MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 5513–5526 doi: 10.1111/mec.13849

Parapatric genetic introgression and phenotypic assimilation: testing conditions for introgression between Hercules beetles (*Dynastes*, Dynastinae)

JEN-PAN HUANG[1]

Museum of Zoology, Department of Ecology and Evolutionary Biology, University of Michigan, 1109 Geddes Ave., Ann Arbor, MI 48109-1079, USA

Abstract

The prevalence and consequences of genetic introgression between species have been intensively debated. I used Hercules beetles as examples to test for conditions that may be associated with the occurrence of introgression. RADseq data were used to reconstruct the species tree and history of introgression between Hercules beetles. Image data from museum specimens were used to investigate the phenotypic similarity of two adaptive traits between species from two distinct climatic realms (Nearctic vs. Neotropical). Genetic introgression was identified between Hercules beetles living in geographic proximity (parapatric). Phylogenetic relatedness and phenotypic similarity did not predict nor preclude genetic introgression between species. Phenotypic assimilation in body coloration was evident between distantly related Hercules beetles codistributed in Central America, where directional introgression was also statistically supported from the putative donor to receiver lineages. The number of introgressed loci was significantly higher between species with than without phenotypic similarity. I discuss the implications of recent studies on adaptive genetic introgression by providing supporting evidence from the Hercules beetle system.

Keywords: *Dynastes*, introgression, phenotypic assimilation, RADseq, species tree

Received 1 July 2016; revision received 30 August 2016; accepted 1 September 2016

Introduction

The prevalence of genetic introgression, the conditions that favour or limit its occurrence, and the consequences of genetic interchange between independently evolving lineages have been intensively debated in evolutionary biology (Harrison & Larson 2014; Payseur & Rieseberg 2016). For example, genetic interchange between animal species has been assumed as a disruptive force for speciation, because it homogenizes diverging evolutionary entities (introgression should thus be rare and its effect on systematics negligible (Mayr 1963)). Recent studies, however, have hypothesized that speciation cannot only proceed in the presence of gene flow, but that gene flow between diverging or diverged taxa can facilitate the formation of genomic islands, or linkage groups, of loci under divergent selection (Wu 2001; see review by Payseur & Rieseberg 2016). The formation of genomic islands may increase genomic incompatibility between species and help to finalize speciation (Seehausen 2013). Identifying the patterns of genetic introgression from shared ancestral polymorphisms using molecular data is unfortunately statistically challenging, and the credibility of positive results that support introgression can be affected by the study design (e.g. the number of sampled loci, the number of sampled individuals per species, the phylogenetic range of taxon sampling and the proposed phylogenetic hypothesis used in the analyses; Cruickshank & Hahn 2014; Eaton *et al*. 2015). That is, claims of genetic introgression between species may still be questionable even when a model for postdivergence genetic interchange is supported by the genetic data.

Correspondence: Jen-Pan Huang,
E-mail: huangjp@umich.edu
[1]Present Address: Integrative Research Center, The Field Museum, 1400 South Lake Shore Drive, Chicago, IL 60605, USA

The study of genetic introgression between animal species has become even more controversial, because recent studies argue for the prevalence of adaptive introgression (Brower 2013; Hedrick 2013; Seehausen 2013; Rheindt *et al*. 2014; Ru *et al*. 2016; Suarez-Gonzalez *et al*. 2016). Introgressed alleles between species can produce novel phenotypes that may facilitate adaptation under novel environmental conditions (e.g. Ru *et al*. 2016). For example, phenotypic convergence in wing patterns between mimetic sympatric *Heliconius* butterflies is due to the introgression of adaptive alleles between species (Pardo-Diaz *et al*. 2012; but see Brower 2013). However, without a well-annotated genomic resource (e.g. Suarez-Gonzalez *et al*. 2016), it is difficult to distinguish patterns of adaptive introgression from patterns of adaptation by new mutations or standing variation (Hedrick 2013; Harrison & Larson 2014; Payseur & Rieseberg 2016). Furthermore, the frequency with which introgression and adaptive phenotypic evolution occur simultaneously is also not clear, that is Is genetic introgression from other codistributed species always a potential resource for adaptive evolution?

The Hercules beetles (genus *Dynastes*; Dynastinae) are composed of 15 species, with distinct adaptive phenotypes between species (Huang & Knowles 2016; Huang 2016; Figs 1 and 2). The beetles are grouped into two major groups of species (Huang & Knowles 2016) distributed in two climatically distinct realms. The White Hercules beetles are North American species, three of which are found in the Nearctic region and two in the Central American Neotropical region (Fig. 1). The Giant Hercules beetles live in the Neotropical region, where nine species are found in South America and one in Central America. Phenotypic assimilation of putative adaptive traits has been proposed between a pair of distantly related but parapatrically distributed White and Giant Hercules species. *Dynastes maya* has been hypothesized as an intermediate form between the Nearctic White Hercules beetles and the Neotropical Giant Hercules beetles (Hardy 2003), where its adaptive body coloration may be due to introgression from a sympatric Giant Hercules species, *Dynastes septentrionalis* (i.e. phenotypic assimilation via introgression). I use the term phenotypic assimilation (e.g. Rheindt *et al*. 2014) instead of phenotypic convergence for patterns of evolving

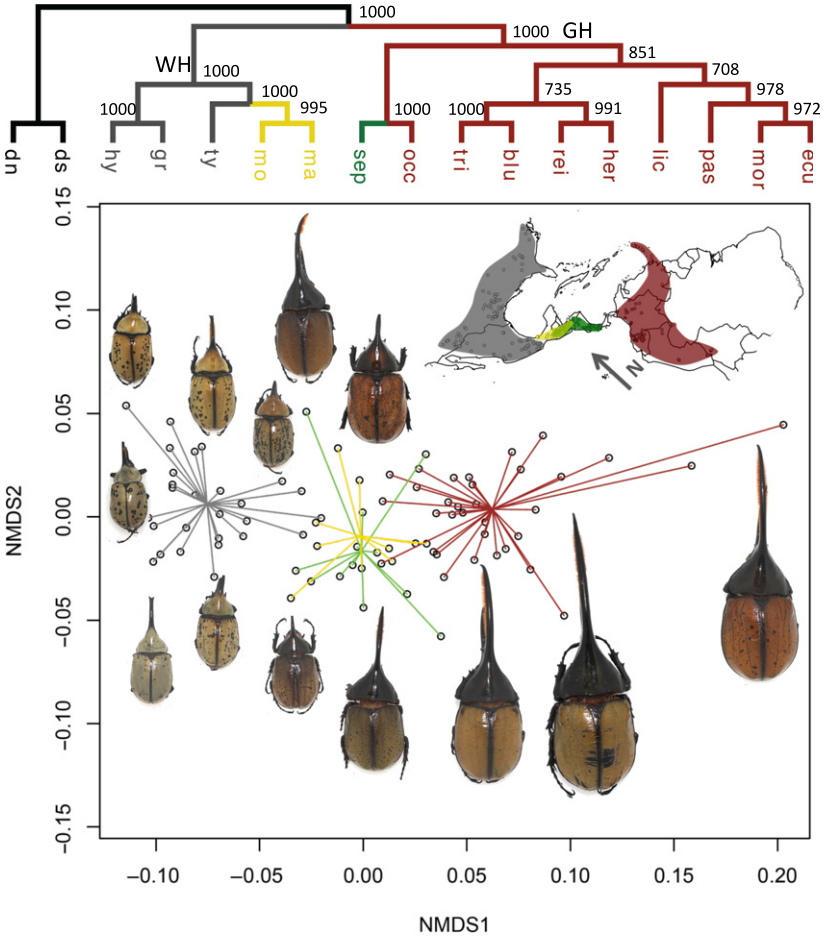

Fig. 1 A *Dynastes* species tree and an NMDS plot based on colour variation among Hercules beetles. Different colours correspond to different geographic groups (grey, Nearctic White Hercules; yellow, Central American Neotropical White Hercules; green, Central American Neotropical Giant Hercules; and brown, Neotropical Giant Hercules). Numbers next to the nodes of the tree are bootstrapping values based on 1000 replicates. See Table 1 for taxon names and abbreviations. Beetle images are not scaled. WH, White Hercules; GH, Giant Hercules. Abbreviations and their corresponding taxonomic names can be found in Table 1. [Colour figure can be viewed at wileyonlinelibrary.com].

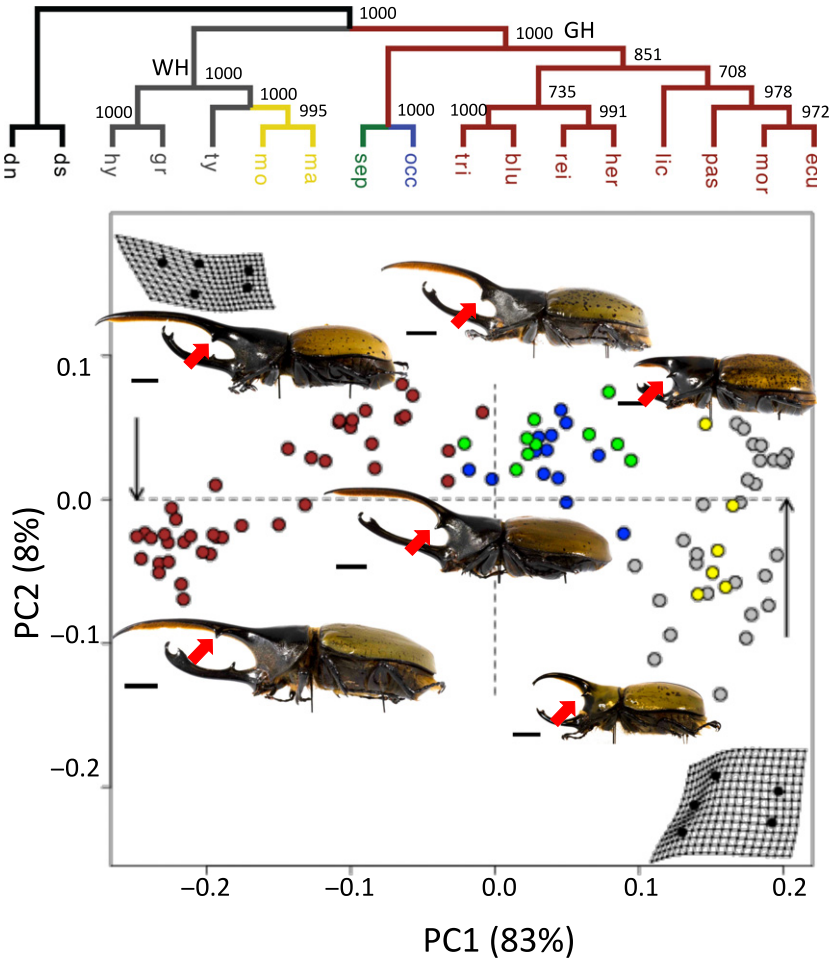

Fig. 2 A *Dynastes* species tree and a PCA plot based on variation in the position of the thoracic denticle (red arrow) across Hercules beetles. Different colours correspond to the different morphological groups (grey, NWS; yellow, CWS; green, CGS; blue, NGS; and brown, NGL [see materials and methods for grouping details]). Numbers next to the nodes of the tree are bootstrapping values based on 1000 replicates. See Table 1 for the taxon names and abbreviations. Scale bar = 1 cm. WH, White Hercules; GH, Giant Hercules. Abbreviations and their corresponding taxonomic names can be found in Table 1. [Colour figure can be viewed at wileyonlinelibrary.com].

similar phenotypes between species in this study because the term convergence implies independent origins. Different Hercules beetle species can produce viable and fertile hybrids (Lai & Ko 2008; Lai 2014), but hybrid swarms are uncommon in natural habitats (Huang 2012). Giant and White Hercules beetles can hybridize under artificial conditions (Lai & Ko 2008), but whether hybridization occurs under natural conditions is unclear. The Hercules beetle system differs from other study systems where the focus is speciation with gene flow. Sister species of Hercules beetles are allopatrically distributed (Huang & Knowles 2016), and gene flow between sister species can be rare without direct contact. Introgressive hybridization, on the other hand, might occur between nonsister species after secondary contact. The Hercules beetles are good subjects for testing the prevalence of introgression between a manageable number of species pairs and for identifying the conditions (i.e. geographic distribution, phylogenetic relationship and phenotypic similarity) that may predict or preclude the occurrence of introgression.

I used restriction-site-associated DNA sequencing (RADseq) to obtain sequence data from thousands of genomic regions to propose a well-resolved and highly supported phylogenetic hypothesis for Hercules beetles. *D*-statistics, which can accurately infer historical introgression from randomly sampled genomic data (Eaton & Ree 2013; Martin *et al.* 2013; Pease & Hahn 2015), were used to test for the significance of introgression between species. Previous studies have identified two putative adaptive traits in Hercules beetles: body coloration and male horn shape (Hinton & Jarman 1973; Jarman & Hinton 1974). The similarity of these traits between groups of species from the Nearctic and Neotropical regions was further investigated using data extracted from digitized images of museum specimens. The main goal was first to determine whether phenotypic assimilation occurs in Central America between distantly related *D. maya* (ma) and *D. septentrionalis* (sep). Subsequently, I tested whether the similarity of the traits coincided with genetic introgression. Specifically, the body coloration of ma and its sister taxon *Dynastes moroni* (mo) differs from that of other White

Hercules beetles; this type of coloration is typical among Giant Hercules beetles of the Neotropics (Fig. 1). On the other hand, sep and its Giant Hercules sister taxon *Dynastes occidentalis* (occ) have male thoracic horn phenotypes that more closely resemble White Hercules beetles (the thoracic denticle is always at the very basal position of the horn; Fig. 2). Using >40 K RAD loci from multiple samples of all the currently recognized species of Hercules beetles, nonparametric multidimensional scaling based on variation in body coloration (represented using red, green, and blue [RGB] values) and a landmark-based geometric morphometric analysis of horn shape, I asked the following questions: (i) Is introgression limited to closely related taxa? (ii) Is introgression limited to phenotypically similar species that are presumably under similar selective pressures (Neotropical vs. Nearctic)? and (iii) Is introgression directional from the putative allele–donor lineages to the allele–receiver lineages in cases of phenotypic assimilation?

Material and methods

Digitizing and processing image data

Images of 82 male Hercules beetles from 12 species vouchered at the Museum of Zoology, University of Michigan, collected between 2008 and 2014 were digitized. Three of the Giant Hercules species lacked complete male specimens (Table 1). Only recent samples were included to avoid the effect of degraded pigments on body coloration. Images were taken using a Nikon D320 digital camera from 50 cm directly above the specimens, along with a standard grey and a series of standard colour plates. Digitized images (in jpg format; see Supporting information) were then imported into Photoshop (CC 2014) to adjust the white balance. After adjustment and standardization of the colour, these images were imported into ImageJ (Schneider *et al.* 2012), and the red, green and blue (RGB) values of selected regions on the dorsal side of the specimens were extracted. A pronotal region and two elytral regions were chosen to test for the differentiation of body coloration (Fig. S1, Supporting information). Note that the apex and base of the elytra sometimes exhibit different colours in Hercules beetles (e.g. olive and brown, respectively), so the RGB values from these two regions of the elytra were recorded separately.

Analysing body coloration data

The RGB values of the sampled images were square-root-transformed, and the Bray–Curtis distances between samples were analysed using nonmetric multidimensional scaling (NMDS). Colour variation in the pronotum was not continuous among samples (black in most Neotropical Giant Hercules beetles and light grey or creamy yellow in many White Hercules beetles), so I used NMDS instead of principle component analysis to prevent overemphasizing the difference in discrete pronotal colour. The statistical significance of colour differences between samples from the following groups

Table 1 Species and sample sizes used in the analyses of the final molecular data set. Groupings of species based on body coloration and horn shape for phenotypic data sets are shown in the two right columns

	Species	Abbreviations	Sample size	Colour	Horn
Outgroups	*Dynastes satanas*	ds	1	N.A.	N.A.*
	Dynastes neptunus	dn	2	N.A.	N.A.
White	*Dynastes granti*	gr	4	NAW	NWS
	Dynastes hyllus	hy	4	NAW	NWS
	Dynastes maya	ma	3	CAW	CWS
	Dynastes moroni	mo	4	CAW	CWS
	Dynastes tityus	ty	2	NAW	NWS
Giant	*Dynastes bleuzeni*	blu	2	N.A.	N.A.
	Dynastes ecuatorianus	ecu	5	ONG	NGL
	Dynastes hercules	her	4	ONG	NGL
	Dynastes lichyi	lic	6	ONG	NGL
	Dynastes morishimai	mor	1	N.A.	N.A.
	Dynastes occidentalis	occ	3	ONG	NGS
	Dynastes paschoali	pas	2	N.A.	N.A.
	Dynastes reidi	rei	2	ONG	N.A.
	Dynastes septentrionalis	sep	4	CAG	CGS
	Dynastes trinidadensis	tri	2	ONG	N.A.

*Specimens were not available or not included for that species.

with different body colours was tested using a multivariate analysis of variance (MANOVA) with 1000 permutations based on Euclidean distances between samples: North American White Hercules (NAW), Central American White Hercules (i.e. taxa ma and mo; CAW), Central American Giant Hercules (i.e. taxon sep; CAG) and other Neotropical Giant Hercules (ONG). Finally, Tukey's honest significant difference (HSD) test was used *post hoc* to test for significant colour differences between pairs of groups. The analyses used the R packages MASS (Venables & Ripley 2002) and vegan (Dixon 2003).

Geometric morphometric analyses

I used image data ($N = 98$) from Huang & Knowles (2016; dryad digital repository http://dx.doi.org/10.5061/dryad.8p6 m0) to investigate the variation in the shape of the thoracic horn among major male Hercules beetles (the relative position of a denticle along the thoracic horn). Only major males exhibit species specific horn shape; minor/satellite males have small horns and their horn shape is similar between species. Five landmarks were identified for each specimen (Fig. S2, Supporting information), and the landmark coordinates from all samples were saved in a single tps file (see Supporting information) using the digitize2D function of the R package geomorph (Adams & Otarola-Castillo 2013). A generalized Procrustes analysis was applied to remove the effect of location, size and rotation of the relative positions of the landmarks among specimens using the function gpagen. The residuals from the mean shape of the landmarks were used to identify shape differences among the taxa using principle component analysis (PCA) and the function plotTangentSpace. Sampled taxa were assigned to five groups: (i) North White Shorthorn (NWS, including *Dynastes granti* [gr], *Dynastes hyllus* [hy] and *Dynastes tityus* [ty]); (ii) Central White Shorthorn (CWS, ma and mo); (iii) Central Giant Shorthorn (CGS, sep); (iv) Neotropical Giant Shorthorn (NGS, occ, which is the sister taxon to sep and exhibits a similar horn phenotype); and (v) Neotropical Giant Longhorn (NGL, the remaining Giant Hercules taxa in the analysis). The statistical significance of the variation of the position of the thoracic denticle along the thoracic horn among horn-shape groups was tested using an analysis of variance (ANOVA) based on the Euclidean distances. Pairwise comparisons were then used to determine whether horn shape differed significantly between pairs of groups. For both ANOVA and pairwise analyses, 999 permutations were used to test for statistical significance.

Rad library preparation

Genomic DNA (see Huang & Knowles 2016) extracted from thoracic muscle tissues of 61 beetle samples (including three samples from two outgroup species; Table 1; sample_info.txt [dryad doi:10.5061/dryad.4c5f9]) was individually barcoded and processed into two reduced-complexity libraries based on restriction fragments (Gompert *et al*. 2012). The DNA was digested with EcoRI and MseI restriction enzymes, and the fragments were then ligated to unique barcode sequences and Illumina adaptors. After ligation, the products were pooled among samples and then amplified by PCR. The PCR products were cleaned using AMPure Beads (Beckman Coulter) and subsequently size selected (350–420 bp) using Pippin Prep (Sage Science). The libraries were sequenced in two lanes on an Illumina HiSeq2000 (San Diego, CA, USA) by the Centre for Applied Genomics at the Hospital for Sick Children (Toronto, ON, Canada) to generate 150-bp single-end reads. Ten Hercules beetle samples were discarded due to low sequencing coverage (<200 000 sequence reads), which reduced the number of samples to 51 (including three outgroup samples).

Processing rad-seq data

The sequence data were assembled *de novo* using the PYRAD pipeline (ver. 3.0.66; Eaton 2014). Base calls with a quality score <20 were converted into Ns, and any read with >2 Ns was discarded. Illumina adaptors and restriction sequences were removed during filtering. Note that I did not filter the sequence reads for possible contamination from bacterial symbionts because these symbionts were not found in thoracic muscle tissues of beetles. Filtered reads within a sample were clustered using a threshold of 90% (a 85% threshold produced a similar number of loci; data in the Supporting information). Error rate and heterozygosity were estimated from the loci clusters for each individual, and the averages were used to establish consensus sequences. Heterozygous sites were included in the consensus using standard IUPAC code. Clusters with a sequencing depth <5 were discarded, and clusters with more than two alleles, perhaps due to paralogous loci, were also discarded (assuming that all species of Hercules beetles were diploid). Consensus loci built within samples were subsequently clustered among samples using a similarity threshold of 90% and then aligned (a maximum of six indels allowed). Loci with heterozygous alleles shared across more than eight individuals (approximately shared between more than two species) were also discarded. Specifically, postdivergence gene flow may have occurred between sister taxa, but shared

heterozygotes across multiple (too many) species may represent fixed differences among clustered paralogs (Eaton & Ree 2013).

A customized R script was applied to visualize the filtering and clustering results (the .loci output file) from PYRAD (cout_n_varsitesfrom_locifile.r; available at https://github.com/airbugs/Dynastes_introgression). The data set contained a systematic increase in sequence variation towards the end of the alignments, after site 110 (Supplementary figures). Another R script (chop_loci_file.r) was therefore used to exclude downstream sites for all aligned clusters. The pairwise sequence divergence based on uncorrected p-distances between samples was calculated for each locus, and loci with a maximum pairwise sequence divergence >15% between taxa were excluded from the data set using another customized R script (EDchopped_locifile.r). The maximum pairwise sequence divergence between an outgroup sample and a Hercules beetle in a mitochondrial COI data set was 15% (data extracted from Huang & Knowles 2016). This level of sequence divergence was thus used as the threshold for removing suspicious clusters of paralogous loci with too many variable sites. Finally, I removed all invariable loci and loci that contained samples from <4 species from the data set. The summarized results before and after data filtering can be found in the supplementary files (Figs S5 and S6, Supporting information). All customized R codes were archived (https://github.com/airbugs/Dynastes_introgression).

Species tree construction

I used a newly developed coalescent-based method (Chifman & Kubatko 2014) to construct the *Dynastes* species tree directly from unlinked SNP data. A customized R script was used to convert the edited .loci file from the previous section to a file with a PHYLIP-like format that contained aligned unlinked SNP data from each locus (concate_SNPs.r; assuming all loci in the final data set were unlinked and randomly distributed in the genome). One SNP, preferably a parsimony-informative one, was randomly chosen from each locus to produce a concatenated data set/alignment of these SNPs. Parsimony-informative SNPs were preferred because the other variable sites contain singletons, and were less informative for phylogenetic reconstruction (data sets without preferring parsimony-informative SNPs can be obtained upon request). The output was then manually edited, saved in NEXUS format and imported into PAUP* (ver. 4.0a147; Swofford 2002). The species tree was constructed using SVDQuartets, with an exhaustive search of all possible quartets (a total of 178 005 quartets), and Quartet FM (Fiduccia & Mattheyses 1982) to build the most likely species tree based on the sampled quartets' likelihood scores implemented in PAUP *. Each SNP was treated as an independent locus even though the input data set was in a concatenated format. The reliability of the branches of the species tree was evaluated using the same search options, with 1000 bootstrapping replicates in PAUP *.

Testing introgression between species

D- (Durand *et al.* 2011) and D_{FOIL} (Pease & Hahn 2015) statistics were used to test for significant genetic introgression between species of Hercules beetles. Based on the level of predicted geographic overlaps using ecological niche modelling published by Huang & Knowles (2016), I estimated the statistical significance of *D*s between species for two groups: (i) allopatric species and (ii) parapatric species (i.e. species pairs with >10% predicted geographic overlap). Genetic introgression was not expected between allopatrically distributed species that do not, or did not based on the last glacial maximum (LGM) distribution, have the opportunity for direct contact. Group (1) tests were to identify any errors in the data inputs that may lead to unexpected or erroneous interpretations (e.g. due to uncertain phylogenetic relationship or shared ancestry among potential allele–donor species; see discussion in Eaton *et al.* [2015]). Pairs of species were assigned to the two geographic distributional groups, with a sister taxon (for D_{FOIL}) and an outgroup taxon (for both *D* and D_{FOIL}) chosen randomly from the corresponding lineages. I used the edited .loci file from the previous section and PYRAD ver. 3.0.66 (Eaton 2014) to perform these tests of introgression. Multiple samples per taxon were included in the calculation, and testing for the significance of *D*s and information from heterozygous sites was incorporated by weighting (using the algorithms from Durand *et al.* 2011).

Four-taxon *D*-statistics were used to detect the significance of excessive derived alleles shared between taxa potentially from genetic introgression (Durand *et al.* 2011). The ancestral allele (A allele) from the outgroup taxon and the derived allele (B allele) from the ingroup taxa of interest can be identified using an outgroup taxon. A four-taxon analysis was labelled based on the phylogeny as (((S1,S2),S3),O), where O is the outgroup taxon, S2 and S3 are taxa of interest with potential introgression, and S1 is the sister taxon to S2. *D*-statistics then compares the frequencies of derived alleles shared among the three ingroup taxa, that is allele pattern ABBA vs. BABA. The expected frequency should be significantly higher for ABBA than BABA loci if non-random events, such as introgression, have occurred but should be equal if these discordant sites arise by

the random sorting of ancestral polymorphisms. Note that *D*-statistics might result in false positives when used to locate genomic regions of introgressive origins (Martin *et al.* 2015). In this study, however, *D*-statistics were used only to test for introgression based on randomly sampled SNPs from the genome, which was less susceptible to false positives.

D-statistics were subsequently extended to infer directional genetic introgression based on the SNP data. For example, partitioned *D*-statistics used an additional taxon from the S3 lineage to help to identify the direction of introgression. A five-taxon partitioned *D*-statistics was therefore labelled as (((S1,S2),($S3_1$,$S3_2$)),O). Partitioned *D*-statistics compared the relative frequencies of the allele patterns ABBBA vs. BABBA, ABBAA vs. BABAA and ABABA vs. BAABA and used these patterns to infer the direction of introgression (Eaton & Ree 2013). Partitioned *D*-statistics, however, can also produce false positives, so another modified version of *D*-statistics for directional introgression was used (D_{FOIL}; Pease & Hahn 2015). D_{FOIL} tests, which have been shown to be statistically robust, accounted for more variations in the relative frequency of bi-allelic loci among the four focal taxa. D_{FOIL} analysis can correctly identify directional introgression even for minimal genetic interchange (Pease & Hahn 2015). D_{FOIL} tests were subsequently applied in the present study to test for directional introgression for the cases of introgression inferred using *D*-statistics. I interpreted the results of directional introgression based on the output signs, following Pease & Hahn (2015).

The standard deviation of the *D*s was measured for each test with 1000 bootstrap replicates, where RAD loci were resampled and replaced with the same number of loci in the original data set (Eaton & Ree 2013). A Z-score, which measures the number of standard deviations that deviate from 0, was calculated from the observed *D* value. The statistical significance was assessed at $P < 0.01$ after Bonferroni correction for multiple testing. Multiple taxa from the outgroup lineage were applied in the analyses, which generates a conservative estimate of the number of derived alleles shared between taxa (Eaton *et al.* 2015).

Results

Differences in body coloration

A total of 27, 10, 12 and 33 Hercules beetles were assigned to the NAW, CAW, CAG and ONG groups, respectively (Fig. 1). A MANOVA identified a significant grouping structure based on body coloration (Table 2). The Tukey's HSD analyses further indicated that the body coloration in the NAW group was significantly distinct from that of the other three groups (Fig. S2, Supporting information and Table 3), but the differences among these groups (CAW, CAG and ONG) were not statistically significant (Table 3). This study therefore identified two body coloration groups in the Hercules beetles: (i) the Nearctic Hercules beetle group (NAW), which includes three species of White Hercules beetles from North America, and (ii) the Neotropical Hercules beetle group (CAW + CAG + ONG), which includes all the remaining species of Hercules beetles from the Neotropical region. Phenotypically, ma and mo have body coloration similar to the parapatric taxon sep but different from the other species of White Hercules beetles, supporting a pattern of phenotypic assimilation of body coloration.

Table 2 MANOVA results based on differences in body coloration

	d.f.	SS	MS	*F*	R^2	*P*
Groups	3	235 437	78 479	64.862	0.71385	< 0.001
Residuals	78	94 375	1210		0.28616	
Total	81	32 9812			1	

Differences in horn shape

The position of the thoracic denticle based on a geometric morphometric analysis differed significantly among horn-shape groups (Table 4). Pairwise comparisons between groups indicated that the position of the thoracic horn denticle differed significantly between NGL and the remaining groups (Table 5). The position of the denticle also differed between NWS and CGS and between NWS and NGS. The differences between CWS and CGS and between CWS and NGS were marginally significant ($P = 0.076$ and 0.060, respectively). The similarity in horn shape between distantly related Hercules beetles in Central America cannot be rejected.

Table 3 Pairwise comparisons* between groups based on colour differences

	NAW	CAW	CAG	ONG
NAW		**0.010**	**0.024**	**0.011**
CAW	**0.013**		0.695	0.455
CAG	**0.023**	0.699		0.693
ONG	**0.008**	0.455	0.696	

Significant results ($P < 0.05$) are in bold.
*Observed (below the diagonal) and permuted *P* values (above the diagonal) are shown.

Table 4 ANOVA results based on differences in horn shape among groups

	d.f.	SS	MS	R^2	F	P
Group	4	2.01751	0.50438	0.76351	75.062	<0.001
Residuals	93	0.62491	0.00672			
Total	97	2.64242				

Table 5 *Pairwise comparisons between horn-shape groups

	NWS	CWS	CGS	NGS	NGL
NWS		0.917	**0.009**	**0.006**	**<0.001**
CWS	0.027		0.076	0.060	**<0.001**
CGS	0.142	0.135		0.731	**<0.001**
NGS	0.141	0.136	0.037		**<0.001**
NGL	0.323	0.312	0.204	0.200	

Significant results are highlighted in bold.
*The calculated observed distances between groups are shown below the diagonal, and estimated probabilities are shown above the diagonal.

Summary of rad data

A total of 90 874 341 raw reads passed the initial quality control, with an average of 1 817 487 good reads per sample (SE = 288 344). Clustering analyses within samples identified an average of 36 114 clusters (SE = 4020) with at least 5× sequencing depth (mean sequencing depth = 32× per cluster within a sample) among the samples (Fig. S4, Supporting information). Sequence clustering, alignment across samples and editing using customized R scripts (Figs S5 and S6, Supporting information) provided a final data set containing 43 205 variable loci (see Supporting information). The average number of loci per sample after filtering was 15 282 (SE = 1411). The number of loci shared between samples from the final data set was not correlated with their phylogenetic relatedness (Fig. S7, Supporting information), which would be expected if the amount of missing data between samples was caused by mutations in the restriction sites.

Species tree

Three distinct phylogenetic lineages were identified with absolute bootstrapping support (Fig. 1): the outgroup (subgenus *Theogenes*, dn and ds), White Hercules (gr, hy, ma, mo, and ty) and Giant Hercules (sep, occ, tri, blu, rei, her, lic, pas, mor and ecu) (see Table 1 for species names and abbreviations). The species gr and hy were White Hercules sister species (100% bp support; Fig. 1); ty was a sister taxon to the Central American White Hercules lineage including ma and mo (100% bp support; Fig. 1). The sister relationship between ma and mo was also highly supported by bootstrapping (Fig. 1). A Giant Hercules lineage including sep and occ diverged first from the remaining species and was geographically constrained to the western side of the Andes (Central America and the Chocó-Darién moist forest; Fig. 1). The remaining species of Giant Hercules beetles formed two geographic lineages: (i) the Amazonian lineage, which included four species (lic, pas, ecu and mor) that were constrained to the eastern side of the Andes, and (ii) the Antilles lineage, which comprised four species (her, rei, tri and blu) from the Lesser Antilles islands and the nearby geographic region (Fig. 1). The Giant Hercules lineages were highly supported (>70% bp support in all branching patterns), where uncertainties were more often found at deeper evolutionary levels (e.g. the support of sister taxa relationships between species was always very high [>95%], and the branching patterns between the main lineages, e.g. the Amazonian and Antilles lineages, were moderately supported [85%]; Fig. 1). When one random SNP (without preferring a parsimony-informative site) was chosen from each locus for the SVDQuartets analysis, the same geographic lineages in White and Giant Hercules were revealed – for example, the Antillean and Amazonian lineages of Giant Hercules beetles. However, the bootstrapping supports for branching patterns among these lineages within Giant Hercules beetles decreased drastically (<50%; data not shown but can be obtained upon request).

Introgression between species

Significant *D*-statistics were identified only between parapatric taxa (Table 6 and Fig. 3). For example, species from the geographically isolated Lesser Antilles lineage provided no significant *D* values. Parapatric distribution, however, did not predict the presence of introgression (e.g. between lic and ecu; $D = 0.029$ and $Z = 1.632$). Phylogenetic distance did not limit the presence of introgression between taxa (i.e. introgression was identified between distantly related lineages, e.g. between the common ancestor of ma + mo and the Giant Hercules taxon sep; $D = 0.179$ and $Z = 7.3838$), nor did the phenotypic similarity in putative adaptive traits (e.g. a significant *D* value was identified between hy and sep [different colour; $D = 0.134$ and $Z = 5.3596$] and between occ and lic [different horn shape; $D = 0.078$ and $Z = 4.7683$]). Genomic regions (number of loci) of putative introgressed origin, however, were significantly more common between species with similar body colour (e.g. ma vs. sep) than between dissimilar species (e.g. hy vs. sep) ($D = 0.189$ and $Z = 4.7488$).

Table 6 Results from the four-taxon *D*-statistic tests.

Group*	S1	S2	S3	O	ABBA/BABA	No. loci	Cross†	*D*	Z‡
1	her	rei	tri	ty + hy	299/272	8288	NO	0.047	0.9912
1	blu	tri	rei	ty + hy	518/586	7593	NO	−0.061	1.817
1	hy	gr	ty	pas + her	814/911	12 038	NO	−0.056	2.1617
1	mor	ecu	pas	ty + hy	1297/1184	13 841	NO	0.046	2.2252
1	mo	ma	ty	pas + her	681/792	9737	NO	−0.075	2.6829
1	ma	ty	her	dn + ds	143/202	2270	YES	−0.172	2.6460
1	rei	her	ty	dn + ds	185/178	6179	YES	0.020	0.3188
1	mor	ecu	blu	sep	845/845	12 375	NO	0.000	0.0172
1	tri	blu	ecu	sep	605/556	9440	NO	0.042	1.4635
1	mor	ecu	blu + tri	sep	895/901	13 109	NO	−0.003	0.1329
1	rei + her	blu + tri	ecu	sep	1188/1063	14 622	NO	0.055	2.3226
1	ecu	pas	blu	sep	1099/1038	14 397	NO	0.028	1.2244
1	tri	blu	pas	sep	721/673	10 494	NO	0.034	1.2753
1	mor + ecu	pas	blu + tri	sep	1397/1336	17 770	NO	0.022	1.1753
2	occ	sep	hy	dn + ds	696/606	9605	YES	0.069	2.7168
2	gr	hy	sep	dn + ds	869/663	10 632	YES	0.134	**5.3596**
2	gr	hy	ma	pas + her	1024/734	12 719	NO	0.165	**6.5174**
2	gr	hy	ma + mo	pas + her	1558/1106	18 108	NO	0.170	**8.6313**
2	ty	ma + mo	hy	pas + her	1033/987	11 009	NO	0.023	0.9281
2	mo	ma	hy	pas + her	885/831	11 003	NO	0.032	1.2571
2	sep	occ	lic	ty + hy	1734/1482	21 119	NO	0.078	**4.7683**
2	pas	ecu	lic	ty + hy	1797/1696	20 938	NO	0.029	1.6326
2	mor	ecu	lic	ty + hy	1063/1212	13 967	NO	−0.066	2.9888
2	mo	ma	sep	dn + ds	694/625	7914	YES	0.052	2.1025
2	occ	sep	ma + mo	dn + ds	1336/1166	15 864	YES	0.068	3.8462
2	ty	ma + mo	sep	dn + ds	1072/747	9543	YES	0.179	**7.3838**
2	ma	hy	sep	dn + ds	249/365	3363	YES	−0.189	**4.7488**

*Allopatric and parapatric comparisons are indicated by 1 and 2, respectively.
†Testing introgression between one Giant and one White species of Hercules beetle.
‡Significant results ($P < 0.01$ after Bonferroni correction) are highlighted in bold.

The D_{FOIL} analyses identified only two cases of significant directional genetic introgression (Table 7 and Fig. 3). The directional introgression was inferred from the putative allele–donor lineage to the allele–receiver lineage based on the pattern of body coloration assimilation (i.e. from sep + occ to ma + mo and from sep to ma). The remaining cases of introgression supported by *D*-statistics were inferred as either bidirectional or ancestral introgression (e.g. between occ + sep and lic; Table 7). The case of introgression between hy and sep + occ could only be supported when gr was used as the sister taxon to hy and not when the other species of White Hercules beetles were used as a sister taxon to hy (e.g. ty). The pattern of introgression between hy and ma + mo (or between hy and ma) inferred from the *D*-statistics became inconclusive based on the D_{FOIL} results (Table 7).

Discussion

The importance of introgression between species has been highly debated in systematics and evolutionary biology (Mayr 1963; Mallet 2007; Harrison & Larson 2014; Payseur & Rieseberg 2016). Only until recently have empirical studies become available that specifically tested for the prevalence of, and conditions that may favour/constrain introgression in an evolutionary lineage (e.g. Streicher *et al*. 2014; Eaton *et al*. 2015). By sampling all recognized species of Hercules beetles (Huang 2016), which minimized the effect of ghost lineages on inferring introgression (extinct lineages were not taken into account; see discussion in Pease & Hahn 2015), and by using two lanes of Illumina sequencing, I was able to reconstruct the history of introgression between species of Hercules beetles. This history demonstrates that introgression, detected by genetic data, occurred between geographically proximate species, as expected. Introgression, though, was not constrained by the phylogenetic relatedness in the Hercules beetle system. Phenotypic similarity in putative adaptive traits between parapatric species of Hercules beetles was not associated with the presence or absence of introgression. Interestingly, phenotypic assimilation in body coloration among the Central American species

was coincided with significant and directional introgression from the putative allele–donor lineage to the allele–receiver lineage. Introgression between species was thus not rare in a system where speciation is predominantly allopatric, implying that introgression may not be a negligible evolutionary force. Conditions that favour or limit the presence of introgression, however, remain elusive. Below, I discuss the significance of genetic interchange between species and the implication for adaptive phenotypic evolution and systematics.

Introgression is not constrained by phylogenetic relatedness nor by phenotypic similarity

My results indicated that introgression may not only be common, but also can occur between distantly related species of Hercules beetles (e.g. between one White Hercules and one Giant Hercules species that diverged around 4 Ma, which was calibrated using a beetle mitochondrial clock; Huang & Knowles 2016). Similar results have been reported between distantly related barking frogs (ca. 4 Ma; Streicher *et al.* 2014) and between *Heliconius* butterflies (ca. 2.5 Ma; Kronforst 2008). Introgression between species may be constrained to closely related and recently diverged taxa, because epistatic interactions can evolve between genes in the genome. Genic incompatibility between fully diverged species can subsequently establish reproductive isolation (Dobzhansky 1936; Mayr 1942). However, the complete formation of reproductive isolation between species is a slow process under the classical model of allopatric speciation (Mayr 1963), and introgression may be possible upon secondary contact for a long time after divergence. Hercules beetle system is a prime example of the progressive accumulation of molecular and phenotypic divergences between allopatric evolutionary lineages (Huang & Knowles 2016) and the results here support the prediction from the classical view of allopatric speciation.

My study also found that genetic introgression can occur between species with distinct phenotypes of putative adaptive traits (Fig. 3; e.g. between occ + sep and lic [horn shape] and between ma + mo and hy [body coloration]). Speciation with gene flow has been demonstrated in many empirical cases (Nosil 2008), where strong selection against maladapted intermediate hybrids has been hypothesized. My results further demonstrated that genetic introgression can occur in Hercules beetles between phenotypically distinct and phylogenetically distant species after secondary contact (only nonsister species overlap geographically), while their evolutionary independence remains well supported (Huang & Knowles 2016). The intermediate forms may have a lower fitness, either because of their increased visibility to potential predators (body coloration functions as camouflage on differently coloured tree trunks; Hinton & Jarman 1973) or the lack of specific phenotypic features to guard against predation and/or competition (horns can be defensive devices in addition to weapons for male–male and/or interspecific competition; Jarman & Hinton 1974).

Introgression was interestingly not detected between lic and ecu, where an extensive geographic overlap between species was predicted and similar phenotypes of both putative adaptive traits were found (Fig. 3). Prezygotic isolation as differentiation in seasonal phenology and altitudinal preference may play important roles in this case. For example, ecu in Ecuador has an annual swarm after the onset of the rainy season in the Amazonian rainforest (from April to June; Huang 2012). The Andean cloud-forest species lic, however, does not typically have a swarming season. The factors constraining the presence of introgression nevertheless remain unclear, and more life history studies are needed to better understand the prezygotic factors isolating lic and ecu.

Implications for genetic introgression

The type of habitat for species ma and mo is distinct from those of the other species of White Hercules beetles. The introgression of alleles from a codistributed species of Giant Hercules beetles into ma and mo under this novel condition could have facilitated rapid adaptation (Hedrick 2013). The assimilation of body coloration was clearly associated with directional introgression from a putative allele–donor to allele–receiver lineages (Fig. 3; from sep to ma and from sep + occ to ma + mo). This association, however, was not well supported for the evolution of male horn shape (no directional introgression from ma and/or mo into the common ancestor of sep and occ), and the assimilation of horn shape was only marginally supported (Table 5). Additionally, introgression, as discussed in the previous section, can occur between phenotypically distinct species, which complicates the predicted patterns of adaptive introgression, that is introgression can simply be stochastic. For example, introgression was detected between sep and hy using *D*-statistics (Fig. 3 and Table 6). Nevertheless, sep and ma shared significantly more derived alleles than sep and hy (Table 6). Directional introgression was therefore found between species with adaptive body coloration assimilation, but the number of shared derived alleles was also significantly higher between the phenotypically similar species pair than the other case of introgression without phenotypic assimilation.

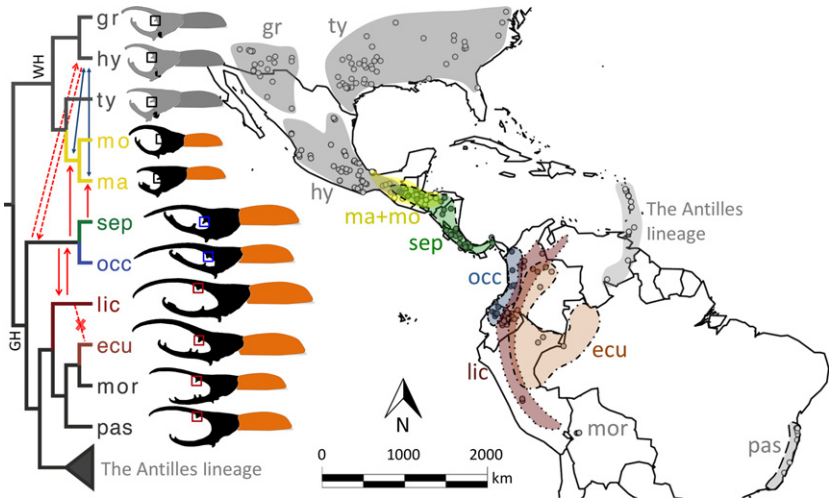

Fig. 3 A species tree of *Dynastes* beetles with inferred directional introgression shown as solid red arrows between branches and a map showing the predicted geographic distributions for the Hercules beetle species. Blue arrows indicate introgression inferred by ABBA/BABA tests, but no clear patterns of directional introgression could be identified by D_{FOIL} analyses. Dashed red arrows indicate inferred directional introgression when only a specific sister taxon (S_2 = gr, see Table 7) was used in the D_{FOIL} analyses. A dashed line connecting taxa lic and ecu with a red cross indicates a lack of significant genetic introgression between the two nonsister species, even with a large area of predicted geographic overlap and similar phenotypes. Overlapped geographic distributions are colour-coded (modified from Huang & Knowles [2016]). Body coloration groups are illustrated using different colours, and red, blue and black boxes indicate the horn-shape groups. See Table 1 for the taxon names and abbreviations. WH, White Hercules; GH, Giant Hercules. Abbreviations and their corresponding taxonomic names can be found in Table 1. [Colour figure can be viewed at wileyonlinelibrary.com].

Similar patterns have been found in *Heliconius* butterflies (Kronforst 2008; Martin *et al*. 2013), where introgressed loci putatively responsible for mimetic patterns have been identified in subsequent studies (Pardo-Diaz *et al*. 2012). My results suggest that introgression could be a potential source for adaptation, and its effect should be tested against the effects of new mutation and standing genetic variation (Hedrick 2013) at the genus level, that is introgression does not have to be between closely related and recently diverged species complexes. Future studies that focus on the functions of the introgressed loci and variations in the genomic architecture (e.g. genomic islands of divergence) between species of Hercules beetles are needed to better understand the nature and consequence of introgression.

Phylogenetic reconstruction can be affected by introgression (e.g. Eaton *et al*. 2015). For example, phylogenetic relationships between species and estimates of divergence times could both be altered among live oaks if introgressed loci were not taken into account. Although the effect of introgression on species tree reconstruction was not accounted for in this study, the species tree for the Hercules beetles based on genome-wide SNP data mirrors the morphological similarities. The main geographic and morphological lineages in White and Giant Hercules beetles were well supported (Fig. 1). Introgressions were inferred between species from different well-supported lineages. For example, significant introgression and phenotypic assimilation was identified between ma + mo and sep + occ. I acknowledge that the reconstructed Hercules beetle species tree may be affected by historical introgression as phylogenetic uncertainties were found among the three main geographic/morphological lineages in Giant Hercules beetles (see results section for details). Introgression between species from different Giant Hercules beetle geographic lineages, for example between lic and occ, may result in discordance in the genealogical histories between loci. Future studies that identify and quantify the genomic regions of introgression and that determine the phylogenetic relationships among species based on these different regions in Hercules beetles (similar to those by Wen *et al*. (2016) and Zhang *et al*. (2016)) are needed to better address this question. Introgression between species pairs reported in this study, however, should be robust because different choices of sister and outgroup taxa/lineages were utilized in the analyses and only consistent results were taken as evidence for introgression.

My study used only *D*- and D_{FOIL} statistics to infer introgression events. Although the pattern of excessive shared derived alleles between a species pairs supports introgression, alternative hypotheses that may explain the same genetic pattern, for example ancestral structured populations (Durand *et al*. 2011), cannot be ruled

Table 7 Results from the D_{FOIL} analyses

Case	$S_1S_2S_3S_4O$	D_{FO}	D_{IL}	D_{FI}	D_{OL}	No. loci	Signs	Inference
sep + occ	sep,occ,ma,mo,dn + ds	0.038 (3.93)*	0.039 (3.95)	0.034 (3.68)	0.035 (3.80)	8428	++++	?
vs.	sep,occ,ma,ty,dn + ds	0.138 (11.79)	0.124 (10.67)	0.039 (3.58)	0.022 (2.00)	6436	+++0	$S_1 \Rightarrow S_3$
ma + mo	sep,occ,mo,ty,dn + ds	0.094 (9.94)	0.084 (8.99)	0.029 (3.31)	0.018 (1.99)	9679	++00	$S_{12} \Rightarrow S_3 \mid S_3 \Rightarrow S_{12}$
	sep,occ,ma + mo,ty,dn + ds	0.107 (10.83)	0.097 (10.03)	0.026 (3.07)	0.015 (1.66)	9822	++00	$S_{12} \Rightarrow S_3 \mid S_3 \Rightarrow S_{12}$
	ma,mo,sep,pas + ecu,dn + ds	0.066 (7.99)	0.064 (7.64)	0.040 (4.23)	0.038 (4.23)	9090	++++	?
	ma,mo,occ,pas + ecu,dn + ds	0.044 (4.92)	0.039 (4.59)	0.042 (4.33)	0.035 (3.91)	8887	++++	?
	ma,mo,sep + occ,pas + ecu,dn + ds	0.055 (7.03)	0.052 (6.79)	0.037 (4.14)	0.033 (4.00)	9690	++++	?
	ma + mo,ty,sep + occ,pas + ecu,dn + ds	0.057 (8.47)	0.022 (3.07)	0.100 (11.87)	0.062 (7.50)	11 150	+0++	$S_3 \Rightarrow S_1$
	ma,mo,sep,her + rei,dn + ds	0.079 (7.77)	0.079 (8.30)	0.036 (3.63)	0.036 (3.74)	8398	++++	?
	ma,mo,occ,her + rei,dn + ds	0.059 (5.92)	0.065 (6.86)	0.033 (3.31)	0.040 (3.99)	8204	++0+	?
	ma,mo,sep + occ,her + rei,dn + ds	0.068 (7.26)	0.072 (7.79)	0.031 (3.38)	0.037 (3.72)	8917	++0+	?
	ma + mo,ty,sep + occ,her + rei,dn + ds	0.068 (7.74)	0.034 (3.34)	0.098 (11.88)	0.061 (6.63)	10 039	+0++	$S_3 \Rightarrow S_1$
sep into hy	sep,occ,hy,gr,dn + ds	0.049 (5.33)	0.041 (4.48)	0.012 (1.34)	0.003 (0.38)	9465	++00	$S_{12} \Rightarrow S_3 \mid S_3 \Rightarrow S_{12}$
	hy,ty,sep,occ,dn + ds	0.029 (2.87)	0.017 (1.68)	0.026 (2.29)	0.017 (1.44)	6982	0000	None
	hy + gr,ty,sep,occ,dn + ds	0.017 (1.94)	0.017 (1.88)	−0.016 (1.79)	−0.016 (1.83)	9772	0000	none
occ into lic	occ,sep,lic,ecu,ty + gr	0.063 (11.82)	0.059 (11.00)	0.012 (2.28)	0.008 (1.37)	22 951	++00	$S_{12} \Rightarrow S_3 \mid S_3 \Rightarrow S_{12}$
	occ,sep,lic,ecu + mor + pas,ty + gr	0.076 (16.21)	0.075 (16.57)	0.016 (2.72)	0.014 (2.52)	27 427	++00	$S_{12} \Rightarrow S_3 \mid S_3 \Rightarrow S_{12}$
ma + mo into hy	ma,mo,hy,gr,ecu + pas	0.098 (12.06)	0.094 (12.22)	0.033 (3.96)	0.029 (3.66)	12 377	++++	?
	ma + mo,ty,hy,gr,ecu + pas	0.097 (12.66)	0.060 (6.83)	0.068 (7.88)	0.034 (4.14)	12 091	++++	?

*Values in parentheses are *Z*-scores.

out using my data set. However, ancestral structure is an unlikely explanation for excessive shared derived alleles between distantly related Hercules beetle species. For example, ancestral structures within White and Giant Hercules beetles that isolated ma + mo and occ + sep from their sister lineages would have existed even before the divergence between Giant and White Hercules beetles (Durand *et al.* 2011). Historical introgression is the most plausible explanation in this case. False positives using *D*-statistics have been shown when significant results were used to locate genomic regions of introgression (Martin *et al.* 2015). Reduced genetic diversity in a genomic region can result in an inflated *D* value, which leads to false positives. The RAD loci were short (110 base pairs) with a few segregating sites per locus and cannot be effectively used to access if those that support introgression have relatively low genetic variation compared to other background loci. Complementary analyses that account for the effect of genetic diversity on inferring introgression, for example by partitioning the genome into short loci blocks (ranging from 1 to 10 kb; Martin *et al.* 2015), can help to validate my results and provide a more detailed picture of introgression patterns in the genome when a linkage map among the RAD loci becomes available.

Conclusion

Genetic introgression was identified in an evolutionary lineage where allopatric divergence had been inferred as the predominant mode of speciation, which supports the prediction from the classical speciation model that the completion of reproductive isolation is a slow process and introgression can be possible upon secondary contact. Introgression was also identified between distantly related species and between species with distinct phenotypes of putative adaptive traits. Geographic overlap between species was the only prerequisite, but

not a predictor, for introgression. The results of this study, however, support the growing evidence that introgression between species could be a potential source for adaptive phenotypic evolution, which warrants further genomic investigation.

Acknowledgements

I thank L. Lacey Knowles and members of the Knowles laboratory, especially Andrea Thomaz, Melisa Olave and Jeet Sukumaran, for helpful discussions and suggestions. I also thank Dr. William Blackhall and Todd Widhelm for English editing. Two anonymous reviewers provided insightful comments and suggestions on previous version of this manuscript. This work was supported by the Doctoral Dissertation Improvement Grant from the National Science Foundation (DEB-15-01462).

References

Adams DC, Otarola-Castillo E (2013) Geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, **4**, 393–399.

Brower AVZ (2013) Introgression of wing pattern alleles and speciation via homoploid hybridization in *Heliconius* butterflies: a review of evidence from the genome. *Proceedings of the Royal Society B-Biological Sciences*, **280**, 20122302.

Chifman J, Kubatko L (2014) Quartet inference from SNP data under coalescent model. *Bioinformatics*, **30**, 3317–3324.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.

Dixon P (2003) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, **14**, 927–930.

Dobzhansky T (1936) Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics*, **21**, 113–135.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**, 2239–2252.

Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology*, **62**, 689–706.

Eaton DAR, Hipp AL, González-Rodríguez A, Cavender-Bares J (2015) Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution*, **69**, 2587–2601.

Fiduccia CM, Mattheyses RM (1982) A linear time heuristic for network partitions. In: *Proceedings 19th IEEE Design Automation Conference*, pp. 175–181.

Gompert Z, Lucas LK, Nice CC, Fordyce JA, Forister ML, Buerkle CA (2012) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, **66**, 2167–2181.

Hardy M (2003) Description of a new species of *Dynastes* Kirby (Coleoptera Scarabaeidae Dynastinae) from North and Central America. *Besoiro*, **9**, 3–7.

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, **105**, 795–809.

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

Hinton HE, Jarman GM (1973) Physiological colour change in the elytra of the Hercules beetle, *Dynastes hercules*. *Journal of Insect Physiology*, **19**, 533–549.

Huang J-P (2012) Ecuador—Land of the giant Hercules Beetles. *Scarabs*, **71**, 1–12.

Huang J-P (2016) Hercules Beetles (Genus *Dynastes*, Dynastidae): a revisionary study based on molecular, morphological, ecological differences, and geographic distribution. *University of Michigan Museum of Zoology Miscellaneous Publication*, in review

Huang J-P, Knowles LL (2016) The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single Bayesian approach in Hercules Beetles. *Systematic Biology*, **65**, 685–699.

Jarman GM, Hinton HE (1974) Some defense mechanisms of the Hercules beetle, *Dynastes hercules*. *Journal of Entomology Series A: General Entomology*, **49**, 71–80.

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

Lai J (2014) Hybridization of *Dynastes h. hercules* and *Dynastes h. paschoali*. *Scarabs*, **76**, 5–9.

Lai J, Ko H-P (2008) *For the Love of Rhinoceros and Stag Beetles*, 2nd edn. Morning Star Publisher, Taipei, Taiwan.

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

Martin SH, Dasmahapatra KK, Nadeau NJ *et al.* (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, **23**, 1817–1828.

Martin SH, Davey JW, Jiggins CD (2015) Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, **32**, 244–257.

Mayr E (1942) *Systematics and the Origin of Species From the Viewpoint of a Zoologist*. Columbia University Press, Cambridge, Massachusetts.

Mayr E (1963) *Animal Species and Evolution*. Harvard University Press, Cambridge, Massachusetts.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103–2106.

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

Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. *Molecular Ecology*, **25**, 2337–2360.

Pease JB, Hahn MW (2015) Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology*, **64**, 651–662.

Rheindt FE, Fujita MK, Wilton PR, Edwards SV (2014) Introgression and phenotypic assimilation in Zimmerius flycatchers (Tyrannidae): population genetic and phylogenetic inferences from genome-wide SNPs. *Systematic Biology*, **63**, 134–152.

Ru D, Mao K, Zhang L, Wang X, Lu Z, Sun Y (2016) Genomic evidence for polyphyletic origins and interlineage gene flow within complex taxa: a case study of *Picea brachytyla* in the Qinghai-Tibet Plateau. *Molecular Ecology*, **25**, 2373–2386.

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH image to imageJ: 25 years of image analysis. *Nature Methods*, **9**, 671–675.

Seehausen O (2013) Conditions when hybridization might predispose populations for adaptive radiation. *Journal of Evolutionary Biology*, **26**, 279–281.

Streicher JW, Devitt TJ, Goldberg CS, Malone JH, Blackmon H, Fujita M (2014) Diversification and asymmetrical gene flow across time and space: lineage sorting and hybridization in polytypic barking frogs. *Molecular Ecology*, **23**, 3273–3291.

Suarez-Gonzalez A, Hefer CA, Christe C *et al.* (2016) Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Molecular Ecology*, **25**, 2427–2442.

Swofford DL (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Ver. 4. Sinauer, Sunderland, Massachusetts.

Venables WN, Ripley BD (2002) *Modern Applied Statistics With S*, 4th edn. Springer, New York, New York.

Wen D, Yu Y, Hahn MW, Nakhlen L (2016) Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Molecular Ecology*, **25**, 2361–2372.

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

Zhang W, Dasmahapatra KK, Mallet J, Moreira GRP, Kronforst MR (2016) Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biology*, **17**, 25.

J-P. H. conceived the study, generated and analyzed the data, and wrote the paper.

Data accessibility

Illumina reads are deposited in NCBI Sequence Read Archive (SRP076992). Supplementary figures and files (including assembled RADseq alignments, image files, tps and RGB value files) are deposited in Dryad (doi:10.5061/dryad.4c5f9). Customized R codes for data processing are also archived (https://github.com/airbugs/Dynastes_introgression).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Examples of how body coloration was measured in Hercules beetles.

Fig. S2 Examples of how landmarks were selected for the thoracic horn shape.

Fig. S3 Results from Tukey's HSD tests based on body coloration between pairs of Hercules beetle groups (see materials and methods).

Fig. S4 Summaries of the number of good reads and their sequencing depths after sequencing for samples that are included in the molecular analyses.

Fig. S5 Results from the original PYRAD output file (.loci output).

Fig. S6 Summaries of molecular data in the final dataset.

Fig. S7 The number of shared loci (probability calculated from a total of 43,205 loci) between sequenced samples.