MOLECULAR ECOLOGY

Molecular Ecology (2012) doi: 10.1111/j.1365-294X.2012.05730.x

Genome-wide patterns of divergence and gene flow across a butterfly radiation

NICOLA J. NADEAU,* SIMON H. MARTIN,* KRZYSZTOF M. KOZAK,* CAMILO SALAZAR,*†‡ KANCHON K. DASMAHAPATRA,§ JOHN W. DAVEY,¶ SIMON W. BAXTER,* ** MARK L. BLAXTER,¶ JAMES MALLET§†† and CHRIS D. JIGGINS*

*Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK, †Smithsonian Tropical Research Institute, Panama City, Naos Island, Causeway Amador, Panama, ‡Facultad de Ciencias Naturales y Matemáticas, Universidad del Rosario, Bogotá, Colombia, §Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, UK, ¶Ashworth Laboratories, Institute of Evolutionary Biology, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JT, UK, **School of Molecular & Biomedical Science, University of Adelaide, SA 5005, Australia, ††Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA, 02138, USA

Abstract

The *Heliconius* butterflies are a diverse recent radiation comprising multiple levels of divergence with ongoing gene flow between species. The recently sequenced genome of *Heliconius melpomene* allowed us to investigate the genomic evolution of this group using dense RAD marker sequencing. Phylogenetic analysis of 54 individuals robustly supported reciprocal monophyly of *H. melpomene* and *Heliconius cydno* and refuted previous phylogenetic hypotheses that *H. melpomene* may be paraphylectic with respect to *H. cydno*. *Heliconius timareta* also formed a monophyletic clade closely related but distinct from *H. cydno* with *Heliconius heurippa* falling within this clade. We find evidence for genetic admixture between sympatric populations of the sister clades *H. melpomene* and *H. cydno*/*timareta*, particularly between *H. cydno* and *H. melpomene* from Central America and between *H. timareta* and *H. melpomene* from the eastern slopes of the Andes. Between races, divergence is primarily explained by isolation by distance and there is no detectable genetic population structure between parapatric races, suggesting that hybrid zones between races are not zones of secondary contact. Our results also support previous findings that colour pattern loci are shared between populations and species with similar colour pattern elements. Furthermore, this pattern is almost unique to these genomic regions, with only a very small number of other loci showing significant similarity between populations and species with similar colour patterns.

Keywords: divergence, genome, *Heliconius*, phylogeny, speciation

Received 28 March 2012; revision received 28 May 2012; accepted 9 June 2012

Introduction

The *Heliconius* butterflies are a neotropical genus that has recently radiated into many species, subspecies and races. Contained within this phenotypic diversity are many examples of phenotypic convergence between both closely and distantly related species. These features combined with the recently completed genome sequence of *Heliconius melpomene* make the genus an excellent biological system in which to address questions about the genetics of speciation, adaptive divergence and phenotypic convergence. In particular, the sister species *H. melpomene* and *H. cydno* are sympatric for much of their range and are generally highly divergent in colour pattern, which is important for mate recognition (Jiggins 2008; Merrill *et al.* 2011a). Although both pre- and postzygotic barriers are present, hybrids are still occasionally found in the wild (Bull *et al.* 2006; Mallet *et al.* 2007; Merrill *et al.* 2011b).

Correspondence: Nicola J. Nadeau, Fax + 44 (0)1223 336676; E-mail njn27@cam.ac.uk

Despite evidence that hybrids between *H. melpomene* and *H. cydno* have reduced fitness (Jiggins 2008; Merrill *et al.* 2011b), there have been instances when selection has favoured the introgression of adaptive alleles between these species. For example, *H. heurippa* is a distinct species closely related to *H. cydno*, whose wing pattern has arisen from a hybridization event between a member of the *H. cydno* clade and *H. melpomene*, producing butterflies with colour patterns intermediate between the two parental species (Mavárez *et al.* 2006; Salazar *et al.* 2008). This hybridization event led to reproductive barriers with both parental species, in part due to a role for colour pattern in mate recognition (Melo *et al.* 2009). *H. timareta* is another species in the *H. cydno* species complex, previously thought to be limited to an isolated population in Ecuador, but shown to be present in multiple populations across the eastern slopes of the Andes from Peru to Colombia (Brower 1996a; Giraldo *et al.* 2008; Jiggins 2008; Mallet 2009). These populations have only recently been identified because many of them have the same colour pattern as local *H. melpomene* populations. Indeed, their patterns have been acquired through introgression of alleles from *H. melpomene* (Pardo-Diaz *et al.* 2012; The Heliconius Genome Consortium 2012).

There are 29 geographic races of *H. melpomene*, many of which have strikingly different aposematic colour patterns. There is evidence that these races are maintained by local frequency-dependent selection driven by predator avoidance of familiar warning colour patterns (Mallet & Barton 1989; Kapan 2001). The genetic control of colour pattern variation in *Heliconius* is well studied and the majority of the diversity is controlled by a small number of genetic loci (Sheppard *et al.* 1985; Jiggins *et al.* 2005). In particular, two clusters of loci control most colour pattern variation in *H. melpomene*, *HmB/D* on chromosome 18 (Baxter *et al.* 2008) and *HmYb/Sb/N* on chromosome 15 (Ferguson *et al.* 2010). These largely control red/orange and yellow/white colour pattern elements, respectively. These colour pattern loci appear to act as divergence islands, with very little genetic differentiation between races except at these genomic regions (Baxter *et al.* 2010; Nadeau *et al.* 2012). However, this has not been investigated at a genome-wide scale.

Several *H. melpomene* races with similar colour patterns show disjunct geographical distributions (e.g. *Heliconius melpomene rosina* and *Heliconius melpomene amaryllis*, Fig. 1). Phylogenetic analyses of several genetic markers unlinked to colour pattern loci have shown that races with similar patterns are not genetically similar. Genetic clustering is generally by geography, leading many to speculate that similar colour patterns have evolved independently (Brower 1996b; Flanagan *et al.* 2004; Quek *et al.* 2010). However, more recent analysis of the red colour pattern locus has shown that alleles at this locus cluster by colour pattern rather than geography (Hines *et al.* 2011). This supports the alternative hypothesis that races with similar colour patterns once had a continuous distribution that has been interrupted by the spread of more derived forms (Mallet 2010). Therefore, within the *H. melpomene/cydno* clade, there have been very few independent origins of novel colour patterns, and instead existing colour pattern alleles have been shared both within and between species. It remains to be seen if other regions of the genome show a signal of having been shared or retained between similar colour pattern forms.

Here, we address the evolutionary history of the *H. cydno/melpomene* radiation using dense RAD marker sequencing (Davey *et al.* 2011). In particular, we investigate the phylogenetic relationships of these species, which have previously been poorly resolved (Flanagan *et al.* 2004; Beltran *et al.* 2007; Quek *et al.* 2010). This data set has previously been used to create a preliminary phylogeny using neighbour-joining methods (The Heliconius Genome Consortium 2012) but here we seek to test this with more robust methodology. We also investigate the prevalence of gene flow between *H. melpomene* and members of the *H. cydno* species complex, the influences of geographic and taxonomic separation on genetic divergence, and address the question of how much of the genome is shared between butterflies that have similar colour patterns.

Methods

RAD genotyping and SNP calling

RAD genotyping of the samples shown in Fig. 1 (details in Table S1, Supporting information) has been described previously (The Heliconius Genome Consortium 2012). Briefly, DNA was digested with the PstI restriction enzyme to achieve a high density of markers before ligation of P1 sequencing adaptors with five-base molecular identifiers to allow multiplexing of three to six individuals per sequencing lane (Baird *et al.* 2008; Baxter *et al.* 2011). Paired-end 100 base sequencing was performed on the libraries using Illumina Genome Analyzer IIx and HiSeq2000 sequencers. The 270-Mb *Heliconius melpomene* genome sequence contains 27K PstI cut sites; therefore, on average, RAD markers are expected every 10 kb.

Reads from each individual were aligned to the *H. melpomene* genome scaffolds (The Heliconius Genome Consortium 2012) using Stampy v1.0.13 (Lunter & Goodson 2011) with default parameters, except expected substitution rate, which was set to 0.01. Geno-

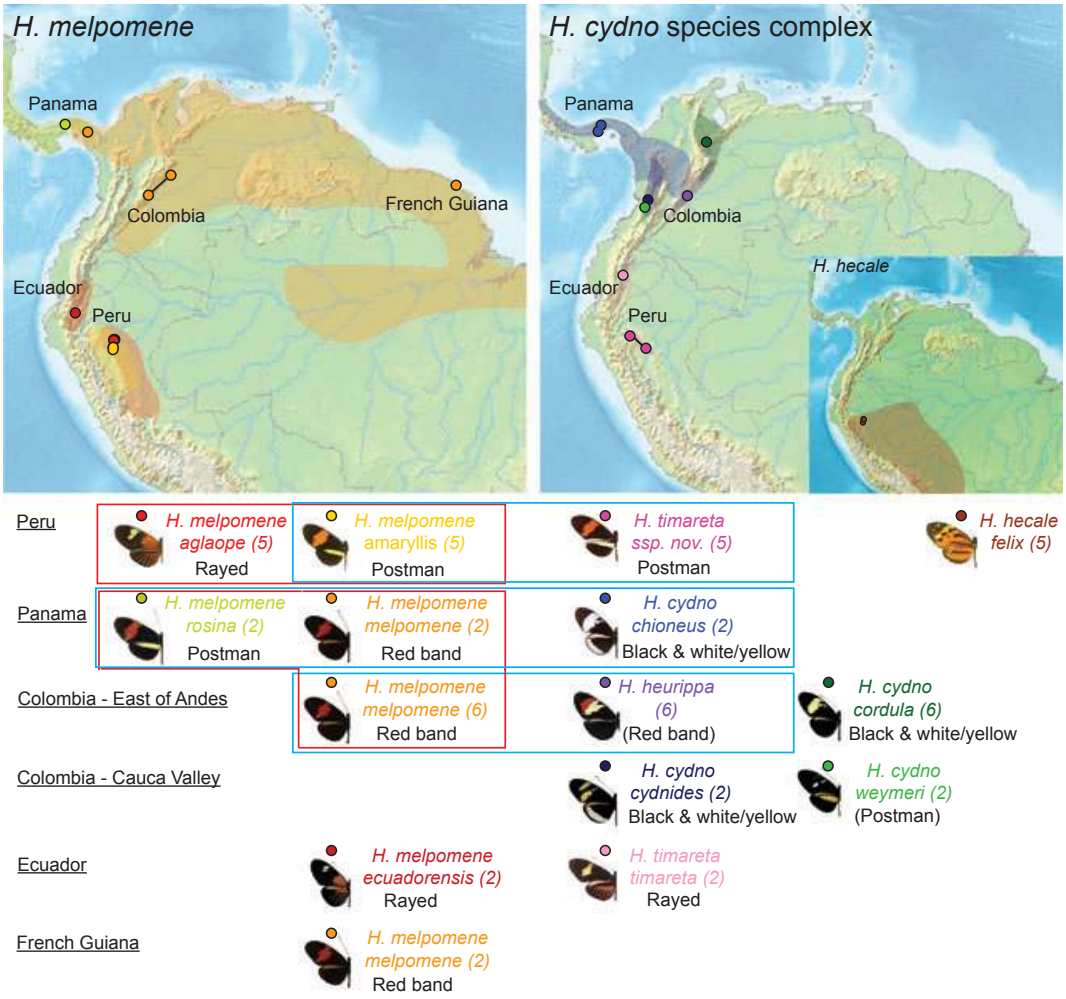

Fig. 1 Population sampling localities in South America. Pale shading shows distributions of the races/subspecies that we sampled (from Rosser *et al.* 2012). Note that distributions shown are for the subspecies sampled in this study only and complete species ranges are not shown. Boxes surround the most important population comparisons, red boxes indicate parapatric races and blue boxes indicate sympatric sister species. Samples sizes are given in parentheses after the species names. Colour pattern classifications used in the analyses are given below species names, those in brackets have only some elements of the given pattern type.

types were called using the Genome Analysis Tool Kit (GATK) v1.1 UnifiedGenotyper (DePristo *et al.* 2011). Genotype calls were only accepted if supported by a genotype quality (GQ) score of at least 30, which corresponds to an error rate of $\leq 1/1000$.

Phylogeny

We used the 480 123 nucleotide sites with confident calls (GQ $\geq$ 30) for all 54 RAD sequenced individuals to generate a phylogeny. Due to the difficulties of aligning sequence reads to divergent genomic regions, taxa that are more divergent from the reference tend to have more missing data (The Heliconius Genome Consortium 2012). Therefore, we chose not to use sites with missing data for some individuals as this might introduce biases and lead to problems with branch length estimation. A maximum likelihood approach was implemented in the program RAxML v. 7.2.8 (Stamatakis 2006). To ensure optimal search through the tree space, we empirically determined that 10 is sufficient as both the initial rearrangement setting and the number of gamma rate categories. We generated constraint trees (shown in Fig. S1, Supporting information) to represent phylogenetic hypotheses presented in previous publications and based on AFLP data (Quek *et al.* 2010), one mitochondrial and two nuclear loci (Beltran *et al.* 2007), separate analyses of the nuclear *tpi* and *mpi* loci (Flanagan *et al.* 2004) and an analysis of *Cytochrome oxidase I/II* data (K. Kozak, unpublished). The ML tree was estimated with and without each of the constraints, using random starting trees and the GTRGAMMA optimiza-

tion. Likelihoods of the results were compared statistically using the Shimodaira–Hasegawa test (Shimodaira & Hasegawa 1999) in RAxML. The result of the unconstrained analysis was determined to be significantly better than the alternative hypotheses at $P = 0.05$. Therefore, we repeated the unconstrained ML tree estimation with 1000 bootstrap replicates using the GTRCAT approximation (Stamatakis 2006) and multithreading onto 10 processors.

Structure analysis

We performed *Structure* analysis (Pritchard *et al.* 2000) using 53 837 nucleotide sites, consisting of the variable sites with minor allele frequency above 1% that have confident calls for all *H. melpomene, H. cydno, H. timareta* and *H. heurippa* individuals. An initial run with $K = 1$ was used to estimate the allele frequency distribution parameter, λ. We then ran short clustering runs (10^3 burn-in, 10^3 data collection) with the obtained value of λ (0.43), for $K = 2$–15, using an admixture model with no prior population information provided. This produced an optimum at $K = 4$. We then ran longer clustering runs (10^4 burn-in, 10^4 data collection) for $K = 2$–7. We also ran separate analyses of *H. melpomene*, *H. cydno* and *H. timareta* (including *H. heurippa*) with independent initial estimates of λ (0.58, 0.95 and 1.14, respectively) for $K = 2$–7 for *H. melpomene* and $K = 1$–4 for *H. timareta* and *H. cydno*.

Identification of RAD loci

Individual RAD loci were identified as regions containing 100 or more high-quality base calls across all individuals and that were separated from other RAD loci by at least 1 kb. When considering only *H. melpomene,* there were 6721 RAD loci with a mean span of 475 bases containing on average 212 called bases. The spacing between loci was 27 kb on average but this was highly variable (standard deviation = 30 kb). When incorporating individuals from the *H. cydno/timarata* clade, this dropped to 4078 RAD loci with a mean span of 420 bases containing on average 182 called bases with a spacing of 39 ± 54 kb. The number of RAD loci decreases as more individuals are added, as loci with fewer than 100 high-quality called bases across all individuals are ignored. However, some of the loci must also drop out due to differences in restriction enzyme cut sites. This variation was not included in the analysis.

F_{ST} analysis

F_{ST} was calculated for individual RAD loci using custom scripts in R (R Development Core Team 2011) with equations as described by Nadeau *et al.* (2012). To investigate genomic patterns of divergence at different levels and in different populations, we initially calculated F_{ST} between a subset of populations from which we had sequenced larger numbers of individuals. These included 5 sympatric species comparisons, *H. m. melpomene* ($n = 6$ individuals)/*H. cydno cordula* ($n = 6$) in Colombia/Venezuela, *H. m. melpomene* ($n = 6$)/*H. heurippa* ($n = 6$) in Colombia, *H. melpomene* ($n = 4$)/*H. cydno chioneus* ($n = 2$) in Panama, *Heliconius melpomene amaryllis* ($n = 5$)/*H. timareta spp nov.* ($n = 5$) and the more distantly related *H. m. amaryllis/H. hecale felix* ($n = 5$) in Peru; two parapatric race comparisons *H. m. amaryllis/aglaope* ($n = 5$) and *H. m. rosina* ($n = 2$)/*H. m. melpomene* ($n = 8$); and one allopatric race comparison between *H. m. amaryllis* and *H. m. melpomene* from Colombia. Pearson's product-moment correlations were calculated between arcsine-transformed RAD locus F_{ST} values in different population comparisons.

We also used these F_{ST} values to calculate the size of divergence hitch-hiking regions by measuring the correlation between these values for RAD loci at increasing genomic distances (following Smadja *et al.* 2012). Pearson's product-moment correlations were calculated for arcsine-transformed F_{ST} values for the pairs of RAD loci at increasing genomic distances. Distance categories were increased in 1-kb intervals up to 100 kb, then at 2-kb intervals up to 300 kb and then at 5-kb intervals up to 500 kb. A model of exponential decline in correlation coefficient with increasing genomic distance was fit to the data (with weightings by the number of data points in each category), summarized by the equation:

$$y = e^{-ax}$$

where y is the correlation coefficient, x is the genomic distance and a is a parameter estimated from the data, with larger values of a indicating smaller hitch-hiking regions.

To further investigate patterns of gene flow and divergence within and between species, two individuals were sampled from five further populations across the *H. melpomene, cydno* and *timareta* ranges: *H. cydno weymeri* and *cydnides* from the Cauca Valley in Colombia, *H. melpomene ecuadorensis* and *H. timareta timareta* from Ecuador and *H. melpomene melpomene* from French Guiana. As F_{ST} is strongly influenced by sample size (Waples 1998), we recalculated F_{ST} for all RAD loci between all possible pairs of populations using the two individuals with the best sequence coverage from each population. Thus, absolute values of F_{ST} may be overestimates, but previous work indicates that relative values calculated from small sample sizes across large numbers of sites are robust for comparisons between populations and gene regions (Nadeau *et al.* 2012). Mean

values of F_{ST} across all loci were used in Mantel tests, using the ecodist R package (Mantel 1967; Goslee & Urban 2007), to calculate the effect of taxonomic and geographic distance on divergence (values in Table S2, Supporting information). Taxonomic distances were scored as 1 between *H. melpomene* and *H. cydno/timareta/heurippa* and between *H. cydno* and *H. timareta/heurippa* and as 2 between *H. hecale* and *H. melpomene/cydno/timareta/heurippa*. Geographic distances were calculated as the straight-line distance between the GPS coordinates of the sampled populations using the Geographic Distance Matrix Generator (Ersts 2007).

AMOVA analysis

RAD loci were processed using custom scripts in R (R Development Core Team 2011) to generate genotype distance matrices, with the distance from a homozygous to heterozygous base call equal to half the distance between two different homozygous bases. These were used to perform AMOVA tests in the R package ade4 (Excoffier *et al.* 1992; Chessel *et al.* 2004), with significance estimates obtained by randomization tests with 99 permutations. AMOVAs were run on the combined *H. cydno/melpomene* data set with groupings by taxon (*H. melpomene* or *H. cydno/timareta/heurippa*), geographic location and colour pattern (both as listed in Fig. 1). They were also run on the data set containing only *H. melpomene*, with groupings by colour pattern and geography. In this instance, the geographic groupings were broadened to encompass more populations given the reduced number of populations being compared (Ecuador/Peru, Panama/Colombia and French Guiana, which correspond to the populations identified in the cluster analysis). A locus was said to have a better fit to one model than another if the difference between the fit of the models at that locus was greater than a threshold of two standard deviations above the mean fit of the model across all RAD loci.

Results

Genome-wide phylogeny and population structure of the Heliconius melpomene/cydno/timareta group

We generated a maximum likelihood phylogeny using the aligned RAD sequence data with five individuals of *H. hecale* as outgroups (Fig. 2). This phylogeny is consistent with the previously published neighbour-joining phylogeny created using the same data (The Heliconius Genome Consortium 2012). There are strongly supported monophyletic clades for each of the species, *Heliconius melpomene, Heliconius cydno, Heliconius timareta* and *Heliconius heurippa* (Fig. 2A). *H. heurippa* is most closely related to *H. timareta*. It should be noted that although we refer to the clade containing *H. timarata* and *H. heurippa* as *H. timareta,* reflecting our wider sampling of this species, in fact *H. heurippa* is the older name and if these species were to be synonymized *H. timareta* would become a subspecies of *H. heurippa*. There is strong support for a *H. cydno/timareta/heurippa* clade, distinct from *H. melpomene*. Relatively low bootstrap support values for most nodes within species suggest a high degree of genetic exchange within species. Conversely, the high support for interspecific splits indicates that levels of gene flow between species are insufficient to obscure the phylogenetic signal of speciation. Consistent with most previously published phylogenies, we find that *H. melpomene* falls into two major, 'eastern' and 'western' clades with respect to the Andes. Butterflies from Panama and Colombia fall in the western clade and butterflies from Amazonian Peru, Ecuador and French Guiana fall in the eastern clade (Flanagan *et al.* 2004; Beltran *et al.* 2007; Quek *et al.* 2010). The Shimodaira–Hasegawa test showed all the previously published phylogenetic hypotheses to be a worse fit to our extensive data set than the results of our unconstrained analysis ($P < 0.05$). In particular, our analysis contradicts the findings of Quek *et al.* (2010), showing no evidence of nesting of *H. cydno* within *H. melpomene*. In general, the quantity of sequence data accumulated in our data set allows for significantly more robust phylogenetic analysis than any of the previous studies based on much shorter sequences or AFLPs.

We also performed an analysis of population structure (Pritchard *et al.* 2000) on the whole group as well as each of *H. melpomene, H. cydno* and *H. timareta* (including *H. heurippa*) separately. For the whole group, the inferred optimum number of clusters was 5 (Fig. 2B), which largely corresponded to *H. cydno*, *H. timareta/heurippa*, eastern and western *H. melpomene* and a third *H. melpomene* cluster that contributed about 70% to the French Guiana population. The eastern *H. melpomene* cluster contributed the remaining 30% of the French Guiana population. Apart from this, the populations identified in the analysis were generally distinct with apparently only small amounts of admixture. The *H. melpomene* population from eastern Colombia, although primarily clustering with the western clade, also has genetic contributions from eastern and French Guiana clades. These results are largely consistent with those found previously using AFLPs (Quek *et al.* 2010). There is also evidence for admixture between species of up to 13% in some individuals. This is particularly strong between eastern clade *H. melpomene* and *H. timareta* but is also evident between western clade *H. melpomene* and *H. cydno*. Notably, when $K = 2$, the clusters correspond to *H. melpomene*, and *cydno/ti-*

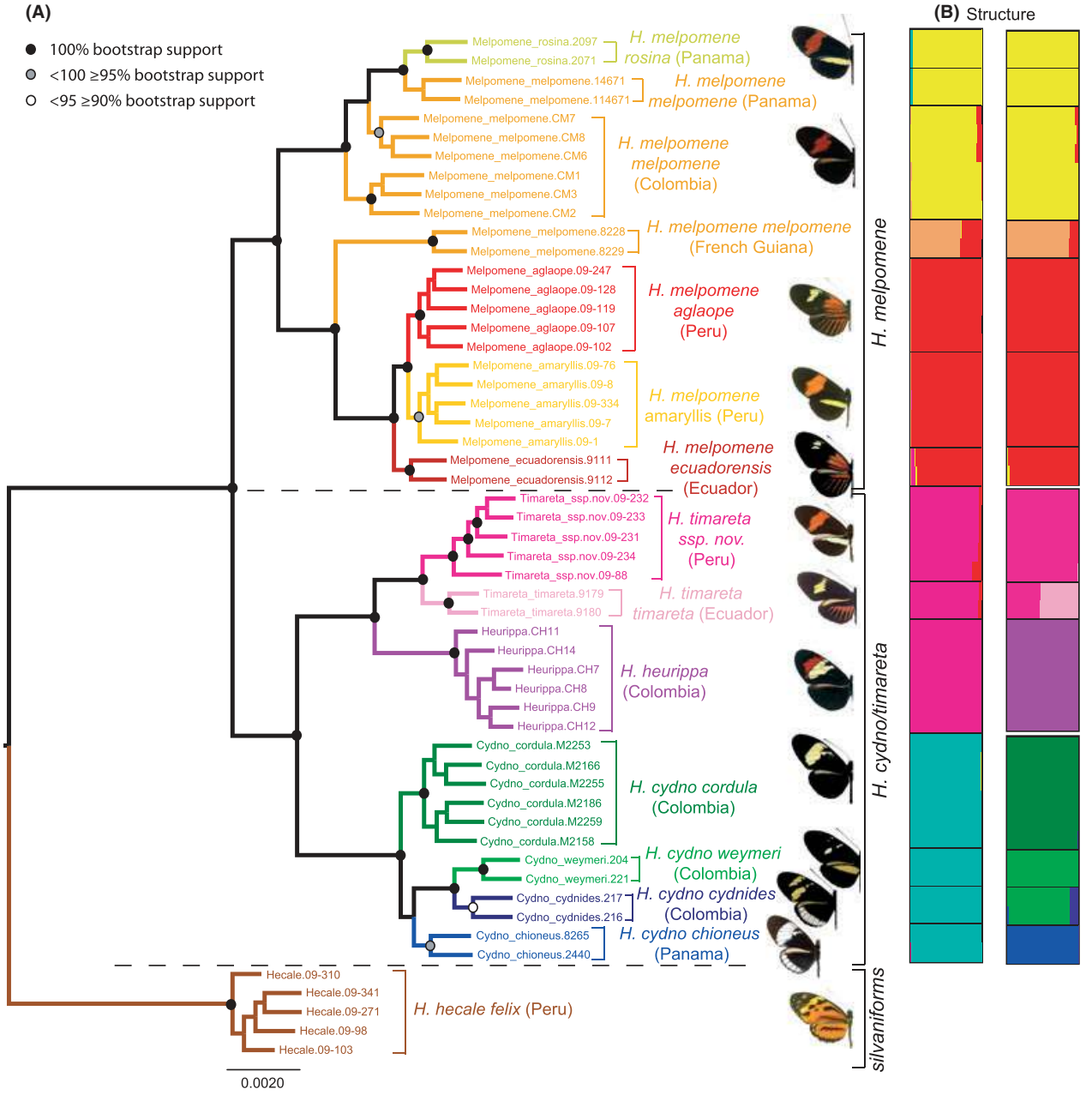

Fig. 2 (A) Maximum likelihood phylogeny with 1000 bootstrap replicates using the GTRCAT approximation generated from 480 123 nucleotide sites. (B) Structure analysis of the *Heliconius melpomene/cydno* clade (left) and for each species separately (right), showing the optimum number of population clusters in each case ($K = 5$ and $K = 10$ respectively). The populations, as in A, are separated with black lines. Each individual is represented by a single horizontal bar with colours indicating the genetic contribution from each of the identified clusters.

mareta and there is a contribution of almost 16% from the *H. cydno* cluster to Panamanian *H. melpomene*. This dropped to just 0.8% at $K = 3$ when the eastern and western melpomene clades are resolved (Fig. S2, Supporting information). This raises the intriguing possibility that differentiation of the *H. melpomene* eastern and western clades may in part be due to introgression from the sympatric species *H. cydno* (for the western clade) and *H. timareta* (for the eastern clade).

We also analysed each of the species separately. For *H. timareta/heurippa*, the optimum number of clusters increased to 3, largely resolving the subspecies and resolving *H. heurippa* as a distinct cluster within this group (Fig. 2B). For *H. cydno*, the optimum number of

clusters was 4, again largely resolving the races but with only a small amount of structure present between *H. c. cydnides* and *H. c. weymeri*. However, the presence of distinct clusters at this level could be due to isolation by distance only, with the clusters being due to the discrete sampling of continuous variation (Pritchard *et al.* 2000). Within *H. melpomene*, the optimum number of clusters stayed at 3. Even when *K* was increased to 7 for these populations, we found no evidence for genetic differentiation between parapatric *H. m. aglaope* and *amaryllis* in Peru or between *H. m. rosina* and *melpomene* in Panama (Fig. S2, Supporting information).

Are divergent genomic regions consistent between races and species?

It has been suggested that the same genomic regions might be involved at different stages of divergence, with the genomic extent of divergence increasing as speciation proceeds (Merrill *et al.* 2011a; Nosil & Feder 2012). To test this, we looked at whether the same regions are divergent at different points along the speciation continuum. F_{ST} values of loci from each of the sympatric *H. melpomene-H. cydno/timareta/heurippa* species pairs were highly significantly correlated (Fig. S3, Supporting information; Colombia, *H. heurippa/melpomene* vs. Peru, *H. timareta/H. melpomene* $r = 0.48$, $P < 2.2e{-}16$; Panama, *H. cydno/melpomene* vs. Peru, $r = 0.382$ $P < 2.2e{-}16$; Colombia vs. Panama $r = 0.508$, $P < 2.2e{-}16$). In contrast, there was no correlation between the two *H. melpomene* parapatric race comparisons ($r = 0.023$, $P = 0.195$). This suggests that correlations in divergence are not simply due to differences in mutation rate or recombination rate across the genome, both of which might lead to correlated divergence in different species pairs. Within geographic regions, however, there was some evidence for a speciation continuum from races to species level divergence (Mallet *et al.* 2007; Merrill *et al.* 2011a). There were weak but significant correlations in RAD locus F_{ST} values between races and species from the same geographic regions (Peru, $r = 0.077$, $P = 8.2e{-}6$; Colombia, $r = 0.072$, $P = 3.4e{-}05$).

Testing for islands of divergence

We used a model of exponential decline in correlation coefficient with increasing genomic distance to look for evidence of islands of divergence on a genome-wide scale. Islands were smallest between parapatic populations of the same species, at 10–20 kb, and larger, at 80–180 kb, between allopatric populations and sympatric sister species (Table 1, Fig. S4, Supporting information). Between distantly related species, correlations again decreased more rapidly with distance, and estimated island size was ~40 kb. However, only three of the comparisons had a significant fit to the model (parapatric races *H. m. rosina/melpomene*, *H. melpomene/cydno* in Colombia and *H. melpomene/cydno* in Panama) and only *H. melpomene* and *H. c. chioneus* in Panama had a good fit with 12.4% of the total variation in correlation coefficient explained by the exponential decline with distance. Therefore, it is perhaps only between these populations, particularly *H. melpomene* and *cydno* in Panama, where divergence hitch-hiking is having a detectable effect on the genome.

Patterns of gene flow between and within species with geographic distance

In addition to the comparisons described previously, we also calculated pairwise F_{ST} using all possible pairs of sampled populations (as indicated in Fig. 1). In general,

Table 1 The extent of divergence hitch-hiking as measured by the decline in correlation between F_{ST} values with increasing genomic distance

	Exponent multiplication factor for decline in correlation coefficient with distance† (±SE)	Fit of the model	Approximate island size‡ (kb)
Parapatric races (*Heliconius melpomene aglaope/amaryllis*, Peru)	5.51e−04 ± 1.46e−04	−0.06	13
Parapatric races (*H. melpomene rosina/melpomene*)	3.55e−04 ± 6.86e−05	0.04*	20
Allopatric races (*H. melpomene amaryllis/melpomene*)	6.97e−05 ± 6.87e−06	−0.15	100
Sister species (*H. melpomene/heurippa*, Colombia)	8.54e−05 ± 9.17e−06	−0.13	81
Sister species (*H. melpomene/cydno*, Colombia)	6.67e−05 ± 6.08e−06	0.08*	104
Sister species (*H. melpomene/timareta*, Peru)	3.96e−05 ± 3.25e−06	−0.22	176
Sister species (*H. melpomene/cydno*, Panama)	4.85e−05 ± 3.89e−06	0.12*	144
Distant species (*H. melpomene/hecale*, Peru)	1.78e−04 ± 2.29e−05	−0.22	39

*significant at $P < 0.01$.

†The value a in the equation $y = e^{-ax}$.

‡Defined as the distance at which the model predicts that the correlation coefficient (r) drops below 0.001.

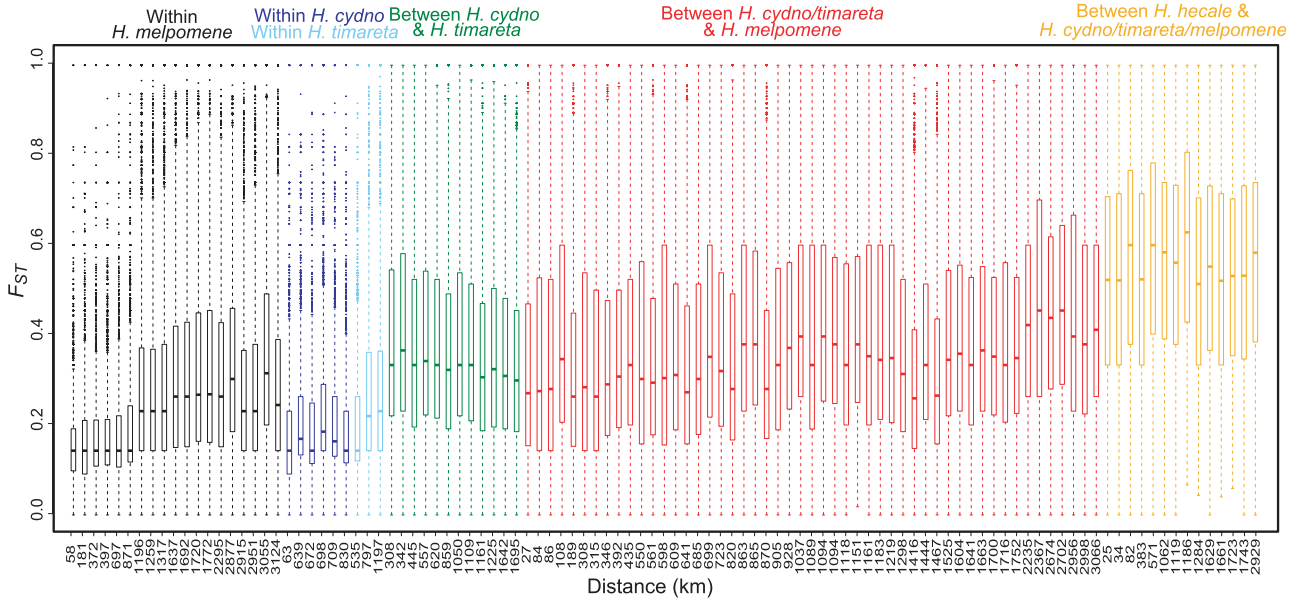

Fig. 3 Distributions of F_{ST} values across all RAD loci for all pairs of populations. Two individuals from each population are used, so F_{ST} values are likely to be inflated by up to 0.25 relative to true values.

populations showed increasing divergence with increasing geographic distance, as well as showing increases with taxonomic distance (Fig. 3). Within *H. melpomene*, there was a significant correlation between linear geographic distance and mean F_{ST} between populations (Mantel's $r = 0.793$, $P = 0.001$). A partial Mantel test on all populations of all species with distance and species dissimilarity as variables was also significant ($r = 0.651$, $P = 0.003$). This indicates increasing gene flow with geographic proximity both within species and between *H. melpomene* and *H. cydno/timareta*.

Genetic associations with geography vs. taxon

The above results suggest that there is gene flow between sympatric populations of *H. melpomene* and *H. cydno/timareta* and that between multiple distinct pairs of these species there may be particular regions of the genome that consistently act as 'islands' of divergence while others flow more freely. To test this further, we measured the strength of genetic associations with taxon (*H. melpomene* or *H. cydno/timareta*) and with geography (country) for each RAD locus using AMOVAS (Fig. 4A, Fig. S5, Supporting information). We found 1297 loci (32%) that were significantly associated with taxon (at $P = 0.01$) but not with geography ($P > 0.05$) and 231 (5.7%) that were significantly associated with geography but not taxon. The null expectation at these significance levels would be 0.95% (39 loci), showing that more of the genome follows these patterns than would be expected by chance. Of these 285 passed our stringent threshold for being more highly associated with taxon than geography (indicated with stars in Fig. 4A), suggesting that these regions are maintained between species. Fifty-three had stronger associations with geography than taxon, which would be consistent with these regions being independently shared between species in different geographic populations, most likely due to gene flow.

If the loci most strongly associated with taxon are those under divergent selection, these should also be loci with high F_{ST} in sympatric *H. melpomene–H. cydno/timareta* pairs. Of the 21 loci that were outliers in the F_{ST} distribution in all four *H. melpomene- H. cydno/timareta/heurippa* pairs, all were also found to be significantly associated with taxon in the AMOVA analysis and 18 were significantly more highly associated with taxon than geography (Fig. S5, Supporting information). Of the 285 loci identified in the AMOVA analysis, all had F_{ST} above 0.4 in at least one of the species pair comparisons (Fig. S3, Supporting information).

Associations with colour pattern

Regions controlling wing pattern are known to show a distinct intraspecific phylogeographic history as compared with neutral markers (Hines *et al.* 2011) and in some cases have introgressed across species boundaries (Salazar *et al.* 2010; The Heliconius Genome Consortium 2012; Pardo-Diaz *et al.* 2012). To investigate first whether these patterns are evident in our data, and second how much of the genome shares this signal, we

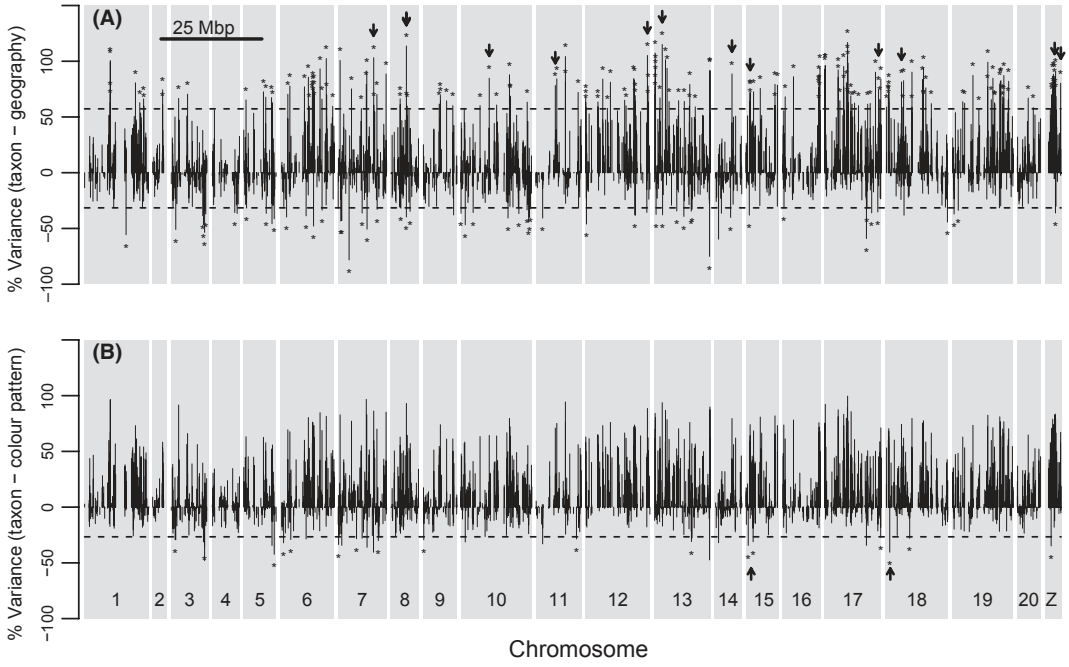

Fig. 4 RAD locus associations from AMOVA tests with: (A) taxon (*H. melpomene* or *H. cydno/timareta*) compared to geography and (B) taxon compared to colour pattern (as indicated in Fig. 1). Dashed lines indicate thresholds that were used to define loci as more highly associated with a particular factor. Starred values are those that pass the threshold and are significantly associated based on permutation tests (and that could also be assigned to mapped genome scaffolds). Arrows indicate the positions of the *HmYb* (left) and *HmB* (right) colour pattern loci.

tested for genetic associations with colour pattern type (as given in Fig. 1) for each of the RAD loci. Fourteen loci (0.3%) were found to be significantly more associated with colour pattern as compared to taxon or geography (Fig. 4B). One of these is within the *HmB* locus controlling red colour pattern (Nadeau *et al.* 2012). Two loci either side of the yellow colour pattern locus, *HmYb*, were also associated (one ~700 kb away and one ~480 kb away). The other 11 loci showing associations with colour pattern are therefore candidate regions for having been introgressed between species along with known wing patterning regions.

We also tested for genetic associations with colour pattern within only *H. melpomene*. Only one locus showed a significant association with colour pattern that also fit our criteria for showing a stronger association with colour than geography, which was in the *HmYb* region (Fig. S6A, Supporting information). The absence of a similar association at the red locus may be because the 'postman' and 'red band' phenotypes only differ in yellow colour pattern elements (controlled by *HmYb*). When these phenotypes were combined, giving only two categories of colour pattern (those with red hind-wing rays and those with a red forewing band), 15 loci were detected with stronger associations with colour pattern than geography (Fig. S6B, Supporting information). These included one at *HmB* as well as that previously detected at *HmYb*.

Discussion

The phylogenetic and phylogeographic history of the *Heliconius melpomene* clade has been the subject of considerable interest over the last twenty years, both for understanding the origins of wing pattern diversity and the speciation history of this group. Previous studies have used mtDNA (Brower 1994) and nuclear phylogenies (Flanagan *et al.* 2004; Beltran *et al.* 2007), or anonymous nuclear markers (Quek *et al.* 2010). Several of these have suggested paraphyly of *H. melpomene* with respect to *Heliconius cydno*, albeit with weak statistical support. Here, we have used an extensive genomic data set to provide greater resolution to this question and directly compared it with these previous phylogenies, demonstrating that *H. melpomene, H. cydno* and *Heliconius timareta/Heliconius heurippa* form three distinct monophyletic clades. The data therefore support the separation of *H. timareta* and *H. cydno* as distinct species.

H. heurippa is closely related to *H. timareta* but forms a distinct cluster in some analyses. The absence of a clear genetic contribution from *H. melpomene* to *H. heurippa*, for example in the Structure analysis, is consistent

with previous work showing that *H. heurippa* does not have a mosaic hybrid genome (Jiggins *et al.* 2008; Salazar *et al.* 2008). Instead, as demonstrated previously, introgression from *H. melpomene* is largely limited to colour pattern controlling loci, which have led to the establishment of this taxon as a partially reproductively isolated form (Salazar *et al.* 2010). Nonetheless, our data do call for a reassessment of the origins of *H. heurippa*. It now seems probably that this species arose from hybridization between an *H. melpomene* and an ancestral *H. timareta* population, rather than *H. cydno* as has been previously assumed. Further analyses of neighbouring *H. timareta* populations from Colombia will be necessary to determine the extent to which *H. heurippa* is reproductively isolated from these populations and the importance of the introgression of colour pattern alleles from *H. melpomene* in causing isolation. Combined with the recent description of *H. timareta* populations in Colombia (Giraldo *et al.* 2008) and Peru (J. Mavarez, unpublished), this suggests that *H. heurippa* lies at one end of a possibly discontinuous, and often cryptic, chain of *H. timareta* populations stretching along the Eastern slopes of the Andes from Peru to Colombia. We find no evidence for gene flow between *H. cydno* and the *H. timareta* clade. This is consistent with previous findings that *H. heurippa* does not readily interbreed with *H. cydno* (Mavárez *et al.* 2006) as these are the only populations of the two clades that abut geographically.

Genome-wide we find virtually no genetic structure between parapatric races of the same species, within either *H. melpomene* or *H. cydno*. This is contrary to another study of the *H. cydno cydnides/weymeri* hybrid zone using larger sample sizes but fewer loci, which found a signal of genetic structure between these races (C. F. Arias, personal communication). However, it is consistent with previous studies of *H. melpomene* hybrid zones using small numbers of genetic markers, which have suggested that races differ only at colour pattern controlling loci (Baxter *et al.* 2010; Nadeau *et al.* 2012). Similar phenotypic clines controlled by single loci have been found in other systems, for example leaf shape in morning glory (Campitelli & Stinchcombe This Issue). Our results suggest that hybrid zones between *H. melpomene* races are not secondary contact zones between populations that were once allopatric (Brower 1996b) and supports the hypothesis that locally selected colour patterns evolve almost independently of the rest of the genome (Mallet 2010; Hines *et al.* 2011). This is also supported by our finding that only colour pattern loci are shared between races with the same colour pattern. Other regions of the genome have not been retained between races that share colour patterns. We do find some other genomic regions that distinguish races with rayed patterns, which suggest some population structure perhaps due to a more recent expansion of the rayed forms (Mallet 2010; Quek *et al.* 2010; Hines *et al.* 2011). The situation in *Heliconius* is therefore quite similar to that seen in stickleback populations adapted to freshwater, where apparently independent origins of the same phenotypes have arisen from fixation of the same alleles (Colosimo *et al.* 2005; P. A. Hohenlohe *et al.* 2010). This has profound implications for comparative studies of phenotypic evolution and suggests that evolutionary biologists need to be cautious when inferring independent evolutionary origins for convergent phenotypes.

Consistent with species delimitations, we find particular genomic regions that are consistently divergent between *H. melpomene/cydno* species pairs. When comparing different sympatric species pairs of *H. melpomene* and *H. cydno/timareta*, there are strong correlations between F_{ST} values across genomic regions. These correlations may be partially due to shared ancestry leading to common patterns of divergence and ancestral polymorphism across the genome. However, when analysing genetic variation across multiple populations, we find particular regions that are highly associated with geographic but not taxonomic differences. This indicates that the correlations are, at least partially, due to ongoing gene flow at particular genomic regions in multiple populations. Therefore, consistent genomic regions are flowing freely, while others differentiate *H. melpomene* from both *H. cydno* and *H. timareta* in different localities. This is most likely due to a combination of their shared evolutionary history and similar ecology. *H. cydno* and *H. timareta* are found primarily at higher altitudes, are more host generalist and are found in more closed canopy forest as compared to *H. melpomene* (Jiggins 2008; Giraldo *et al.* 2008). Nonetheless, there are differences, most notably in wing pattern and perhaps also in pheromonal courtship signals (Giraldo *et al.* 2008). Therefore, comparing regions that are consistently divergent across populations to those that are uniquely divergent in particular populations might provide a useful means to identify genomic regions that have recently evolved adaptive differences.

There is considerable interest in the idea that islands of the genome might diverge during speciation, with surrounding regions remaining more homogeneous due to gene flow (Wu 2001; Nosil *et al.* 2009; Gompert *et al.* 2012; Stolting *et al.* This Issue). The *H. melpomene/H. cydno* species group is a good system for investigating this phenomenon, as gene flow between these species is known to occur (Bull *et al.* 2006; Kronforst 2008), despite their divergence over 1.5 MYA and their ecological and morphological distinctness in sympatry. We find evidence that islands of divergence get larger with increasing levels of divergence or decreasing gene flow.

Hitch-hiking regions appear to extend 10–20 kb between parapatric races and 80–180 kb between sympatric sister species. These are smaller than hitch-hiking regions reported from whitefish (Rogers & Bernatchez 2007; Renaut *et al.* 2012) and stickleback (Hohenlohe *et al.* 2010, 2012) but larger than those reported in *Littorina* winkles (Wood *et al.* 2008). Differences between study species could be due to obvious differences in demography, life history and selection, such as the parthenogenetic reproduction of aphids, which might act to increase the size of genomic islands through increased linkage disequilibrium. However, there might also be differences due to the wide variety of methods by which hitch-hiking regions are measured. The method we use was developed by Smadja *et al.* (2012) and led to estimation of much smaller hitch-hiking regions in the pea aphid as compared to previous estimates (Via & West 2008). Our estimates of islands of divergence in *Heliconius* may be downwardly biased because they are based on a genome-wide average, rather than being focussed solely on divergent regions. However, Smadja *et al.* found no significant difference when considering only regions containing outlier loci. The advantages of this method are that it does not rely on detecting outlier or background levels of divergence or in determining whether linked outliers are due to hitch-hiking or independent divergent selection (Nosil & Feder 2012; Via 2012).

The power of high-throughput sequencing of RAD markers has allowed us to provide answers to many long-standing questions about the evolution of the *Heliconius* butterflies. We have demonstrated ongoing gene flow between species, which appears to be occurring independently in multiple populations. Together with the multiple levels of divergence that are present in this system, this clearly offers great potential for addressing questions about the process of divergence with gene flow. What is now needed is a clear theoretical framework of expectations for genome-wide patterns of divergence under different scenarios of selection and gene flow (Feder *et al.* 2012). The *Heliconius* butterflies would then be an ideal empirical system in which to test these hypotheses.

Acknowledgements

We thank the governments of Colombia, Peru and Panama for permission to collect and export butterflies and Mauricio Linares for collecting and providing us will samples of *H. cydno cordula*. This work was funded primarily by a Leverhulme Trust award to CDJ and a BBSRC grant to JM, CDJ and MB. Sequencing of the RAD libraries was performed at the GenePool, Edinburgh.

References

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Baxter SW, Papa R, Chamberlain N *et al.* (2008) Convergent evolution in the genetic basis of mullerian mimicry in *Heliconius* butterflies. *Genetics*, **180**, 1567–1577.

Baxter SW, Nadeau NJ, Maroja LS *et al.* (2010) Genomic hotspots for adaptation: the population genetics of mullerian mimicry in the *Heliconius* melpomene clade. *PLoS Genetics*, **6**, e1000794.

Baxter SW, Davey JW, Johnston JS *et al.* (2011) Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS ONE*, **6**, e19315.

Beltran M, Jiggins CD, Brower AV, Bermingham E, 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**, 221–239.

Brower AV (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius* erato inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 6491–6495.

Brower AV (1996a) 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**, 317–332.

Brower AV (1996b) Parallel race formation and the evolution of mimicry in *Heliconius* butterflies: a phylogenetic hypothesis from mitochondrial dna sequences. *Evolution*, **50**, 195–221.

Bull V, Beltran M, Jiggins CD, McMillan WO, Bermingham E, Mallet J (2006) Polyphyly and gene flow between non-sibling *Heliconius* species. *Bmc Biology*, **4**, 11. doi:10.1186/1741-7007-4-11.

Campitelli BE, Stinchcombe JR (2012) Natural selection maintains a single-locus leaf shape cline in Ivyleaf morning glory, *Ipomoea hederacea*. *Molecular Ecology*, this issue.

Chessel D, Dufour AB, Thioulouse J (2004) The Ade4 package-I- one-table methods. *R News*, **4**, 5–10.

Colosimo PF, Hosemann KE, Balabhadra S *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928–1933.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.

Ersts PJ (2007) Geographic distance matrix generator. American Museum of Natural History, Center for Biodiversity and Conservation. Available from: http://biodiversityinformatics.amnh.org/open_source/gdmg.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial dna restriction data. *Genetics*, **131**, 479–491.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350. doi:10.1016/j.tig.2012.03.009. Available from: http://www.sciencedirect.com/science/article/pii/S0168952512000492.

Ferguson L, Lee SF, Chamberlain N *et al.* (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.

Flanagan NS, Tobler A, Davison A *et al.* (2004) Historical demography of müllerian mimicry in the neotropical *Heliconius* butterflies. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 9704–9709.

Giraldo N, Salazar C, Jiggins C, Bermingham E, Linares M (2008) Two sisters in the same dress: *Heliconius* cryptic species. *BMC Evolutionary Biology*, **8**, 324.

Gompert Z, Parchman TL, Buerkle CA (2012) Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 439–450.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1–19.

Hines HM, Counterman BA, Papa R *et al.* (2011) Wing patterning gene redefines the *Mimetic* history of *Heliconius* butterflies. *Proceedings of the National Academy of Sciences*, **108**, 19666–19671.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Plos Genetics*, **6**, e1000862.

Hohenlohe PA, Bassham S, Currey M, Cresko WA (2012) Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 395–408.

Jiggins CD (2008) Ecological speciation in *Mimetic* butterflies. *BioScience*, **58**, 541–548.

Jiggins CD, Mavarez J, Beltrán M, McMillan WO, Johnston JS, Bermingham E (2005) A genetic linkage map of the *Mimetic* butterfly *Heliconius* melpomene. *Genetics*, **171**, 557–570.

Jiggins CD, Salazar C, Linares M *et al.* (2008) Hybrid trait speciation and *Heliconius* butterflies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 3047–3054.

Kapan DD (2001) Three-butterfly system provides a field test of mullerian mimicry. *Nature*, **409**, 338–340.

Kronforst MR (2008) Gene flow persists millions of years after speciation in *Heliconius* butterflies. *Bmc Evolutionary Biology*, **8**, 98.

Lunter G, Goodson M (2011) Stampy: a statistical algorithm for sensitive and fast mapping of illumina sequence reads. *Genome Research*, **21**, 936–939.

Mallet J (2009) Rapid Speciation, Hybridization and Adaptive Radiation in the Heliconius Melpomene Group. In: Speciation and Patterns of Diversity (eds Butlin R, Bridle J, Schluter D), pp. 177–194. Cambridge University Press, Cambridge, UK.

Mallet J (2010) Shift happens! shifting balance and the evolution of diversity in warning colour and mimicry. *Ecological Entomology*, **35**, 90–104.

Mallet J, Barton NH (1989) Strong natural selection in a warning-color hybrid zone. *Evolution*, **43**, 421–431.

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

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

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

Melo MC, Salazar C, Jiggins CD, Linares M (2009) Assortative mating preferences among hybrids offers a route to hybrid speciation. *Evolution*, **63**, 1660–1665.

Merrill RM, Gompert Z, Dembeck LM, Kronforst MR, McMillan WO, Jiggins CD (2011a) Mate preference across the speciation continuum in a clade of *Mimetic* Butterflies. *Evolution*, **65**, 1489–1500.

Merrill RM, Van Schooten B, Scott JA, Jiggins CD (2011b) Pervasive Genetic Associations Between Traits Causing Reproductive Isolation in *Heliconius* Butterflies. *Proceedings of The Royal Society B: Biological Sciences*, **278**, 511–518.

Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic Islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 343–353.

Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 332–342.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Pardo-Diaz C, Salazar C, Baxter SW *et al.* (2012) Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics*, **8**, e1002752.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Quek S-P, Counterman BA, Albuquerque de Moura P *et al.* (2010) Dissecting comimetic radiations in *Heliconius* reveals divergent histories of convergent butterflies. *Proceedings of the National cademy of Sciences of the United States of America*, **107**, 7365–7370.

R Development Core Team (2011) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Available from: http://www.R-project.org/.

Renaut S, Maillet N, Normandeau E *et al.* (2012) Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 354–363.

Rogers SM, 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**, 1423–1438.

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

Salazar C, Jiggins CD, Taylor JE, Kronforst MR, Linares M (2008) Gene flow and the genealogical history of *Heliconius heurippa*. *BMC Evolutionary Biology*, **8**, 132.

Salazar C, Baxter SW, Pardo-Diaz C *et al.* (2010) Genetic evidence for hybrid trait speciation in *Heliconius* butterflies. *PLoS Genetics*, **6**, e1000930.

Sheppard PM, Turner JRG, Brown KS, Benson WW, Singer MC (1985) Genetics and the evolution of muellerian mimicry in *Heliconius* butterflies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **308**, 433–610.

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114–1116.

Smadja CM, Canbäck B, Vitalis R *et al.* (2012) Large-scale candidate gene scan reveals the role of Chemoreceptor genes in host plant specialisation and speciation in the pea aphid. *Evolution*, doi:10.1111/j.1558-5646.2012.01612.x.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Stolting K, Nipper R, Lindtke D *et al.* (2012) Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology*, this issue.

The Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94–98.

Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 451–460.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.

Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**, 438–450.

Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*, **17**, 3123–3135.

Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

N.J.N. is interested in the genetics of adaptation, speciation and sexual selection. S.H.M. is a PhD student investigating the role of selection in shaping genomic diversity. K.M.K. is a PhD student interested in molecular phylogenetics and macroevolutionary patterns. C.S. is interested in using tools from population genetics and genomics to study the speciation process in recent adaptive radiations. K.K.D is interested in speciation, with an emphaisis on neotropical butterflies. J.W.D. is interested in applying high throughput sequencing technologies to the study of butterfly genome architecture. S. W. B. is interested in detecting molecular variation that underlies adaptive traits. M.L.B. uses genomic data to investigate the evolutionary biology of a wide range of non-vertebrate animal species. J.M. and C.D.J. have both been instrumental in establishing Heliconius as a system for studying speciation and divergence as well as the genetics of adaptation and mimicry.

Data accessibility

DNA sequence reads: Submitted to the European Nucleotide Archive, Accession ERP000991.

Sampling locations: Supporting Information.

Scripts and input files: DRYAD entry doi:10.5061/dryad.j7q8p.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Fig. S1 Constraint trees used in the phylogenetic analysis, based on alternative phylogenetic hypotheses.

Fig. S2 *Structure* results for different values of *K* (between 1 and 7) for all species together and for a *K* value of 7 for *H. melpomene* only.

Fig. S3 Comparisons of F_{ST} values calculated for all RAD tags in different population and species pairs.

Fig. S4 Strength of correlations between F_{ST} values at genomic separations between 1 and 500kb. Fitted lines indicate exponential declines with genomic distance.

Fig. S5 AMOVA associations with taxon (*H. cyndo/melpomene*) and geography.

Fig. S6 AMOVA associations with colour pattern as compared to geography for RAD tags from *H. melpomene*.

Table S1. Samples used for RAD sequencing.

Table S2 Mean F_{ST} and geographic distances for 2 individuals from all pairs of populations.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.