. M., Ritchie, M. G., Seifert, B., Smadja, C. M., Stelkens, R., Szymura, J. M., Väinölä, R., Wolf, J. B. W., and Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26(2):229–246.
- [Alexandrou et al., 2011] Alexandrou, M. A., Oliveira, C., Maillard, M., McGill, R. A. R., Newton, J., Creer, S., and Taylor, M. I. (2011). Competition and phylogeny determine community structure in Mullerian co-mimics. *Nature*, 469(7328):84–88.
- [Anderson, 1949] Anderson, E. (1949). Introgressive hybridization. *Introgressive hybridization*.
- [Anderson, 2001] Anderson, M. J. (2001). Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(3):626–639.
- [Anderson et al., 2009] Anderson, T. M., Candille, S. I., Musiani, M., Greco, C., Stahler, D. R., Smith, D. W., Padhukasahasram, B., Randi, E., Leonard, J. A., and Bustamante, C. D. (2009). Molecular and evolutionary history of melanism in North American gray wolves. *Science*, 323(5919):1339–1343.
- [Andersson et al., 2007] Andersson, J., Borg-Karlson, A.-K., Vongvanich, N., and Wiklund, C. (2007). Male sex pheromone release and female mate choice in a butterfly. *Journal of Experimental Biology*, 210(6):964–970.
- [Andersson, 1994] Andersson, M. B. (1994). *Sexual selection*. Princeton University Press.
- [Arendt and Reznick, 2008] Arendt, J. and Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology & Evolution*, 23(1):26–32.
- [Arias et al., 2008] Arias, C. F., Muñoz, A. G., Jiggins, C. D., Mavárez, J., Bermingham, E., and Linares, M. (2008). A hybrid zone provides evidence for incipient ecological speciation in *Heliconius* butterflies. *Molecular Ecology*, 17(21):4699–4712.
- [Arias et al., 2012] Arias, C. F., Rosales, C., Salazar, C., Castaño, J., Bermingham, E., Linares, M., and McMillan, W. O. (2012). Sharp genetic discontinuity across a unimodal *Heliconius* hybrid zone. *Molecular Ecology*, 21(23):5778–5794.
- [Arn et al., 1992] Arn, H., Tóth, M., Priesner, E., and WPRS, I. (1992). *List of sex pheromones of Lepidoptera and related attractants*. Organisation internationale de lutte biologique Section régionale ouest paléarctique (OILB-SROP).
- [Arnold, 1992] Arnold, M. L. (1992). Natural Hybridization as an Evolutionary Process. *Annual Review of Ecology and Systematics*, 23(1):237–261.
- [Azevedo et al., 2006] Azevedo, R. B., Lohaus, R., Srinivasan, S., Dang, K. K., and Burch, C. L. (2006). Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature*, 440(7080):87–90.
- [Baer, 2003] Baer, B. (2003). Bumblebees as model organisms to study male sexual selection in social insects. *Behavioral Ecology and Sociobiology*, 54(6):521–533.

- [Barluenga et al., 2006] Barluenga, M., Stölting, K. N., Salzburger, W., Muschick, M., and Meyer, A. (2006). Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, 439(7077):719–723.
- [Barton, 2013] Barton, N. (2013). Does hybridization influence speciation? *Journal of evolutionary biology*, 26(2):267–269.
- [Barton, 1995] Barton, N. H. (1995). A general model for the evolution of recombination. *Genetical research*, 65(2):123–145.
- [Barton and Gale, 1993] Barton, N. H. and Gale, K. S. (1993). *Genetic analysis of hybrid zones*. New York, harrison, r. g. (ed.). hybrid zones and the evolutionary process edition.
- [Barton and Hewitt, 1981] Barton, N. H. and Hewitt, G. M. (1981). Hybrid Zones and Speciation. Evolution and Speciation and Says in Honour of MJD White.
- [Barton and Hewitt, 1985] Barton, N. H. and Hewitt, G. M. (1985). Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics*, 16(1):113–148.
- [Barton and Hewitt, 1989] Barton, N. H. and Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, 341(6242):497–503.
- [Bates, 1862] Bates, H. W. (1862). Contributions to an Insect Fauna of the Amazon Valley. Lepidoptera: Heliconidæ. *Transactions of the Linnean Society of London*, 23(3):495–566.
- [Baxter et al., 2008a] Baxter, S. W., Johnston, S. E., and Jiggins, C. D. (2008a). Butterfly speciation and the distribution of gene effect sizes fixed during adaptation. *Heredity*, 102(1):57–65.
- [Baxter et al., 2010] Baxter, S. W., Nadeau, N. J., Maroja, L. S., Wilkinson, P., Counterman, B. A., Dawson, A., Beltran, M., Perez-Espona, S., Chamberlain, N., Ferguson, L., Clark, R., Davidson, C., Glithero, R., Mallet, J., McMillan, W. O., Kronforst, M., Joron, M., French Constant, R. H., and Jiggins, C. D. (2010). Genomic Hotspots for Adaptation: The Population Genetics of Müllerian Mimicry in the *Heliconius melpomene* Clade. *PLoS Genet*, 6(2):e1000794.
- [Baxter et al., 2008b] Baxter, S. W., Papa, R., Chamberlain, N., Humphray, S. J., Joron, M., Morrison, C., French Constant, R. H., McMillan, W. O., and Jiggins, C. D. (2008b). Convergent Evolution in the Genetic Basis of Müllerian Mimicry in *Heliconius* Butterflies. *Genetics*, 180(3):1567–1577.
- [Beaumont, 2005] Beaumont, M. A. (2005). Adaptation and speciation: what can F<sub>ST</sub> tell us? *Trends in Ecology & Evolution*, 20(8):435–440.
- [Belkhir et al., 1996] Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., and Bonhomme, F. (1996). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. *Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France)*.
- [Beltran et al., 2007] Beltran, M., Jiggins, C. D., Brower, A. V., Bermingham, E., and Mallet, J. (2007). Do pollen feeding, pupal-mating and larval gregariousness have a single origin in *Heliconius* butterflies? Inferences from multilocus DNA sequence data. *Biological Journal of the Linnean Society*, 92(2):221–239.
- [Benson, 1972] Benson, W. W. (1972). Natural Selection for Müllerian Mimicry in *Heliconius erato* in Costa Rica. *Science*, 176(4037):936–939.
- [Benson, 1982] Benson, W. W. (1982). Alternative models for infrageneric diversification in the humid tropics: tests with passion vine butterflies. In Prance, G. T., editor, *Biological Diversification in the Tropics*, pages 608–640. Columbia Univ. Press, New York, NY.
- [Benson et al., 1975] Benson, W. W., Brown Jr, K. S., and Gilbert, L. E. (1975). Coevolution of plants and herbivores: passion flower butterflies. *Evolution*, pages 659–680.
- [Bergström and Lundgren, 1973] Bergström, G. and Lundgren, L. (1973). Androconial secretion of three species of butterflies of the genus *Pieris* (Lep., Pieridae). *Zoon Suppl*, 1:67–75.
- [Bernatchez and Dodson, 1990] Bernatchez, L. and Dodson, J. J. (1990). Allopatric origin of sympatric populations of lake whitefish (*Coregonus clupeaformis*) as revealed by mitochondrial-DNA restriction analysis. *Evolution*, pages 1263–1271.
- [Bierne et al., 2003] Bierne, N., Daguin, C., Bonhomme, F., David, P., and Borsa, P. (2003). Direct selection on allozymes is not required to explain heterogeneity among marker loci across a *Mytilus* hybrid zone. *Molecular Ecology*, 12(9):2505–2510.

- [Bierne et al., 2011] Bierne, N., Welch, J., Loire, E., Bonhomme, F., and David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, 20(10):2044–2072.
- [Blum, 2002] Blum, M. J. (2002). Rapid movement of a *Heliconius* hybrid zone: evidence for phase III of Wright's shifting balance theory? *Evolution*, 56(10):1992–1998.
- [Blum, 2008] Blum, M. J. (2008). Ecological and genetic associations across a *Heliconius* hybrid zone. *Journal of Evolutionary Biology*, 21(1):330–341.
- [Boppré, 1984] Boppré, M. (1984). Chemically mediated interactions between butterflies. In *Symposia of the Royal Entomological Society of London*.
- [Boughman et al., 2005] Boughman, J. W., Rundle, H. D., and Schluter, D. (2005). Parallel evolution of sexual isolation in sticklebacks. *Evolution*, 59(2):361–373.
- [Boyko, 2005] Boyko, A. R. (2005). *The evolution of warning coloration in Heliconius butterflies*. PhD thesis, Purdue University, West Lafayette, IN.
- [Bradshaw et al., 1995] Bradshaw, H., Wilbert, M., and Otto, K. (1995). Genetic mapping of floral traits associated with reproductive isolation in. *Nature*, 376:31.
- [Brideau et al., 2006] Brideau, N. J., Flores, H. A., Wang, J., Maheshwari, S., Wang, X., and Barbash, D. A. (2006). Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *science*, 314(5803):1292–1295.
- [Broman and Sen, 2009] Broman, K. and Sen, S. (2009). *A Guide to QTL Mapping with R/qlt*. Springer.
- [Broman et al., 2002] Broman, K. W., Rowe, L. B., Churchill, G. A., and Paigen, K. (2002). Crossover interference in the mouse. *Genetics*, 160(3):1123–1131.
- [Broman and Speed, 2002] Broman, K. W. and Speed, T. P. (2002). A model selection approach for the identification of quantitative trait loci in experimental crosses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):641–656.
- [Broman et al., 2003] Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qlt: QTL mapping in experimental crosses. *Bioinformatics (Oxford, England)*, 19(7):889–890.
- [Bronson et al., 2003] Bronson, C. L., Grubb Jr, T. C., Sattler, G. D., and Braun, M. J. (2003). Mate preference: a possible causal mechanism for a moving hybrid zone. *Animal Behaviour*, 65(3):489–500.
- [Brower, 1996] Brower, A. V. Z. (1996). A new mimetic species of *Heliconius* (Lepidoptera: Nymphalidae), from southeastern Colombia, revealed by cladistic analysis of mitochondrial DNA sequences. *Zoological Journal of the Linnean Society*, 116(3):317–332.
- [Brown, 1976] Brown, K. S. (1976). An illustrated key to the silvaniform *Heliconius* (Lepidoptera: Nymphalidae) with descriptions of new subspecies. *Trans. Am. Entomol. Soc.*, 102:373–484.
- [Brown, 1981] Brown, K. S. (1981). The Biology of *Heliconius* and Related Genera. *Annual Review of Entomology*, 26(1):427–457.
- [Brown and Mielke, 1972] Brown, K. S. and Mielke, O. H. H. (1972). The heliconians of Brazil (Lepidoptera: Nymphalidae). Part II. Introduction and general comments, with a supplementary revision of the tribe. *Zoologica, New York*, 57:1–40.
- [Brown and Benson, 1974] Brown, Jr., K. S. and Benson, W. W. (1974). Adaptive Polymorphism Associated with Multiple Müllerian Mimicry in *Heliconius numata* (Lepid. Nymph.). *Biotropica*, 6(4):205–228.
- [Browning and Browning, 2009] Browning, B. L. and Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American journal of human genetics*, 84(2):210–223.
- [Brunton and Majerus, 1995] Brunton, C. F. A. and Majerus, M. E. N. (1995). Ultraviolet Colours in Butterflies: Intra- or Inter-Specific Communication? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 260(1358):199–204.
- [Bull et al., 2006] Bull, V., Beltrán, M., Jiggins, C. D., McMillan, W. O., Bermingham, E., and Mallet, J. (2006). Polyphyly and gene flow between non-sibling *Heliconius* species. *BMC Biology*, 4(1):11.

- [Bush, 1975] Bush, G. L. (1975). Modes of animal speciation. *Annual Review of Ecology and Systematics*, pages 339–364.
- [Butlin et al., 2012] Butlin, R., Debelle, A., Kerth, C., Snook, R. R., Beukeboom, L. W., Castillo, C. R., Diao, W., Maan, M. E., Paolucci, S., and Weissing, F. J. (2012). What do we need to know about speciation? *Trends in Ecology & Evolution*, 27(1):27–39.
- [Butlin and Ritchie, 1991] Butlin, R. and Ritchie, M. (1991). Variation in female mate preference across a grasshopper hybrid zone. *Journal of Evolutionary Biology*, 4(2):227–240.
- [Bímová et al., 2011] Bímová, B. V., Macholán, M., Baird, S. J., Munclinger, P., Dufková, P., Laukaitis, C. M., Karn, R. C., Luzynski, K., Tucker, P. K., and Piálek, J. (2011). Reinforcement selection acting on the European house mouse hybrid zone. *Molecular Ecology*, 20(11):2403–2424.
- [Cardé et al., 1978] Cardé, R., Roelofs, W., Harrison, R., Vawter, A., Brussard, P., Mutuura, A., and Munroe, E. (1978). European corn borer: pheromone polymorphism or sibling species? *Science*, 199(4328):555–556.
- [Carlborg and Haley, 2004] Carlborg, O. and Haley, C. S. (2004). Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics*, 5(8):618–625.
- [Carneiro et al., 2010] Carneiro, M., Blanco-Aguilar, J. A., Villafuerte, R., Ferrand, N., and Nachman, M. W. (2010). Speciation in the European rabbit (*Oryctolagus cuniculus*): islands of differentiation on the X chromosome and autosomes. *Evolution; International Journal of Organic Evolution*, 64(12):3443–3460.
- [Carneiro et al., 2009] Carneiro, M., Ferrand, N., and Nachman, M. W. (2009). Recombination and Speciation: Loci Near Centromeres Are More Differentiated Than Loci Near Telomeres Between Subspecies of the European Rabbit (*Oryctolagus cuniculus*). *Genetics*, 181(2):593–606.
- [Carter et al., 2005] Carter, A. J. R., Hermisson, J., and Hansen, T. F. (2005). The role of epistatic gene interactions in the response to selection and the evolution of evolvability. *Theoretical population biology*, 68(3):179–196.
- [Catchen et al., 2013] Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology*, 22(11):3124–3140.
- [Chamberlain et al., 2011] Chamberlain, N. L., Hill, R. I., Baxter, S. W., Jiggins, C. D., and Kronforst, M. R. (2011). Comparative population genetics of a mimicry locus among hybridizing *Heliconius* butterfly species. *Heredity*, 107(3):200–204.
- [Chamberlain et al., 2009] Chamberlain, N. L., Hill, R. I., Kapan, D. D., Gilbert, L. E., and Kronforst, M. R. (2009). Polymorphic Butterfly Reveals the Missing Link in Ecological Speciation. *Science*, 326(5954):847–850.
- [Chan et al., 2010] Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Shapiro, M. D., Brady, S. D., Southwick, A. M., Absher, D. M., Grimwood, J., and Schmutz, J. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *science*, 327(5963):302–305.
- [Charlesworth, 1998] Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, 15(5):538–543.
- [Churchill and Doerge, 1994] Churchill, G. A. and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, 138(3):963–971.
- [Clarkson et al., 2014] Clarkson, C. S., Weetman, D., Essandoh, J., Yawson, A. E., Maslen, G., Manske, M., Field, S. G., Webster, M., Antão, T., MacInnis, B., Kwiatkowski, D., and Donnelly, M. J. (2014). Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. *Nature Communications*, 5.
- [Colosimo et al., 2005] Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D., and Kingsley, D. M. (2005). Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. *Science*, 307(5717):1928–1933.
- [Colosimo et al., 2004] Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., and Kingsley, D. M. (2004). The Genetic Architecture of Parallel Armor Plate Reduction in Threespine Sticklebacks. *PLoS Biol*, 2(5):e109.

- [Costanzo and Monteiro, 2007] Costanzo, K. and Monteiro, A. (2007). The use of chemical and visual cues in female choice in the butterfly *Bicyclus anynana*. *Proceedings of the Royal Society B: Biological Sciences*, 274(1611):845–851.
- [Counterman et al., 2010] Counterman, B. A., Araujo-Perez, F., Hines, H. M., Baxter, S. W., Morrison, C. M., Lindstrom, D. P., Papa, R., Ferguson, L., Joron, M., French Constant, R. H., Smith, C. P., Nielsen, D. M., Chen, R., Jiggins, C. D., Reed, R. D., Halder, G., Mallet, J., and McMillan, W. O. (2010). Genomic Hotspots for Adaptation: The Population Genetics of Müllerian Mimicry in *Heliconius erato*. *PLoS Genet*, 6(2):e1000796.
- [Coyle et al., 2007] Coyle, S. M., Huntingford, F. A., and Peichel, C. L. (2007). Parallel evolution of *Pitx1* underlies pelvic reduction in Scottish threespine stickleback (*Gasterosteus aculeatus*). *Journal of Heredity*, 98(6):581–586.
- [Coyne and Orr, 1997] Coyne, J. A. and Orr, H. A. (1997). "Patterns of speciation in *Drosophila*" revisited. *Evolution*, pages 295–303.
- [Coyne and Orr, 1998] Coyne, J. A. and Orr, H. A. (1998). The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 353(1366):287–305.
- [Coyne and Orr, 2004] Coyne, J. A. and Orr, H. A. (2004). *Speciation*, volume 37. Sinauer Associates Sunderland, MA.
- [Crane, 1955] Crane, J. (1955). Imaginal behaviour of a Trinidad butterfly, *Heliconius erato* hydra Hewitson, with special reference to the social use of color. *Zoologica, New York*, 40:167–196.
- [Crane, 1957] Crane, J. (1957). Imaginal behaviour in butterflies of the family Heliconiidae: changing social patterns and irrelevant actions. *Zoologica, New York*, 42:135–145.
- [Cresko et al., 2004] Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M. A., Kimmel, C. B., and Postlethwait, J. H. (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16):6050–6055.
- [Crespi, 1989] Crespi, B. J. (1989). Causes of assortative mating in arthropods. *Animal Behaviour*, 38(6):980–1000.
- [Cruickshank and Hahn, 2014] Cruickshank, T. E. and Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23(13):3133–3157.
- [Danecek et al., 2011] Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., and Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158.
- [Dapporto, 2007] Dapporto, L. (2007). Cuticular lipid diversification in *Lasiommata megera* and *Lasiommata paramegaura*: the influence of species, sex, and population (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society*, 91(4):703–710.
- [Darwin, 1859] Darwin, C. (1859). On the origins of species by means of natural selection. *London: Murray*, page 247.
- [de Visser et al., 1997] de Visser, J. A., Hoekstra, R. F., and van den Ende, H. (1997). An experimental test for synergistic epistasis and its application in *Chlamydomonas*. *Genetics*, 145(3):815–819.
- [DePristo et al., 2011] DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernysky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., and Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5):491–498.
- [Derryberry et al., 2014] Derryberry, E. P., Derryberry, G. E., Maley, J. M., and Brumfield, R. T. (2014). hzar: hybrid zone analysis using an R software package. *Molecular Ecology Resources*, 14(3):652–663.
- [Diehl and Bush, 1989] Diehl, S. R. and Bush, G. L. (1989). The role of habitat preference in adaptation and speciation. *Speciation and its Consequences*, pages 345–365.

- [Dilda and Mackay, 2002] Dilda, C. L. and Mackay, T. F. (2002). The genetic architecture of *Drosophila* sensory bristle number. *Genetics*, 162(4):1655–1674.
- [Dobler et al., 2012] Dobler, S., Dalla, S., Wagschal, V., and Agrawal, A. A. (2012). Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na, K-ATPase. *Proceedings of the National Academy of Sciences*, 109(32):13040–13045.
- [Dobzhansky and Dobzhansky, 1937] Dobzhansky, T. and Dobzhansky, T. G. (1937). *Genetics and the Origin of Species*. Columbia University Press.
- [Dopman et al., 2005] Dopman, E. B., Pérez, L., Bogdanowicz, S. M., and Harrison, R. G. (2005). Consequences of reproductive barriers for genealogical discordance in the European corn borer. *Proceedings of the National Academy of Sciences of the United States of America*, 102(41):14706–14711.
- [Dopman et al., 2010] Dopman, E. B., Robbins, P. S., and Seaman, A. (2010). Components of reproductive isolation between North American pheromone strains of the European corn borer. *Evolution*, 64(4):881–902.
- [Dowling et al., 1997] Dowling, T. E., Secor, and L. C. (1997). The Role of Hybridization and Introgression in the Diversification of Animals. *Annual Review of Ecology and Systematics*, 28(1):593–619.
- [Drès and Mallet, 2002] Drès, M. and Mallet, J. (2002). Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 357(1420):471–492.
- [Durand et al., 2011] Durand, E. Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for Ancient Admixture between Closely Related Populations. *Molecular Biology and Evolution*, 28(8):2239–2252.
- [Edwards, 1972] Edwards, A. W. F. (1972). *Likelihood*. Cambridge, UK: Cambridge University Press.
- [Ehrenreich et al., 2009] Ehrenreich, I. M., Hanzawa, Y., Chou, L., Roe, J. L., Kover, P. X., and Purugganan, M. D. (2009). Candidate gene association mapping of *Arabidopsis* flowering time. *Genetics*, 183(1):325–335.
- [Eizirik et al., 2003] Eizirik, E., Yuhki, N., Johnson, W. E., Menotti-Raymond, M., Hannah, S. S., and O'Brien, S. J. (2003). Molecular genetics and evolution of melanism in the cat family. *Current Biology*, 13(5):448–453.
- [Ellegren et al., 2012] Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., Uebbing, S., and Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, 491(7426):756–760.
- [Ellers and Boggs, 2003] Ellers, J. and Boggs, C. L. (2003). The evolution of wing color: male mate choice opposes adaptive wing color divergence in *Colias* butterflies. *Evolution*, 57(5):1100–1106.
- [Elmer et al., 2009] Elmer, K. R., Lehtonen, T. K., and Meyer, A. (2009). Color assortative mating contributes to sympatric divergence of neotropical cichlid fish. *Evolution*, 63(10):2750–2757.
- [Elmer and Meyer, 2011] Elmer, K. R. and Meyer, A. (2011). Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol Evol*, 26:298–306.
- [Emelianov et al., 2001] Emelianov, I., Drès, M., Baltensweiler, W., and Mallet, J. (2001). Host-induced assortative mating in host races of the larch budmoth. *Evolution; International Journal of Organic Evolution*, 55(10):2002–2010.
- [Emsley, 1963] Emsley, M. G. (1963). A morphological study of imagine Heliconiinae (Lep., Nymphalidae) with a consideration of the evolutionary relationships within the group. *Zoologica, New York*, 48:85–130.
- [Emsley, 1964] Emsley, M. G. (1964). The geographical distribution of the color-pattern components of *Heliconius erato* and *Heliconius melpomene* with genetical evidence for the systematic relationship between the two species. *Zoologica, New York*, 49:245–286.
- [Endler, 1977] Endler, J. A. (1977). *Geographic variation, speciation, and clines*. Princeton University Press.
- [Endler, 1986] Endler, J. A. (1986). *Natural selection in the wild*. Princeton University Press.



- [Estrada and Jiggins, 2008] Estrada, C. and Jiggins, C. D. (2008). Interspecific sexual attraction because of convergence in warning colouration: is there a conflict between natural and sexual selection in mimetic species? *Journal of Evolutionary Biology*, 21(3):749–760.
- [Estrada et al., 2011] Estrada, C., Schulz, S., Yildizhan, S., and Gilbert, L. E. (2011). Sexual selection drives the evolution of antiaphrodisiac pheromones in butterflies. *Evolution*, 65(10):2843–2854.
- [Etges and Ahrens, 2001] Etges, W. J. and Ahrens, M. A. (2001). Premating Isolation Is Determined by Larval-Rearing Substrates in Cactophilic *Drosophila mojavensis*. V. Deep Geographic Variation in Epicuticular Hydrocarbons among Isolated Populations\*. *The American Naturalist*, 158(6):585–598.
- [Feder, 1998] Feder, J. L. (1998). The apple maggot fly, *Rhagoletis pomonella*. *Endless forms: species and speciation*. Oxford Univ. Press, New York, pages 130–144.
- [Feder et al., 2003] Feder, J. L., Berlocher, S. H., Roethele, J. B., Dambroski, H., Smith, J. J., Perry, W. L., Gavrilovic, V., Filchak, K. E., Rull, J., and 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 et al., 2012] Feder, J. L., Egan, S. P., and Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28(7):342–350.
- [Feder and Filchak, 1999] Feder, J. L. and Filchak, K. E. (1999). It's about time: the evidence for host plant-mediated selection in the apple maggot fly, *rhagoletis pomonella*, and its implications for fitness trade-offs in phytophagous insects. In *Proceedings of the 10th International Symposium on Insect-Plant Relationships*, pages 211–225. Springer.
- [Felsenstein, 1981] Felsenstein, J. (1981). Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution*, pages 124–138.
- [Ferguson et al., 2010] Ferguson, L., Lee, S. F., Chamberlain, N., Nadeau, N., Joron, M., Baxter, S., Wilkinson, P., Papanicolaou, A., Kumar, S., Kee, T.-J., Clark, R., Davidson, C., Glithero, R., Beasley, H., Vogel, H., Ffrench-Constant, R., and Jiggins, C. (2010). Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the HmYb/Sb locus. *Molecular Ecology*, 19:240–254.
- [Ferveur, 2005] Ferveur, J.-F. (2005). Cuticular hydrocarbons: their evolution and roles in *Drosophila* pheromonal communication. *Behavior genetics*, 35(3):279–295.
- [Fierst and Hansen, 2010] Fierst, J. L. and Hansen, T. F. (2010). Genetic architecture and postzygotic reproductive isolation: evolution of bateson–dobzhansky–muller incompatibilities in a polygenic model. *Evolution*, 64(3):675–693.
- [Fisher, 1930] Fisher, R. A. (1930). *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press.
- [Ford, 1966] Ford, E. B. (1966). Genetic Polymorphism. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 164(995):350–361.
- [Fordyce et al., 2002] Fordyce, J. A., Nice, C. C., Forister, M. L., and 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.
- [Frankel et al., 2012] Frankel, N., Wang, S., and Stern, D. L. (2012). Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proceedings of the National Academy of Sciences*, 109(51):20975–20979.
- [Freeland and Boag, 1999] Freeland, J. R. and Boag, P. T. (1999). The mitochondrial and nuclear genetic homogeneity of the phenotypically diverse Darwin's ground finches. *Evolution*, pages 1553–1563.
- [Friberg et al., 2007] Friberg, M., Vongvanich, N., Borg-Karlson, A.-K., Kemp, D. J., Merilaita, S., and Wiklund, C. (2007). Female mate choice determines reproductive isolation between sympatric butterflies. *Behavioral Ecology and Sociobiology*, 62(6):873–886.
- [Gallant et al., 2014] Gallant, J. R., Imhoff, V. E., Martin, A., Savage, W. K., Chamberlain, N. L., Pote, B. L., Peterson, C., Smith, G. E., Evans, B., Reed, R. D., Kronforst, M. R., and Mullen, S. P. (2014). Ancient homology underlies adaptive mimetic diversity across butterflies. *Nature Communications*, 5.

- [Gavrilets, 2004] Gavrilets, S. (2004). *Fitness landscapes and the origin of species (MPB-41)*. Princeton University Press Princeton, NJ.
- [Gilbert, 2003] Gilbert, L. E. (2003). Adaptive novelty through introgression in *Heliconius* wing patterns: evidence for shared genetic "tool box" from synthetic hybrid zones and a theory of diversification. In Boggs, C. L., Watt, W. B., and Ehrlich, P., editors, *Butterflies: Ecology and Evolution Taking Flight*. Univ. of Chicago Press, Chicago.
- [Gillespie, 2004] Gillespie, R. (2004). Community assembly through adaptive radiation in Hawaiian spiders. *Science*, 303(5656):356–359.
- [Giraldo et al., 2008] Giraldo, N., Salazar, C., Jiggins, C. D., Bermingham, E., and Linares, M. (2008). Two sisters in the same dress: *Heliconius* cryptic species. *BMC Evolutionary Biology*, 8(1):324.
- [Grant and Grant, 1992] Grant, P. R. and Grant, B. R. (1992). Hybridization of Bird Species. *Science*, 256(5054):193–197.
- [Grant and Grant, 2008] Grant, P. R. and Grant, B. R. (2008). How and why species multiply. *The Radiation of Darwin's Finches. Princeton Series in Evolutionary Biology*.
- [Green et al., 2010] Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspina, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gušić, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox, C., Rasilla, M. d. I., Fortea, J., Rosas, A., Schmitz, R. W., Johnson, P. L. F., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., and Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, 328(5979):710–722.
- [Griswold, 2006] Griswold, C. K. (2006). Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses. *Heredity*, 96(6):445–453.
- [Gross et al., 2009] Gross, J. B., Borowsky, R., and Tabin, C. J. (2009). A novel role for Mc1r in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Genet*, 5(1):e1000326.
- [Haley and Knott, 1992] Haley, C. S. and Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 69(4):315–324.
- [Hansen, 2006] Hansen, T. F. (2006). The Evolution of Genetic Architecture. *Annual Review of Ecology, Evolution, and Systematics*, 37(1):123–157.
- [Harr, 2006] Harr, B. (2006). Genomic islands of differentiation between house mouse subspecies. *Genome Research*, 16(6):730–737.
- [Harrison, 1993] Harrison, R. G. (1993). *Hybrid Zones and the Evolutionary Process*. Oxford University Press.
- [Harrison and Bogdanowicz, 1997] Harrison, R. G. and Bogdanowicz, S. M. (1997). Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. *Evolution*, pages 493–505.
- [Hastings, 1970] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.
- [Hatfield and Schluter, 1999] Hatfield, T. and Schluter, D. (1999). Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution*, pages 866–873.
- [Hay-Roe et al., 2007] Hay-Roe, M. M., Lamas, G., and Nation, J. L. (2007). Pre- and postzygotic isolation and Haldane rule effects in reciprocal crosses of *Danaus erippus* and *Danaus plexippus* (Lepidoptera: Danainae), supported by differentiation of cuticular hydrocarbons, establish their status as separate species. *Biological Journal of the Linnean Society*, 91(3):445–453.
- [Heliconius Genome Consortium, 2012] Heliconius Genome Consortium (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, advance online publication.

- [Hendry et al., 2009] Hendry, A. P., Bolnick, D. I., Berner, D., and Peichel, C. L. (2009). Along the speciation continuum in sticklebacks. *Journal of Fish Biology*, 75(8):2000–2036.
- [Heuskin et al., 2014] Heuskin, S., Vanderplanck, M., Bacquet, P., Holveck, M.-J., Kaltenpoth, M., Engl, T., Pels, C., Taverne, C., Lognay, G., and Nieberding, C. M. (2014). The composition of cuticular compounds indicates body parts, sex and age in the model butterfly *Bicyclus anynana* (Lepidoptera). *Chemical Ecology*, 2:37.
- [Hill et al., 2013] Hill, R. I., Gilbert, L. E., and Kronforst, M. R. (2013). Cryptic genetic and wing pattern diversity in a mimetic *Heliconius* butterfly. *Molecular Ecology*, 22(10):2760–2770.
- [Hines et al., 2011] Hines, H. M., Countermand, B. A., Papa, R., Moura, P. A. d., Cardoso, M. Z., Linares, M., Mallet, J., Reed, R. D., Jiggins, C. D., Kronforst, M. R., and McMillan, W. O. (2011). Wing patterning gene redefines the mimetic history of *Heliconius* butterflies. *Proceedings of the National Academy of Sciences*, 108(49):19666–19671.
- [Hodges and Arnold, 1994] Hodges, S. A. and Arnold, M. L. (1994). Columbines: a geographically widespread species flock. *Proceedings of the National Academy of Sciences*, 91(11):5129–5132.
- [Hoekstra, 2006] Hoekstra, H. E. (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity*, 97(3):222–234.
- [Hoekstra et al., 2006] Hoekstra, H. E., Hirschmann, R. J., Bunday, R. A., Insel, P. A., and Crossland, J. P. (2006). A Single Amino Acid Mutation Contributes to Adaptive Beach Mouse Color Pattern. *Science*, 313(5783):101–104.
- [Hoekstra and Nachman, 2003] Hoekstra, H. E. and Nachman, M. W. (2003). Different genes underlie adaptive melanism in different populations of rock pocket mice. *Molecular Ecology*, 12(5):1185–1194.
- [Hohenlohe et al., 2010] Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., and Cresko, W. A. (2010). Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genet*, 6(2):e1000862.
- [Holzinger and Holzinger, 1994] Holzinger, H. and Holzinger, R. (1994). *Heliconius* and related genera. *Venette, Sciences Nat.* 328p. [Links].
- [Huber et al., 2015] Huber, B., Whibley, A., Poul, Y. L., Navarro, N., Martin, A., Baxter, S., Shah, A., Gilles, B., Wirth, T., McMillan, W. O., and Joron, M. (2015). Conservatism and novelty in the genetic architecture of adaptation in *Heliconius* butterflies. *Heredity*, 114(5):515–524.
- [Jakobsson et al., 2013] Jakobsson, M., Edge, M. D., and Rosenberg, N. A. (2013). The Relationship Between FST and the Frequency of the Most Frequent Allele. *Genetics*, 193(2):515–528.
- [James and Abbott, 2005] James, J. K. and Abbott, R. J. (2005). Recent, allopatric, homoploid hybrid speciation: the origin of *Senecio squalidus* (Asteraceae) in the British Isles from a hybrid zone on Mount Etna, Sicily. *Evolution*, 59(12):2533–2547.
- [Jiang et al., 2013] Jiang, Y., Bolnick, D. I., and Kirkpatrick, M. (2013). Assortative mating in animals. *The American Naturalist*, 181(6):E125–E138.
- [Jiggins, 2008] Jiggins, C. D. (2008). Ecological speciation in mimetic butterflies. *BioScience*, 58(6):541–548.
- [Jiggins et al., 2005a] Jiggins, C. D., Emelianov, I., and Mallet, J. (2005a). Assortative mating and speciation as pleiotropic effects of ecological adaptation: examples in moths and butterflies. In *SYMPOSIUM-ROYAL ENTOMOLOGICAL SOCIETY OF LONDON*, volume 22, page 451.
- [Jiggins et al., 2004] Jiggins, C. D., Estrada, C., and Rodrigues, A. (2004). Mimicry and the evolution of premating isolation in *Heliconius melpomene* Linnaeus. *Journal of Evolutionary Biology*, 17(3):680–691.
- [Jiggins and Mallet, 2000] Jiggins, C. D. and Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology & Evolution*, 15(6):250–255.
- [Jiggins et al., 2005b] Jiggins, C. D., Mavarez, J., Beltrán, M., McMillan, W. O., Johnston, J. S., and Bermingham, E. (2005b). A Genetic Linkage Map of the Mimetic Butterfly *Heliconius melpomene*. *Genetics*, 171(2):557–570.

- [Jiggins and McMillan, 1997] Jiggins, C. D. and McMillan, W. O. (1997). The genetic basis of an adaptive radiation: warning colour in two *Heliconius* species. *Proceedings of the Royal Society B: Biological Sciences*, 264(1385):1167–1175.
- [Jiggins et al., 1997] Jiggins, C. D., McMillan, W. O., King, P., and Mallet, J. (1997). The maintenance of species differences across a *Heliconius* hybrid zone. *Heredity*, 79(5):495–505.
- [Jiggins et al., 1996] Jiggins, C. D., McMillan, W. O., Neukirchen, W., and Mallet, J. (1996). What can hybrid zones tell us about speciation? The case of *Heliconius erato* and *H. himera* (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society*, 59(3):221–242.
- [Jiggins et al., 2001] Jiggins, C. D., Naisbit, R. E., Coe, R. L., and Mallet, J. (2001). Reproductive isolation caused by colour pattern mimicry. *Nature*, 411(6835):302–305.
- [Johnson et al., 2000] Johnson, T. C., Kelts, K., and Odada, E. (2000). The holocene history of Lake Victoria. *AMBIO: A Journal of the Human Environment*, 29(1):2–11.
- [Jones et al., 2012] Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., Brady, S. D., Zhang, H., Pollen, A. A., Howes, T., Amemiya, C., Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, Lander, E. S., Di Palma, F., Lindblad-Toh, K., and Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392):55–61.
- [Jones et al., 2011] Jones, R. T., Salazar, P. A., French Constant, R. H., Jiggins, C. D., and Joron, M. (2011). Evolution of a mimicry supergene from a multilocus architecture. *Proceedings. Biological sciences / The Royal Society*, 279(1727):316–325.
- [Jordan et al., 2005] Jordan, M. A., Snell, H. L., Snell, H. M., and Jordan, W. C. (2005). Phenotypic divergence despite high levels of gene flow in Galápagos lava lizards (*Microlophus albemarlensis*). *Molecular Ecology*, 14(3):859–867.
- [Joron et al., 2011] Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., Whibley, A., Becuwe, M., Baxter, S. W., Ferguson, L., Wilkinson, P. A., Salazar, C., Davidson, C., Clark, R., Quail, M. A., Beasley, H., Glithero, R., Lloyd, C., Sims, S., Jones, M. C., Rogers, J., Jiggins, C. D., and French Constant, R. H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, 477(7363):203–206.
- [Joron and Mallet, 1998] Joron, M. and Mallet, J. L. (1998). Diversity in mimicry: paradox or paradigm? *Trends in Ecology & Evolution*, 13(11):461–466.
- [Joron et al., 2006] Joron, M., Papa, R., Beltrán, M., Chamberlain, N., Mavárez, J., Baxter, S., Abanto, M., Bermingham, E., Humphray, S. J., Rogers, J., Beasley, H., Barlow, K., French Constant, R., Mallet, J., McMillan, W. O., and Jiggins, C. D. (2006). A Conserved Supergene Locus Controls Colour Pattern Diversity in *Heliconius* Butterflies. *PLoS Biology*, 4(10):e303.
- [Joron et al., 1999] Joron, M., Wynne, I. R., Lamas, G., and Mallet, J. (1999). Variable Selection and the Coexistence of Multiple mimetic forms of the Butterfly *Heliconius numata*. *Evolutionary Ecology*, 13(7-8):721–754.
- [Jurenka et al., 2003] Jurenka, R. A., Subchev, M., Abad, J.-L., Choi, M.-Y., and Fabrias, G. (2003). Sex pheromone biosynthetic pathway for disparlure in the gypsy moth, *Lymantria dispar*. *Proceedings of the National Academy of Sciences of the United States of America*, 100(3):809–814.
- [Kapan, 1998] Kapan, D. D. (1998). Divergent natural selection and müllerian mimicry in polymorphic *Heliconius cydno* (Lepidoptera: Nymphalidae). Doctoral dissertation, University of British Columbia.
- [Kapan, 2001] Kapan, D. D. (2001). Three-butterfly system provides a field test of müllerian mimicry. *Nature*, 409(6818):338–340.
- [Kapan et al., 2006] Kapan, D. D., Flanagan, N. S., Tobler, A., Papa, R., Reed, R. D., Gonzalez, J. A., Restrepo, M. R., Martinez, L., Maldonado, K., Ritschoff, C., Heckel, D. G., and McMillan, W. O. (2006). Localization of Müllerian Mimicry Genes on a Dense Linkage Map of *Heliconius erato*. *Genetics*, 173(2):735–757.
- [Karlson and Lüscher, 1959] Karlson, P. and Lüscher, M. (1959). ‘pheromones’: a new term for a class of biologically active substances. *Nature*, 183:55–56.

- [Kearsey and Farquhar, 1998] Kearsey, M. and Farquhar, A. (1998). QTL analysis in plants; where are we now? *Heredity*, 80(2):137–142.
- [Kemp, 2007] Kemp, D. J. (2007). Female butterflies prefer males bearing bright iridescent ornamentation. *Proceedings of the Royal Society B: Biological Sciences*, 274(1613):1043–1047.
- [Kirkpatrick and Ravigné, 2002] Kirkpatrick, M. and Ravigné, V. (2002). Speciation by natural and sexual selection: models and experiments. *The American Naturalist*, 159(S3):S22–S35.
- [Klein and Araújo, 2010] Klein, A. L. and Araújo, A. M. d. (2010). Courtship behavior of *Heliconius erato phyllis* (Lepidoptera, Nymphalidae) towards virgin and mated females: conflict between attraction and repulsion signals? *Journal of Ethology*, 28(3):409–420.
- [Knott and Haley, 2000] Knott, S. A. and Haley, C. S. (2000). Multitrait least squares for quantitative trait loci detection. *Genetics*, 156(2):899–911.
- [Kondrashov and Mina, 1986] Kondrashov, A. S. and Mina, M. V. (1986). Sympatric speciation: when is it possible? *Biological Journal of the Linnean Society*, 27(3):201–223.
- [Kondrashov and Shpak, 1998] Kondrashov, A. S. and Shpak, M. (1998). On the origin of species by means of assortative mating. *Proceedings of the Royal Society B: Biological Sciences*, 265(1412):2273–2278.
- [Kopp and Hermisson, 2006] Kopp, M. and Hermisson, J. (2006). The evolution of genetic architecture under frequency-dependent disruptive selection. *Evolution; international journal of organic evolution*, 60(8):1537–1550.
- [Kozak et al., 2015] Kozak, K. M., Wahlberg, N., Neild, A. F. E., Dasmahapatra, K. K., Mallet, J., and Jiggins, C. D. (2015). Multilocus Species Trees Show the Recent Adaptive Radiation of the Mimetic *Heliconius* Butterflies. *Systematic Biology*, 64(3):505–524.
- [Kraus et al., 2012] Kraus, R. H., Kerstens, H. H., Hooft, P. v., Megens, H.-J., Elmberg, J., Tsvey, A., Sartakov, D., Soloviev, S. A., Crooijmans, R. P., Groenen, M. A., Ydenberg, R. C., and Prins, H. H. (2012). Widespread horizontal genomic exchange does not erode species barriers among sympatric ducks. *BMC Evolutionary Biology*, 12(1):45.
- [Kronforst et al., 2013] Kronforst, M., Hansen, M. B., Crawford, N., Gallant, J., Zhang, W., Kulathinal, R., Kapan, D., and Mullen, S. (2013). Hybridization Reveals the Evolving Genomic Architecture of Speciation. *Cell Reports*, 5(3):666–677.
- [Kronforst et al., 2012] Kronforst, M. R., Barsh, G. S., Kopp, A., Mallet, J., Monteiro, A., Mullen, S. P., Protas, M., Rosenblum, E. B., Schneider, C. J., and Hoekstra, H. E. (2012). Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. *Pigment cell & melanoma research*, 25(4):411–433.
- [Kronforst and Gilbert, 2008] Kronforst, M. R. and Gilbert, L. E. (2008). The population genetics of mimetic diversity in *Heliconius* butterflies. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1634):493–500.
- [Kronforst et al., 2006a] Kronforst, M. R., Kapan, D. D., and Gilbert, L. E. (2006a). Parallel Genetic Architecture of Parallel Adaptive Radiations in Mimetic *Heliconius* Butterflies. *Genetics*, 174(1):535–539.
- [Kronforst et al., 2007] Kronforst, M. R., Young, L., and Gilbert, L. (2007). Reinforcement of mate preference among hybridizing *Heliconius* butterflies. *Journal of evolutionary biology*, 20(1):278–285.
- [Kronforst et al., 2006b] Kronforst, M. R., Young, L. G., Blume, L. M., and Gilbert, L. E. (2006b). Multilocus Analyses of Admixture and Introgression Among Hybridizing *Heliconius* Butterflies. *Evolution*, 60(6):1254–1268.
- [Kronforst et al., 2006c] Kronforst, M. R., Young, L. G., Kapan, D. D., McNeely, C., O'Neill, R. J., and Gilbert, L. E. (2006c). Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proceedings of the National Academy of Sciences*, 103(17):6575–6580.
- [Kroymann and Mitchell-Olds, 2005] Kroymann, J. and Mitchell-Olds, T. (2005). Epistasis and balanced polymorphism influencing complex trait variation. *Nature*, 435(7038):95–98.

- [Lair et al., 1997] Lair, K. P., Bradshaw, W. E., and Holzapfel, C. M. (1997). Evolutionary Divergence of the Genetic Architecture Underlying Photoperiodism in the Pitcher-Plant Mosquito, *Wyeomyia Smithii*. *Genetics*, 147(4):1873.
- [Lande, 1980] Lande, R. (1980). The Genetic Covariance Between Characters Maintained by Pleiotropic Mutations. *Genetics*, 94(1):203–215.
- [Lander and Botstein, 1989] Lander, E. S. and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121(1):185–199.
- [Landry et al., 2007] Landry, L., Vincent, W., and Bernatchez, L. (2007). Parallel evolution of lake whitefish dwarf ecotypes in association with limnological features of their adaptive landscape. *Journal of evolutionary biology*, 20(3):971–984.
- [Laurie et al., 2004] Laurie, C. C., Chasalow, S. D., LeDeaux, J. R., McCarroll, R., Bush, D., Hauge, B., Lai, C., Clark, D., Rocheford, T. R., and Dudley, J. W. (2004). The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics*, 168(4):2141–2155.
- [Lawniczak et al., 2010] Lawniczak, M. K. N., Emrich, S. J., Holloway, A. K., Regier, A. P., Olson, M., White, B., Redmond, S., Fulton, L., Appelbaum, E., Godfrey, J., Farmer, C., Chinwalla, A., Yang, S.-P., Minx, P., Nelson, J., Kyung, K., Walenz, B. P., Garcia-Hernandez, E., Aguiar, M., Viswanathan, L. D., Rogers, Y.-H., Strausberg, R. L., Sasaki, C. A., Lawson, D., Collins, F. H., Kafatos, F. C., Christophides, G. K., Clifton, S. W., Kirkness, E. F., and Besansky, N. J. (2010). Widespread Divergence Between Incipient *Anopheles gambiae* Species Revealed by Whole Genome Sequences. *Science*, 330(6003):512–514.
- [Le Poul, 2014] Le Poul, Y. (2014). *Selection for mimicry in butterflies: Quantitative approaches on colour pattern resemblance and diversity*. PhD thesis, Museum National d'Histoire Naturelle, Paris.
- [Le Poul et al., 2014] Le Poul, Y., Whibley, A., Chouteau, M., Prunier, F., Llaurens, V., and Joron, M. (2014). Evolution of dominance mechanisms at a butterfly mimicry supergene. *Nature Communications*, 5.
- [Leamy et al., 2008] Leamy, L., Klingenberg, C., Sherratt, E., Wolf, J., and Cheverud, J. (2008). A search for quantitative trait loci exhibiting imprinting effects on mouse mandible size and shape. *Heredity*, 101(6):518–526.
- [Lecocq et al., 2013] Lecocq, T., Vereecken, N. J., Michez, D., Dellicour, S., Lhomme, P., Valterová, I., Rasplus, J.-Y., and Rasmont, P. (2013). Patterns of Genetic and Reproductive Traits Differentiation in Mainland vs. Corsican Populations of Bumblebees. *PLoS ONE*, 8(6):e65642.
- [Lewontin and Birch, 1966] Lewontin, R. and Birch, L. (1966). Hybridization as a source of variation for adaptation to new environments. *Evolution*, pages 315–336.
- [Li et al., 2009] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.
- [Linares, 1996] Linares, M. (1996). The Genetics of the Mimetic Coloration in the Butterfly *Heliconius cydno weymeri*. *Journal of Heredity*, 87(2):142–149.
- [Linares, 1997] Linares, M. (1997). Origin of neotropical mimetic biodiversity from a threeway hybrid zone of *Heliconius cydno* butterflies. In H. Ulrich, pages 93–108. Proceedings of the International Symposium on Biodiversity and Systematics in Tropical Ecosystems. Zoologisches Forschungsinstitut und Museum Alexander Koenig, Bonn, tropical biodiversity and systematics. proceedings of the international symposium on biodiversity and systematics in tropical ecosystems. edition.
- [Linnen and Hoekstra, 2009] Linnen, C. R. and Hoekstra, H. E. (2009). Measuring natural selection on genotypes and phenotypes in the wild. volume 74, pages 155–168. Cold Spring Harbor Laboratory Press.
- [Linnen et al., 2009] Linnen, C. R., Kingsley, E. P., Jensen, J. D., and Hoekstra, H. E. (2009). On the Origin and Spread of an Adaptive Allele in Deer Mice. *Science*, 325(5944):1095–1098.
- [Llaurens et al., 2014] Llaurens, V., Joron, M., and Théry, M. (2014). Cryptic differences in colour among Müllerian mimics: how can the visual capacities of predators and prey shape the evolution of wing colours? *Journal of Evolutionary Biology*.

- [Lofstedt, 1993] Lofstedt, C. (1993). Moth pheromone genetics and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 340(1292):167–177.
- [Losos, 1992] Losos, J. B. (1992). The Evolution of Convergent Structure in Caribbean Anolis Communities. *Systematic Biology*, 41(4):403–420.
- [Losos, 2011] Losos, J. B. (2011). Convergence, Adaptation, and Constraint. *Evolution*, 65(7):1827–1840.
- [Losos and De Queiroz, 1997] Losos, J. B. and De Queiroz, K. (1997). Evolutionary consequences of ecological release in Caribbean Anolis lizards. *Biological Journal of the Linnean Society*, 61(4):459–483.
- [Loudon and Koehl, 2000] Loudon, C. and Koehl, M. A. (2000). Sniffing by a silkworm moth: wing fanning enhances air penetration through and pheromone interception by antennae. *Journal of Experimental Biology*, 203(19):2977–2990.
- [Lundgren and Bergström, 1975] Lundgren, L. and Bergström, G. (1975). Wing scents and scent-released phases in the courtship behavior of *Lycaeides argyrognomon* (Lepidoptera: Lycaenidae). *Journal of Chemical Ecology*, 1(4):399–412.
- [Lunter and Goodson, 2011] Lunter, G. and Goodson, M. (2011). Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research*, 21(6):936–939.
- [MacCallum et al., 1998] MacCallum, C. J., Nürnberger, B., Barton, N., and Szymura, J. (1998). Habitat preference in the *Bombina* hybrid zone in Croatia. *Evolution*, pages 227–239.
- [Magnus, 1958] Magnus, D. B. (1958). Experimental analysis of some ‘over-optimal’ sign-stimuli in the mating behavior of the fritillary butterfly *Argynnis paphia* L. (Lepidoptera, Nymphalidae). *Proceedings of the 10th International Congress on Entomology*, 2:405–418.
- [Malausa et al., 2005] Malausa, T., Bethenod, M.-T., Bontemps, A., Bourguet, D., Cornuet, J.-M., and Ponsard, S. (2005). Assortative Mating in Sympatric Host Races of the European Corn Borer. *Science*, 308(5719):258–260.
- [Mallet, 1986] Mallet, J. (1986). Heredity - Abstract of article: Hybrid zones of *Heliconius* butterflies in Panama and the stability and movement of warning colour clines. *Heredity*, 56(2):191–202.
- [Mallet, 1989] Mallet, J. (1989). The Genetics of Warning Colour in Peruvian Hybrid Zones of *Heliconius erato* and *H. melpomene*. *Proceedings of the Royal Society of London. B. Biological Sciences*, 236(1283):163–185.
- [Mallet, 1993] Mallet, J. (1993). Speciation, Racialization, and Color Pattern Evolution in *Heliconius* Butterflies: Evidence from Hybrid Zones. In *Hybrid Zones and the Evolutionary Process.*, pages 226–260. Oxford University Press, New York, r.g. harrison edition.
- [Mallet, 1995] Mallet, J. (1995). A species definition for the modern synthesis. *Trends in Ecology & Evolution*, 10(7):294–299.
- [Mallet, 2005] Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, 20(5):229–237.
- [Mallet, 2009] Mallet, J. (2009). Rapid speciation, hybridization and adaptive radiation in the *Heliconius melpomene* group. In Butlin, R., Bridle, J., and Schluter, D., editors, *Speciation and patterns of diversity*, pages 177–194. Cambridge Univ Pr, Cambridge.
- [Mallet, 2010] Mallet, J. (2010). Shift happens! Shifting balance and the evolution of diversity in warning colour and mimicry. *Ecological Entomology*, 35(s1):90–104.
- [Mallet and Barton, 1989a] Mallet, J. and Barton, N. (1989a). Inference from clines stabilized by frequency-dependent selection. *Genetics*, 122(4):967–976.
- [Mallet and Barton, 1989b] Mallet, J. and Barton, N. H. (1989b). Strong natural selection in a warning-color hybrid zone. *Evolution*, pages 421–431.
- [Mallet et al., 2007] Mallet, J., Beltrán, M., Neukirchen, W., and Linares, M. (2007). Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology*, 7(1):28.

- [Mallet and Gilbert, 1995] Mallet, J. and Gilbert, L. E. (1995). Why are there so many mimicry rings? Correlations between habitat, behaviour and mimicry in *Heliconius* butterflies. *Biological Journal of the Linnean Society*, 55(2):159–180.
- [Mallet et al., 1998a] Mallet, J., McMillan, W. O., and Jiggins, C. (1998a). Mimicry and Warning Color at the Boundary between Races and Species. In *Endless Forms: Species and Speciation*, pages 390–403.
- [Mallet et al., 1998b] Mallet, J., McMillan, W. O., and Jiggins, C. D. (1998b). Estimating the mating behavior of a pair of hybridizing *Heliconius* species in the wild. *Evolution*, pages 503–510.
- [Mallet et al., 2009] Mallet, J., Meyer, A., Nosil, P., and Feder, J. L. (2009). Space, sympatry and speciation. *Journal of evolutionary biology*, 22(11):2332–2341.
- [Manceau et al., 2010] Manceau, M., Domingues, V. S., Linnen, C. R., Rosenblum, E. B., and Hoekstra, H. E. (2010). Convergence in pigmentation at multiple levels: mutations, genes and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1552):2439–2450.
- [Manichaikul et al., 2006] Manichaikul, A., Dupuis, J., Sen, Ś., and Broman, K. W. (2006). Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. *Genetics*, 174(1):481–489.
- [Marco et al., 1998] Marco, A., Chivers, D. P., Kiesecker, J. M., and Blaustein, A. R. (1998). Mate choice by chemical cues in western redback (*Plethodon vehiculum*) and Dunn’s (*P. dunni*) salamanders. *Ethology*, 104(9):781–788.
- [Marek and Bond, 2009] Marek, P. E. and Bond, J. E. (2009). A Müllerian mimicry ring in Appalachian millipedes. *Proceedings of the National Academy of Sciences*, 106(24):9755–9760.
- [Martin et al., 2014] Martin, A., McCulloch, K. J., Patel, N. H., Briscoe, A. D., Gilbert, L. E., and Reed, R. D. (2014). Multiple recent co-options of *Optix* associated with novel traits in adaptive butterfly wing radiations. *EvoDevo*, 5(1):7.
- [Martin and Orgogozo, 2013] Martin, A. and Orgogozo, V. (2013). The Loci of Repeated Evolution: A Catalog of Genetic Hotspots of Phenotypic Variation. *Evolution*, 67(5):1235–1250.
- [Martin et al., 2012] Martin, A., Papa, R., Nadeau, N. J., Hill, R. I., Counterman, B. A., Halder, G., Jiggins, C. D., Kronforst, M. R., Long, A. D., McMillan, W. O., and 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):12632–12637.
- [Martin and Reed, 2014] Martin, A. and Reed, R. D. (2014). Wnt signaling underlies evolution and development of the butterfly wing pattern symmetry systems. *Developmental Biology*, 395(2):367–378.
- [Martin et al., 2013] Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., and Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, 23(11):1817–1828.
- [Martin et al., 2015] Martin, S. H., Davey, J. W., and Jiggins, C. D. (2015). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution*, 32(1):244–257.
- [Martín and López, 2000] Martín, J. and López, P. (2000). Chemoreception, symmetry and mate choice in lizards. *Proceedings of the Royal Society of London B: Biological Sciences*, 267(1450):1265–1269.
- [Mavárez and González, 2006] Mavárez, J. and González, M. (2006). A set of microsatellite markers for *Heliconius melpomene* and closely related species. *Molecular Ecology Notes*, 6(1):20–23.
- [Mavárez et al., 2006] Mavárez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., and Linares, M. (2006). Speciation by hybridization in *Heliconius* butterflies. *Nature*, 441(7095):868–871.
- [Mayr, 1942] Mayr, E. (1942). *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.
- [McBride and Singer, 2010] McBride, C. S. and Singer, M. C. (2010). Field studies reveal strong postmating isolation between ecologically divergent butterfly populations. *PLoS biology*, 8(10):2360.



- [McDonald, 1994] McDonald, J. H. (1994). Detecting Natural Selection by Comparing Geographic Variation in Protein and DNA Polymorphisms. In Golding, B., editor, *Non-Neutral Evolution*, pages 88–100. Springer US.
- [Mega and Araújo, 2010] Mega, N. O. and Araújo, A. M. d. (2010). Analysis of the mating behavior and some possible causes of male copulatory success in *Dryas iulia alcionea* (Lepidoptera, Nymphalidae, Heliconiinae). *Journal of Ethology*, 28(1):123–132.
- [Melo et al., 2009] Melo, M. C., Salazar, C., Jiggins, C. D., and Linares, M. (2009). Assortative mating preferences among hybrids offers a route to hybrid speciation. *Evolution*, 63(6):1660–1665.
- [Merchán et al., 2005] Merchán, H. A., Jiggins, C. D., and Linares, M. (2005). A narrow *Heliconius cydno* (Nymphalidae; Heliconiini) hybrid zone with differences in morph sex ratios. *Biotropica*, 37:119–128.
- [Merrill et al., 2015] Merrill, R. M., Dasmahapatra, K. K., Davey, J. W., Dell’Aglio, D. D., Hanly, J. J., Huber, B., Jiggins, C. D., Joron, M., Kozak, K. M., Llaurens, V., Martin, S. H., Montgomery, S. H., Morris, J., Nadeau, N. J., Pinharanda, A. L., Rosser, N., Thompson, M. J., Vanjari, S., Wallbank, R. W. R., and Yu, Q. (2015). The diversification of *Heliconius* butterflies: what have we learned in 150 years? *Journal of Evolutionary Biology*, pages n/a–n/a.
- [Merrill et al., 2011a] Merrill, R. M., Gompert, Z., Dembeck, L. M., Kronforst, M. R., McMillan, W. O., and Jiggins, C. D. (2011a). Mate preference across the speciation continuum in a clade of mimetic butterflies. *Evolution*, 65(5):1489–1500.
- [Merrill et al., 2011b] Merrill, R. M., Van Schooten, B., Scott, J. A., and Jiggins, C. D. (2011b). Pervasive genetic associations between traits causing reproductive isolation in *Heliconius* butterflies. *Proceedings of the Royal Society of London B: Biological Sciences*, 278(1705):511–518.
- [Merrill et al., 2012] Merrill, R. M., Wallbank, R. W. R., Bull, V., Salazar, P. C. A., Mallet, J., Stevens, M., and Jiggins, C. D. (2012). Disruptive ecological selection on a mating cue. *Proceedings of the Royal Society of London B: Biological Sciences*, 279(1749):4907–4913.
- [Metropolis et al., 1953] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092.
- [Meyer et al., 1990] Meyer, A., Kocher, T. D., Basasibwaki, P., and Wilson, A. C. (1990). Monophyletic origin of lake victoria cichlid fishes suggested by mitochondrial dna sequences. *Nature*.
- [Michel et al., 2010] Michel, A. P., Sim, S., Powell, T. H. Q., Taylor, M. S., Nosil, P., and Feder, J. L. (2010). Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences*, 107(21):9724–9729.
- [Millar, 2000] Millar, J. G. (2000). Polyene hydrocarbons and epoxides: A Second Major Class of Lepidopteran Sex Attractant Pheromones. *Annual Review of Entomology*, 45(1):575–604.
- [Mullen and Hoekstra, 2008] Mullen, L. M. and Hoekstra, H. E. (2008). Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution; International Journal of Organic Evolution*, 62(7):1555–1570.
- [Mundy, 2005] Mundy, N. I. (2005). A window on the genetics of evolution: MC1r and plumage colouration in birds. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1573):1633–1640.
- [Mérot et al., 2015] Mérot, C., Frérot, B., Leppik, E., and Joron, M. (2015). Beyond magic traits: multimodal mating cues in *Heliconius* butterflies. *Evolution*.
- [Mérot et al., 2013] Mérot, C., Mavárez, J., Evin, A., Dasmahapatra, K. K., Mallet, J., Lamas, G., and Joron, M. (2013). Genetic differentiation without mimicry shift in a pair of hybridizing *Heliconius* species (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society*, 109(4):830–847.
- [Müller, 1879] Müller, F. (1879). *Ituna* and *Thyridia*: a remarkable case of mimicry in butterflies. *Trans. Entomol. Soc. Lond.*, 1879:20–29.
- [Nabours, 1933] Nabours, R. K. (1933). Inheritance of Color Patterns in the Grouse Locust *ACRYDIUM ARENOSUM* Burmeister (Tettigidae). *Genetics*, 18(2):159–171.

- [Nadeau et al., 2015] Nadeau, N., Pardo-Diaz, C., Whibley, A., Supple, M. A., Wallbank, R., Wu, G. C., Maroja, L., Ferguson, L., Hines, H., Salazar, C., French Constant, R., Joron, M., McMillan, W. O., and Jiggins, C. (2015). The origins of a novel butterfly wing patterning gene from within a family of conserved cell cycle regulators. *bioRxiv*, page 016006.
- [Nadeau and Jiggins, 2010] Nadeau, N. J. and Jiggins, C. D. (2010). A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends Genet*, 26:484–492.
- [Nadeau et al., 2013] Nadeau, N. J., Martin, S. H., Kozak, K. M., Salazar, C., Dasmahapatra, K. K., Davey, J. W., Baxter, S. W., Blaxter, M. L., Mallet, J., and Jiggins, C. D. (2013). Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology*, 22(3):814–826.
- [Nadeau et al., 2014] Nadeau, N. J., Ruiz, M., Salazar, P., Counterman, B., Medina, J. A., Ortiz-Zuazaga, H., Morrison, A., McMillan, W. O., Jiggins, C. D., and Papa, R. (2014). Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Research*, page gr.169292.113.
- [Nadeau et al., 2012] Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., Quail, M. A., Joron, M., French Constant, R. H., Blaxter, M. L., Mallet, J., and Jiggins, C. D. (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1587):343–353.
- [Nagl et al., 2000] Nagl, S., Tichy, H., Mayer, W. E., Takezaki, N., Takahata, N., and Klein, J. (2000). The origin and age of haplochromine fishes in Lake Victoria, East Africa. *Proceedings of the Royal Society of London B: Biological Sciences*, 267(1447):1049–1061.
- [Naisbit et al., 2002] Naisbit, R. E., Jiggins, C. D., Linares, M., Salazar, C., and Mallet, J. (2002). Hybrid sterility, Haldane’s rule and speciation in *Heliconius cydno* and *H. melpomene*. *Genetics*, 161(4):1517.
- [Naisbit et al., 2003] Naisbit, R. E., Jiggins, C. D., and Mallet, J. (2003). Mimicry: developmental genes that contribute to speciation. *Evolution & development*, 5(3):269–280.
- [Nieberding et al., 2008] Nieberding, C. M., de Vos, H., Schneider, M. V., Lassance, J.-M., Estramil, N., Andersson, J., Bång, J., Hedenström, E., Löfstedt, C., and Brakefield, P. M. (2008). The Male Sex Pheromone of the Butterfly *Bicyclus anynana*: Towards an Evolutionary Analysis. *PLoS ONE*, 3(7):e2751.
- [Nielsen, 2005] Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.*, 39:197–218.
- [Nijhout et al., 1990] Nijhout, H. F., Wray, G. A., and 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.
- [Noor and Feder, 2006] Noor, M. A. and Feder, J. L. (2006). Speciation genetics: evolving approaches. *Nature Reviews Genetics*, 7(11):851–861.
- [Noor et al., 2000] Noor, M. A., Johnson, N. A., and Hey, J. (2000). Gene flow between *Drosophila pseudoobscura* and *D. persimilis*. *Evolution*, 54(6):2174–2175.
- [Noor and Bennett, 2009] Noor, M. A. F. and Bennett, S. M. (2009). Islands of Speciation or Mirages in the Desert? Examining the Role of Restricted Recombination in Maintaining Species. *Heredity*, 103(6):439–444.
- [Nosil, 2004] Nosil, P. (2004). Reproductive isolation caused by visual predation on migrants between divergent environments. *Proceedings of the Royal Society of London B: Biological Sciences*, 271(1547):1521–1528.
- [Nosil, 2007] Nosil, P. (2007). Divergent host plant adaptation and reproductive isolation between ecotypes of *Timema cristinae* walking sticks. *The American Naturalist*, 169(2):151–162.
- [Nosil and Crespi, 2006] Nosil, P. and Crespi, B. J. (2006). Experimental evidence that predation promotes divergence in adaptive radiation. *Proceedings of the National Academy of Sciences*, 103(24):9090–9095.

- [Nosil et al., 2008] Nosil, P., Egan, S. P., and Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution; International Journal of Organic Evolution*, 62(2):316–336.
- [Nosil et al., 2009] Nosil, P., Harmon, L. J., and Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, 24(3):145–156.
- [Nosil and Schluter, 2011] Nosil, P. and Schluter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology & Evolution*, 26(4):160–167.
- [Obara, 1964] Obara, Y. (1964). Mating behavior of the cabbage butterfly, *Pieris rapae crucivora*. II. The 'mate-refusal posture' of the female. *Zool. Mag. (Dobutsugaku-Zasshi)*, 73:175–178.
- [Orr, 1998] Orr, H. A. (1998). Testing Natural Selection vs. Genetic Drift in Phenotypic Evolution Using Quantitative Trait Locus Data. *Genetics*, 149(4):2099–2104.
- [Orr, 2005] Orr, H. A. (2005). The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, 6(2):119–127.
- [Orr and Turelli, 2001] Orr, H. A. and Turelli, M. (2001). The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution*, 55(6):1085–1094.
- [Orr and Smith, 1998] Orr, M. R. and Smith, T. B. (1998). Ecology and speciation. *Trends in Ecology & Evolution*, 13(12):502–506.
- [Papa et al., 2013] Papa, R., Kapan, D. D., Counterman, B. A., Maldonado, K., Lindstrom, D. P., Reed, R. D., Nijhout, H. F., Hrbek, T., and McMillan, W. O. (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 et al., 2008] Papa, R., Martin, A., and Reed, R. D. (2008). Genomic hotspots of adaptation in butterfly wing pattern evolution. *Current Opinion in Genetics & Development*, 18(6):559–564.
- [Pardo-Diaz et al., 2012] Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W. O., and Jiggins, C. D. (2012). Adaptive Introgression across Species Boundaries in *Heliconius* Butterflies. *PLoS Genet*, 8(6):e1002752.
- [Parsons et al., 1993] Parsons, T. J., Olson, S. L., and Braun, M. J. (1993). Unidirectional spread of secondary sexual plumage traits across an avian hybrid zone. *Science (New York, N.Y.)*, 260(5114):1643–1646.
- [Paterson, 1985] Paterson, H. E. (1985). The recognition concept of species.
- [Pearson, 2000] Pearson, S. F. (2000). Behavioral asymmetries in a moving hybrid zone. *Behavioral Ecology*, 11(1):84–92.
- [Pfennig et al., 2001] Pfennig, D. W., Harcombe, W. R., and Pfennig, K. S. (2001). Frequency-dependent Batesian mimicry. *Nature*, 410(6826):323–323.
- [Phillips, 2008] Phillips, P. C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, 9(11):855–867.
- [Pillai, 1967] Pillai, K. S. (1967). Upper percentage points of the largest root of a matrix in multivariate analysis. *Biometrika*, 54(1-2):189–194.
- [Pivnick et al., 1992] Pivnick, K. A., Lavoie-Dornik, J., and McNeil, J. N. (1992). The role of the androconia in the mating behaviour of the European skipper, *Thymelicus lineola*, and evidence for a male sex pheromone. *Physiological entomology*, 17(3):260–268.
- [Pliske, 1975] Pliske, T. E. (1975). Courtship Behavior of the Monarch Butterfly, *Danaus plexippus* L. *Annals of the Entomological Society of America*, 68(1):143–151.
- [Podos, 2001] Podos, J. (2001). Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature*, 409(6817):185–188.
- [Powell et al., 2013] Powell, T. H. Q., Hood, G. R., Murphy, M. O., Heilveil, J. S., Berlocher, S. H., Nosil, P., and Feder, J. L. (2013). Genetic divergence along the speciation continuum: the transition from host race to species in *rhagoletis* (Diptera: tephritidae). *Evolution; International Journal of Organic Evolution*, 67(9):2561–2576.

- [Pritchard et al., 2000] Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959.
- [Protas et al., 2006] Protas, M. E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W. R., Zon, L. I., Borowsky, R., and Tabin, C. J. (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics*, 38(1):107–111.
- [Puebla et al., 2014] Puebla, O., Bermingham, E., and McMillan, W. O. (2014). Genomic atolls of differentiation in coral reef fishes (*Hypoplectrus* spp., Serranidae). *Molecular Ecology*, 23(21):5291–5303.
- [Pyke et al., 2001] Pyke, C. R., Condit, R., Aguilar, S., and Lao, S. (2001). Floristic Composition across a Climatic Gradient in a Neotropical Lowland Forest. *Journal of Vegetation Science*, 12(4):553–566.
- [Pérez i de Lanuza G, 2013] Pérez i de Lanuza G, E. F. (2013). Color-assortative mating in a color-polymorphic lacertid lizard. *Behavioral Ecology*, 24(1):273–279.
- [Rajon and Plotkin, 2013] Rajon, E. and Plotkin, J. B. (2013). The evolution of genetic architectures underlying quantitative traits. *Proceedings of the Royal Society B: Biological Sciences*, 280(1769):20131552.
- [Ramsey et al., 2003] Ramsey, J., Bradshaw, H., and Schemske, D. W. (2003). Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution*, 57(7):1520–1534.
- [Reed et al., 2011] Reed, R. D., Papa, R., Martin, A., Hines, H. M., Counterman, B. A., Pardo-Diaz, C., Jiggins, C. D., Chamberlain, N. L., Kronforst, M. R., Chen, R., Halder, G., Nijhout, H. F., and McMillan, W. O. (2011). *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science (New York, N.Y.)*, 333(6046):1137–1141.
- [Rhymer and Simberloff, 1996] Rhymer, J. M. and Simberloff, D. (1996). Extinction by Hybridization and Introgression. *Annual Review of Ecology and Systematics*, 27(1):83–109.
- [Rice and Hostert, 1993] Rice, W. R. and Hostert, E. E. (1993). Laboratory experiments on speciation: what have we learned in 40 years? *Evolution*, pages 1637–1653.
- [Rieseberg et al., 1999] Rieseberg, L. H., Archer, M. A., and Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity*, 83(4):363–372.
- [Robertson, 1967] Robertson, A. (1967). The nature of quantitative genetic variation. *Heritage from Mendel*, pages 265–280.
- [Robertson and Monteiro, 2005] Robertson, K. A. and Monteiro, A. (2005). Female *Bicyclus anynana* butterflies choose males on the basis of their dorsal UV-reflective eyespot pupils. *Proceedings of the Royal Society B: Biological Sciences*, 272(1572):1541–1546.
- [Roelofs and Cardé, 1971] Roelofs, W. L. and Cardé, R. T. (1971). Hydrocarbon sex pheromone in tiger moths (Arctiidae). *Science (New York, N.Y.)*, 171(3972):684–686.
- [Rogers and Bernatchez, 2007] Rogers, S. M. and Bernatchez, L. (2007). The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp. Salmonidae) species pairs. *Molecular Biology and Evolution*, 24(6):1423–1438.
- [Rogers et al., 2013] Rogers, S. M., Mee, J. A., and Bowles, E. (2013). The consequences of genomic architecture on ecological speciation in postglacial fishes. *Current Zoology*, 59(1):53–71.
- [Rolán-Alvarez et al., 1997] Rolán-Alvarez, E., Johannesson, K., and Erlandsson, J. (1997). The maintenance of a cline in the marine snail *Littorina saxatilis*: the role of home site advantage and hybrid fitness. *Evolution*, pages 1838–1847.
- [Rosenblum et al., 2004] Rosenblum, E. B., Hoekstra, H. E., and Nachman, M. W. (2004). Adaptive reptile color variation and the evolution of the MC1R gene. *Evolution*, 58(8):1794–1808.
- [Rosser et al., 2014] Rosser, N., Dasmahapatra, K. K., and Mallet, J. (2014). Stable *Heliconius* butterfly hybrid zones are correlated with a local rainfall peak at the edge of the Amazon basin. *Evolution*, pages n/a–n/a.

- [Rosser et al., 2012] Rosser, N., Phillimore, A. B., Huertas, B., Willmott, K. R., and Mallet, J. (2012). Testing historical explanations for gradients in species richness in heliconiine butterflies of tropical America. *Biological journal of the linnean society*, 105(3):479–497.
- [Rueffler et al., 2006] Rueffler, C., Van Dooren, T. J., Leimar, O., and Abrams, P. A. (2006). Disruptive selection and then what? *Trends in Ecology & Evolution*, 21(5):238–245.
- [Rundle et al., 2003] Rundle, H. D., Vamossi, S. M., and Schluter, D. (2003). Experimental test of predation's effect on divergent selection during character displacement in sticklebacks. *Proceedings of the National Academy of Sciences*, 100(25):14943–14948.
- [Rutowski, 1977a] Rutowski, R. L. (1977a). Chemical communication in the courtship of the small sulphur butterfly *Eurema lisa* (lepidoptera, pieridae). *Journal of Comparative Physiology ? A*, 115(1):75–85.
- [Rutowski, 1977b] Rutowski, R. L. (1977b). The use of visual cues in sexual and species discrimination by males of the small sulphur butterfly *Eurema lisa* (lepidoptera, pieridae). *Journal of comparative physiology*, 115(1):61–74.
- [Rutowski, 1991] Rutowski, R. L. (1991). The evolution of male mate-locating behavior in butterflies. *American Naturalist*, pages 1121–1139.
- [Rutowski and Schaefer, 1984] Rutowski, R. L. and Schaefer, J. (1984). Courtship behavior of the gulf fritillary *agraulis - vanillae* nymphalidae. *Journal of The Lepidopterists' Society*, 38:23–31.
- [Römpler et al., 2006] Römpler, H., Rohland, N., Lalueza-Fox, C., Willerslev, E., Kuznetsova, T., Rabeder, G., Bertranpetit, J., Schöneberg, T., and Hofreiter, M. (2006). Nuclear gene indicates coat-color polymorphism in mammoths. *Science*, 313(5783):62–62.
- [Saint-Laurent et al., 2003] Saint-Laurent, R., Legault, M., and Bernatchez, L. (2003). Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax* Mitchell). *Molecular Ecology*, 12(2):315–330.
- [Salzburger et al., 2002] Salzburger, W., Baric, S., and Sturmbauer, C. (2002). Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology*, 11(3):619–625.
- [Schemske, 2000] Schemske, D. W. (2000). Understanding the origin of species. *Evolution*, 54(3):1069–1073.
- [Schliewen et al., 1994] Schliewen, U. K., Tautz, D., and Pääbo, S. (1994). Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature*, 368(6472):629–632.
- [Schluter, 1994] Schluter, D. (1994). Experimental evidence that competition promotes divergence in adaptive radiation. *Science*, 266(5186):798–801.
- [Schluter, 2000] Schluter, D. (2000). *The Ecology of Adaptive Radiation*. Oxford Series in Ecology and Evolution.
- [Schluter, 2001] Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology & Evolution*, 16(7):372–380.
- [Schulz et al., 1993] Schulz, S., Boppre, M., and Vane-Wright, R. (1993). Specific mixtures of secretions from male scent organs of African milkweed butterflies (Danainae). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 342(1300):161–181.
- [Schulz et al., 2008] Schulz, S., Estrada, C., Yildizhan, S., Boppré, M., and Gilbert, L. E. (2008). An antiaphrodisiac in *Heliconius melpomene* butterflies. *Journal of Chemical Ecology*, 34(1):82–93.
- [Schwenk et al., 2008] Schwenk, K., Brede, N., and Streit, B. (2008). Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1505):2805–2811.
- [Schön et al., 2004] Schön, C. C., Utz, H. F., Groh, S., Truberg, B., Openshaw, S., and Melchinger, A. E. (2004). Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics*, 167(1):485–498.
- [Scott, 1972] Scott, J. (1972). Mating of butterflies. *Journal of Research on the Lepidoptera*, 11(2):99–127.
- [Seehausen, 2004] Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, 19(4):198–207.

- [Seehausen et al., 2014] Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, A., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., Lindholm, A. K., Lucek, K., Maan, M. E., Marques, D. A., Martin, S. H., Matthews, B., Meier, J. I., Möst, M., Nachman, M. W., Nonaka, E., Rennison, D. J., Schwarzer, J., Watson, E. T., Westram, A. M., and Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, 15(3):176–192.
- [Seehausen et al., 1997] Seehausen, O., Van Alphen, J. J., and Witte, F. (1997). Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science*, 277(5333):1808–1811.
- [Servedio and Noor, 2003] Servedio, M. R. and Noor, M. A. (2003). The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution, and Systematics*, pages 339–364.
- [Servedio et al., 2011] Servedio, M. R., Van Doorn, G. S., Kopp, M., Frame, A. M., and Nosil, P. (2011). Magic traits in speciation: ‘magic’ but not rare? *Trends in Ecology & Evolution*, 26(8):389–397.
- [Shapiro et al., 2004] Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jónsson, B., Schluter, D., and Kingsley, D. M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, 428(6984):717–723.
- [Shapiro et al., 2009] Shapiro, M. D., Summers, B. R., Balabhadra, S., Aldenhoven, J. T., Miller, A. L., Cunningham, C. B., Bell, M. A., and Kingsley, D. M. (2009). The genetic architecture of skeletal convergence and sex determination in ninespine sticklebacks. *Current Biology*, 19(13):1140–1145.
- [Shaw, 2002] Shaw, K. L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences*, 99(25):16122–16127.
- [Sheppard et al., 1985] Sheppard, P. M., Turner, J. R. G., Brown, K. S., Benson, W. W., and Singer, M. C. (1985). Genetics and the Evolution of Mullerian Mimicry in *Heliconius* Butterflies. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 308(1137):433–610.
- [Shikano et al., 2013] Shikano, T., Laine, V. N., Herczeg, G., Vilkki, J., and Merilä, J. (2013). Genetic Architecture of Parallel Pelvic Reduction in Ninespine Sticklebacks. *G3: Genes|Genomes|Genetics*, 3(10):1833–1842.
- [Shine et al., 2002] Shine, R., Reed, R. N., Shetty, S., Lemaster, M., and Mason, R. T. (2002). Reproductive isolating mechanisms between two sympatric sibling species of sea snakes. *Evolution*, 56(8):1655–1662.
- [Silberglied and Taylor, 1973] Silberglied, R. and Taylor, O. (1973). Ultraviolet differences between the sulphur butterflies, *Colias eurytheme* and *C. philodice*, and a possible isolating mechanism. *Nature*, 241:406–408.
- [Silberglied, 1979] Silberglied, R. E. (1979). Communication in the Ultraviolet. *Annual Review of Ecology and Systematics*, 10(1):373–398.
- [Silberglied, 1984] Silberglied, R. E. (1984). Visual communication and sexual selection among butterflies.
- [Silberglied and Taylor, 1978] Silberglied, R. E. and Taylor, O. R. (1978). Ultraviolet reflection and its behavioral role in the courtship of the sulfur butterflies *Colias eurytheme* and *C. philodice* (Lepidoptera, Pieridae). *Behavioral Ecology and Sociobiology*, 3(3):203–243.
- [Smiley, 1978] Smiley, J. (1978). Plant chemistry and the evolution of host specificity New evidence from *Heliconius* and *Passiflora*. *Science*, 201:745–747.
- [Smith and Kronforst, 2013] Smith, J. and Kronforst, M. R. (2013). Do *Heliconius* butterfly species exchange mimicry alleles? *Biology Letters*, 9(4):20130503.
- [Sobel et al., 2010] Sobel, J. M., Chen, G. F., Watt, L. R., and Schemske, D. W. (2010). The biology of speciation. *Evolution*, 64(2):295–315.
- [Soltis and Soltis, 2000] Soltis, P. S. and Soltis, D. E. (2000). The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences*, 97(13):7051–7057.
- [Song et al., 2011] Song, Y., Endepols, S., Klemann, N., Richter, D., Matuschka, F.-R., Shih, C.-H., Nachman, M. W., and Kohn, M. H. (2011). Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Current Biology*, 21(15):1296–1301.

- [Soria-Carrasco et al., 2014] Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., Buerkle, C. A., Feder, J. L., Bast, J., Schwander, T., Egan, S. P., Crespi, B. J., and Nosil, P. (2014). Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation. *Science*, 344(6185):738–742.
- [Spolsky et al., 1992] Spolsky, C., Phillips, C. A., and Uzzell, T. (1992). Gynogenetic reproduction in hybrid mole salamanders (genus *Ambystoma*). *Evolution*, pages 1935–1944.
- [Stebbins, 1959] Stebbins, G. L. (1959). The role of hybridization in evolution. *Proceedings of the American Philosophical Society*, pages 231–251.
- [Stein and Uy, 2006] Stein, A. C. and Uy, J. A. C. (2006). Unidirectional introgression of a sexually selected trait across an avian hybrid zone: a role for female choice? *Evolution; International Journal of Organic Evolution*, 60(7):1476–1485.
- [Stenson et al., 2002] Stenson, A. G., Malhotra, A., and Thorpe, R. S. (2002). Population differentiation and nuclear gene flow in the Dominican anole (*Anolis oculatus*). *Molecular Ecology*, 11(9):1679–1688.
- [Stern, 2013] Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics*, 14(11):751–764.
- [Stern and Orgogozo, 2009] Stern, D. L. and Orgogozo, V. (2009). Is Genetic Evolution Predictable? *Science (New York, N.Y.)*, 323(5915):746–751.
- [Stinchcombe and Hoekstra, 2007] Stinchcombe, J. R. and Hoekstra, H. E. (2007). Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, 100(2):158–170.
- [Strecker et al., 2012] Strecker, U., Hausdorf, B., and Wilkens, H. (2012). Parallel speciation in *Astyanax* cave fish (Teleostei) in Northern Mexico. *Molecular phylogenetics and evolution*, 62(1):62–70.
- [Stride, 1957] Stride, G. O. (1957). Investigations into the courtship behaviour of the male of *Hypolimnas misippus* L. (Lepidoptera, Nymphalidae), with special reference to the role of visual stimuli. *The British Journal of Animal Behaviour*, 5(4):153–167.
- [Stuckas et al., 2009] Stuckas, H., Stoof, K., Quesada, H., and Tiedemann, R. (2009). Evolutionary implications of discordant clines across the Baltic *Mytilus* hybrid zone (*Mytilus edulis* and *Mytilus trossulus*). *Heredity*, 103(2):146–156.
- [Sucena et al., 2003] Sucena, E., Delon, I., Jones, I., Payre, F., and Stern, D. L. (2003). Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature*, 424(6951):935–938.
- [Supple et al., 2013] Supple, M. A., Hines, H. M., Dasmahapatra, K. K., Lewis, J. J., Nielsen, D. M., Lavoie, C., Ray, D. A., Salazar, C., McMillan, W. O., and Counterman, B. A. (2013). Genomic architecture of adaptive color pattern divergence and convergence in *Heliconius* butterflies. *Genome research*, 23(8):1248–1257.
- [Svensson, 1996] Svensson, M. (1996). Sexual selection in moths: the role of chemical communication. *Biological Reviews*, 71(1):113–135.
- [Sweeney et al., 2003] Sweeney, A., Jiggins, C., and Johnsen, S. (2003). Insect communication: polarized light as a butterfly mating signal. *Nature*, 423(6935):31–32.
- [Sweigart and Willis, 2003] Sweigart, A. L. and Willis, J. H. (2003). Patterns of Nucleotide Diversity in Two Species of *Mimulus* Are Affected by Mating System and Asymmetric Introgression. *Evolution*, 57(11):2490–2506.
- [Symonds and Elgar, 2004] Symonds, M. R. and Elgar, M. A. (2004). The mode of pheromone evolution: evidence from bark beetles. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 271(1541):839–846.
- [Szymura and Barton, 1986] Szymura, J. M. and Barton, N. H. (1986). Genetic Analysis of a Hybrid Zone Between the Fire-Bellied Toads, *Bombina orientalis* and *B. variegata*, Near Cracow in Southern Poland. *Evolution*, 40(6):1141.
- [Szymura and Barton, 1991] Szymura, J. M. and Barton, N. H. (1991). The Genetic Structure of the Hybrid Zone between the Fire-Bellied Toads *Bombina orientalis* and *B. variegata*: Comparisons between Transects and between Loci. *Evolution*, 45(2):237.

- [Tajima, 1989] Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3):585–595.
- [Tanksley, 1993] Tanksley, S. D. (1993). Mapping polygenes. *Annual review of genetics*, 27(1):205–233.
- [Taylor et al., 2006] Taylor, E., Boughman, J., Groenenboom, M., Sniatynski, M., Schluter, D., and Gow, J. (2006). Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, 15(2):343–355.
- [Templeton, 2008] Templeton, A. R. (2008). The reality and importance of founder speciation in evolution. *Bioessays*, 30(5):470–479.
- [Theron et al., 2001] Theron, E., Hawkins, K., Bermingham, E., Ricklefs, R. E., and Mundy, N. I. (2001). The molecular basis of an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, *Coereba flaveola*. *Current Biology*, 11(8):550–557.
- [Thomas et al., 2003] Thomas, Y., Bethenod, M.-T., Pelozuelo, L., Frérot, B., and Bourguet, D. (2003). Genetic Isolation Between Two Sympatric Host-Plant Races of the European Corn Borer, *Ostrinia Nubilalis* Hubner I. Sex Pheromone, Moth Emergence Timing, and Parasitism. *Evolution*, 57(2):261–273.
- [Thompson and Jiggins, 2014] Thompson, M. J. and Jiggins, C. D. (2014). Supergenes and their role in evolution. *Heredity*, 113(1):1–8.
- [Tinbergen et al., 1942] Tinbergen, N., Meeuse, B. J. D., and Varossieau, L. K. B. u. W. W. (1942). Die Balz des Samtfalters, *Eumenis* (= *Satyrus*) *semele* (L.). *Zeitschrift für Tierpsychologie*, 5(1):182–226.
- [Tobler et al., 2004] Tobler, A., Kapan, D., Flanagan, N. S., Gonzalez, C., Peterson, E., Jiggins, C. D., Johnstson, J. S., Heckel, D. G., and McMillan, W. O. (2004). First-generation linkage map of the warningly colored butterfly *Heliconius erato*. *Heredity*, 94(4):408–417.
- [Turner, 1977] Turner, J. R. (1977). Butterfly mimicry: the genetical evolution of an adaptation. *Evolutionary biology*, 10.
- [Turner, 1985] Turner, J. R. (1985). Fisher's evolutionary faith and the challenge of mimicry. *Oxford Surveys in Evolutionary Biology*, 2:159–196.
- [Turner, 1971] Turner, J. R. G. (1971). Two Thousand Generations of Hybridisation in a *Heliconius* Butterfly. *Evolution*, 25(3):471–482.
- [Turner, 1972] Turner, J. R. G. (1972). The genetics of some polymorphic forms of the butterflies *Heliconius melpomene* (Linnaeus) and *Heliconius erato* (Linnaeus). II. The hybridization of subspecies of *H. melpomene* from Surinam and Trinidad. *Zoologica*, 56:125–157.
- [Turner and Crane, 1962] Turner, J. R. G. and Crane, J. (1962). The genetics of some polymorphic forms of the butterflies *Heliconius melpomene* Linnaeus and *Heliconius erato* Linnaeus. I. Major genes. *Zoologica, New York*, 47:141–152.
- [Turner and Sheppard, 1975] Turner, J. R. G. and Sheppard, P. M. (1975). Absence of crossing-over in female butterflies (*Heliconius*). *Heredity*, 34(2):265–269.
- [Turner et al., 2005] Turner, T. L., Hahn, M. W., and Nuzhdin, S. V. (2005). Genomic Islands of Speciation in *Anopheles gambiae*. *PLoS Biol*, 3(9):e285.
- [Twomey et al., 2013] Twomey, E., Yeager, J., Brown, J. L., Morales, V., Cummings, M., and Summers, K. (2013). Phenotypic and Genetic Divergence among Poison Frog Populations in a Mimetic Radiation. *PLoS ONE*, 8(2):e55443.
- [Van Oijen and Voorrips, 2001] Van Oijen, J. W. and Voorrips, R. E. (2001). Joinmap Version 3.0, software for the calculation of genetic linkage maps. *Plant Research International, Wageningen, The Netherlands*.
- [Vane-Wright and Boppre, 1993] Vane-Wright, R. and Boppre, M. (1993). Visual and Chemical Signalling in Butterflies: Functional and Phylogenetic Perspectives. *Phil. Trans. R. Soc. Lond. B*, 340:197–205.



- [Vasemägi and Primmer, 2005] Vasemägi, A. and Primmer, C. (2005). Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, 14(12):3623–3642.
- [Veen et al., 2001] Veen, T., Borge, T., Griffith, S. C., Saetre, G.-P., Bures, S., Gustafsson, L., and Sheldon, B. C. (2001). Hybridization and adaptive mate choice in flycatchers. *Nature*, 411(6833):45–50.
- [Via, 1999] Via, S. (1999). Reproductive Isolation between Sympatric Races of Pea Aphids. I. Gene Flow Restriction and Habitat Choice. *Evolution*, 53(5):1446–1457.
- [Via, 2001] Via, S. (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution*, 16(7):381–390.
- [Via, 2002] Via, S. (2002). The ecological genetics of speciation. *The American Naturalist*, 159(S3):S1–S7.
- [Via, 2009] Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(Suppl 1):9939–9946.
- [Via et al., 2000] Via, S., Bouck, A. C., and Skillman, S. (2000). Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution; International Journal of Organic Evolution*, 54(5):1626–1637.
- [Vignieri et al., 2010] Vignieri, S. N., Larson, J. G., and Hoekstra, H. E. (2010). The selective advantage of crypsis in mice. *Evolution*, 64(7):2153–2158.
- [Wagner and Altenberg, 1996] Wagner, G. P. and Altenberg, L. (1996). Perspective: complex adaptations and the evolution of evolvability. *Evolution*, pages 967–976.
- [Wago, 1978] Wago, H. (1978). Studies on the mating behavior of the pale grass blue, *Zizeeria maha argia* (Lepidoptera: Lycaenidae) III. Olfactory cues in sexual discrimination by males. *Applied entomology and zoology*, 13(4):283–289.
- [Weir and Cockerham, 1984] Weir, B. S. and Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6):1358.
- [Whinnett et al., 2005] Whinnett, A., Zimmermann, M., Willmott, K. R., Herrera, N., Mallarino, R., Simpson, F., Joron, M., Lamas, G., and Mallet, J. (2005). Strikingly variable divergence times inferred across an Amazonian butterfly ‘suture zone’. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1580):2525–2533.
- [Willing et al., 2012] Willing, E.-M., Dreyer, C., and van Oosterhout, C. (2012). Estimates of Genetic Differentiation Measured by  $F_{ST}$  Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. *PLoS ONE*, 7(8):e42649.
- [Wilson et al., 2000] Wilson, A. B., Noack-Kunmann, K., and Meyer, A. (2000). Incipient speciation in sympatric Nicaraguan crater lake cichlid fishes: sexual selection versus ecological diversification. *Proceedings of the Royal Society of London B: Biological Sciences*, 267(1458):2133–2141.
- [Wood and Keese, 1990] Wood, T. K. and Keese, M. C. (1990). Host-Plant-Induced Assortative Mating in *Enchenopa* Treehoppers. *Evolution*, 44(3):619–628.
- [Wright, 1932] Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proc of the 6th International Congress of Genetics*, volume 1, pages 356–366.
- [Wright, 1977] Wright, S. (1977). *Evolution and the Genetics of Populations, Volume 3*. Chicago, IL: University of Chicago Press.
- [Wu, 2001] Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14(6):851–865.
- [Wu et al., 2010] Wu, G. C., Joron, M., and Jiggins, C. D. (2010). Signatures of selection in loci governing major colour patterns in *Heliconius* butterflies and related species. *BMC evolutionary biology*, 10:368.
- [Wyatt, 2003] Wyatt, T. D. (2003). *Pheromones and animal behaviour: communication by smell and taste*. Cambridge University Press.
- [Yeaman and Whitlock, 2011] Yeaman, S. and Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, 65-7(7):1897–1911.

- 
- [Zhen et al., 2012] Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M., and Andolfatto, P. (2012). Parallel molecular evolution in an herbivore community. *Science*, 337(6102):1634–1637.
- [Ômura and Honda, 2011] Ômura, H. and Honda, K. (2011). Pungent odor of the adult skipper butterfly *Erynnis montanus* (Lepidoptera: Hesperidae). *Applied Entomology and Zoology*, 46(2):281–286.