Multiple Loci Control Eyespot Number Variation on the Hindwings of *Bicyclus anynana* Butterflies

Angel G. Rivera-Colón,*,† Erica L. Westerman,‡ Steven M. Van Belleghem,† Antónia Monteiro,§,,1 and Riccardo Papa†,††,1**

*Department of Evolution, Ecology, and Behavior, University of Illinois, Urbana-Champaign, Illinois 61801, †Department of Biology, University of Puerto Rico, Rio Piedras Campus, San Juan, 00925, Puerto Rico, ‡Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas 72701, §Department of Biological Sciences, National University of Singapore, Singapore 117543, **Yale-NUS College, Singapore 138609, and ††Molecular Sciences and Research Center, University of Puerto Rico, San Juan, 00926, Puerto Rico

ORCID IDs: 0000-0001-9097-3241 (A.G.R.-C.); 0000-0002-3575-8298 (E.L.W.); 0000-0001-9399-1007 (S.M.V.B.); 0000-0001-9696-459X (A.M.); 0000-0002-7986-9993 (R.P.)

ABSTRACT The underlying genetic changes that regulate the appearance and disappearance of repeated traits, or serial homologs, remain poorly understood. One hypothesis is that variation in genomic regions flanking master regulatory genes, also known as input–output genes, controls variation in trait number, making the locus of evolution almost predictable. Another hypothesis implicates genetic variation in up- or downstream loci of master control genes. Here, we use the butterfly *Bicyclus anynana*, a species that exhibits natural variation in eyespot number on the dorsal hindwing, to test these two hypotheses. We first estimated the heritability of dorsal hindwing eyespot number by breeding multiple butterfly families differing in eyespot number and regressing eyespot numbers of offspring on midparent values. We then estimated the number and identity of independent genetic loci contributing to eyespot number variation by performing a genome-wide association study with restriction site-associated DNA sequencing from multiple individuals varying in number of eyespots sampled across a freely breeding laboratory population. We found that dorsal hindwing eyespot number has a moderately high heritability of ~0.50 and is characterized by a polygenic architecture. Previously identified genomic regions involved in eyespot development, and novel ones, display high association with dorsal hindwing eyespot number, suggesting that homolog number variation is likely determined by regulatory changes at multiple loci that build the trait, and not by variation at single master regulators or input–output genes.

KEYWORDS serial homology; genetics; *apterous*; eyespot number; *Bicyclus anynana*; genetic architecture

BODY plans often evolve through changes in the number of repeated parts or serial homologs by either addition or subtraction. For instance, the pelvic fins of vertebrates are inferred to have originated by addition to a body plan displaying only pectoral fins, perhaps via cooption of the pectoral or caudal fin developmental programs to a novel location in the body (Ruvinsky and Gibson-Brown 2000; Larouche *et al.* 2017). In insects, the absence of limbs and wings in the abdomen is inferred to be due to a process of subtraction, *i.e.*, via the repression or modification of limbs and wings in these segments by *hox* genes (Galant and Carroll 2002; Ronshaugen *et al.* 2002; Tomoyasu *et al.* 2005; Ohde *et al.* 2013). Targets of abdominal *hox* genes, belonging to the limb or wing gene regulatory networks, are likely to underlie loss of limb/wing number in these body segments (Tomoyasu *et al.* 2005; Ohde *et al.* 2013), although these mutations have not yet been identified. Thus, while serial homolog number variation is a common feature in the evolution of organisms' body plans, the underlying genetic changes that regulate the appearance and disappearance of these repeated traits remain poorly understood.

Studies in *Drosophila* have contributed most to the identification of the genetic basis underlying the evolution of

doi: https://doi.org/10.1534/genetics.120.303059
Manuscript received May 24, 2019; accepted for publication December 26, 2020; published Early Online February 4, 2020.
Supplemental material available at figshare: https://doi.org/10.25386/genetics.11766624.
[1]Corresponding author: Department of Biological Sciences, National University of Singapore, 14 Science Dr. 4, Singapore 117543. E-mail: antonia.monteiro@nus.edu.sg; and Julio Garcia Diaz Biology Bldg. #213, Department of Biology, University of Puerto Rico, Río Piedras Campus, San Juan, PR 00931. E-mail: rpapa.lab@gmail.com

serial homolog number. Larvae of different species have different numbers of small hairs, or trichomes, in their bodies, and variation in regulatory DNA around the gene *shavenbaby* appears to be largely responsible for this variation (McGregor *et al.* 2007). Moreover, *shavenbaby* has been labeled a master regulatory gene because its ectopic expression in bare regions of the body leads to trichomes (Payre *et al.* 1999). However, a more complex genetic architecture seems to underlie variation in the number of larger bristles found in the thorax of adults. In this case, variation around *achaete-scute*, a gene complex required for bristle differentiation, plays a role in controlling bristle number variation across species (Marcellini and Simpson 2006). Interestingly, genetic variation in upstream regulatory factors, whose spatial expression overlap some but not all bristles, is also known to impact bristle number in laboratory mutants (García-Bellido and de Celis 2009). Finally, *shavenbaby* and *scute* genes are also known as input–output genes due to their central "middle of the hourglass" position in regulatory networks (Stern and Orgogozo 2008). These genes respond to the input of multiple upstream protein signals, present at distinct locations in the body, and in turn control the regulation of the same battery of downstream genes, to affect the same output (trichome or bristle development) at each of these body locations. Mutations in the regulatory regions of these genes are thus expected to have minimal pleiotropic effects and to lead to changes in the number of times the network is deployed, and thus to evolution in the number of trichomes or bristles in the bodies of these flies. While this type of regulatory network architecture points to predictable regions in the genome that will readily evolve leading to trait number evolution, *i.e.*, hotspots of evolution, it might represent only one type of architecture among others that are still unexplored. Thus, more systems need to be investigated for a more thorough understanding of the genetic basis underlying the variation of repeated traits in bodies.

One promising system for investigating the genetic basis of serial homolog number evolution is the group of eyespot patterns on the wings of nymphalid butterflies. Eyespots originally appeared on the ventral hindwing in a lineage of nymphalid butterflies, sister to the Danainae, and have subsequently been added to the forewings and dorsal surfaces of both wings (Oliver *et al.* 2012, 2014; Schachat *et al.* 2015). Furthermore, within a single species, eyespot number can vary significantly between individuals or sexes (Brakefield and van Noordwijk 1985; Owen 1993; Tokita *et al.* 2013), allowing for population genetic approaches to identify the underlying genetic basis of such variation. Genes controlling eyespot number variation within a species might also be involved in promoting this type of variation across species.

One of the best model species for studying the genetic basis of eyespot number variation is the nymphalid butterfly *Bicyclus anynana*. This species exhibits natural variation and sexual dimorphism in eyespot number on the dorsal hindwing surface, which plays a possible role in mate choice (Westerman *et al.* 2014). The observed variation consists of males averaging 0.75 dorsal hindwing eyespots, with a range of 0–3, and females averaging 1.5 dorsal hindwing eyespots, with a range of 0–5 (Westerman *et al.* 2014) (Figure 1). Dorsal hindwing spot number (DHSN) variation is positively correlated with butterfly size and, in general, spots are added sequentially on the wing: Cu1, M3, then either Cu2 or M2, and, rarely, Pc (wing sectors/spot locations, as described by the Nymphalid Groundplan) (Nijhout 1991). Laboratory populations of this species also display a series of mutant variants that affect eyespot number on other wing surfaces. Genetic and developmental studies on eyespot number variation in this species suggest the existence of at least two different underlying molecular mechanisms. Spontaneous mutants such as Spotty (Brakefield and French 1993; Monteiro *et al.* 1997, 2013), Missing (Monteiro *et al.* 2007), P- and A- (Beldade *et al.* 2008), or X-ray-induced mutations such as 3 + 4 (Monteiro *et al.* 2003) segregate as single Mendelian alleles, and cause discrete and obvious changes in eyespot number, or affect the size of very specific eyespots. On the other hand, multiple alleles of small effect likely regulate the presence or absence of small eyespots that sometimes appear between the typical two eyespots on the forewing, or on the most posterior wing sector of the ventral hindwing. This type of eyespot number variation is positively correlated with eyespot size variation, responds readily to artificial selection on eyespot size (Holloway *et al.* 1993; Monteiro *et al.* 1994; Beldade and Brakefield 2003), and is likely under the regulation of a threshold-type mechanism (Brakefield and van Noordwijk 1985).

Interestingly, eyespot number variation within *B. anynana* can involve changes to single eyespots or to several eyespots at a time, on one or both wing surfaces. For instance, Spotty introduces two eyespots on the dorsal and ventral surfaces of the forewing, whereas A- and P- primarily reduce the size of the single anterior (A-) or the posterior (P-) eyespot of the dorsal surface exclusively, without affecting eyespot size or number on the ventral surface. The genetic basis for these differences is still unknown.

Recently, the gene *apterousA* (*apA*) was shown to regulate wing pattern differences between dorsal and ventral surfaces in *B. anynana*, including differences in eyespot number (Prakash and Monteiro 2018). This gene is expressed exclusively on the dorsal wing surfaces and its mutation via clustered regularly interspaced short palindromic repeats-Cas9 led to dorsal wing surfaces acquiring a ventral identity, which included additional eyespots. This study indicated that *apA* is a repressor of eyespots on the dorsal surface. However, *B. anynana* has eyespots on dorsal wing surfaces, and their presence and variation in number appears to be correlated with variation in the number of small circular patches, positioned at future eyespot centers, lacking *apA* expression (Prakash and Monteiro 2018). This suggests that genetic variation at loci that modulate the expression of *apA* in eyespot centers on the dorsal surface, or genetic variation in regulatory regions of *apA* itself, might be involved in regulating eyespot number specifically on the dorsal surfaces of wings.

The genetic architecture of eyespot number variation in any butterfly species remains unknown. Here, we examine the genetic basis of DHSN variation in *B. anynana*. We carried out two sets of experiments. We first estimated the heritability for this trait by breeding multiple butterfly families differing in eyespot number and regressing eyespot number of offspring on midparent values. Then we estimated the number and identity of independent genetic loci that are contributing to variation in this trait by performing a genome-wide association study with restriction site-associated DNA sequencing (RAD-seq) from multiple individuals varying in number of eyespots sampled across a freely breeding laboratory population.

Materials and Methods

Study organism

B. anynana is a Nymphalid butterfly common to subtropical Africa for which a colony has been maintained in the laboratory since 1988. All *B. anynana* butterflies used in this study were collected from a colony established in New Haven, CT (Yale University), composed of an admixed population of numerous generations of freely breeding individuals with variable DHSN phenotypes. Individuals from this colony originated from an artificial colony established in Leiden University, from numerous gravid females collected in Malawi in 1988. These laboratory populations have been maintained at relatively high population sizes and females preferentially avoid mating with inbred males (van Bergen *et al.* 2013), which facilitates the relatively high levels of heterozygosity detected several years postestablishment (Saccheri and Bruford 1993; Saccheri *et al.* 1996). The colony was kept in controlled conditions of 12-hr light/dark cycles, 80% relative humidity, and a temperature of 27°. Larvae were fed on corn plants and adult butterflies on mashed banana, as described in previous publications (Westerman *et al.* 2014).

Heritability of DHSN

We examined the DHSNs of all offspring from 18 separately reared families whose parents differed in eyespot number: six families where both parents had a DHSN of zero (0F × 0M); six where both parents had a DHSN of one (1F × 1M); and six where both parents had a DHSN of two (2F × 2M). All generations were reared in the conditions described above. We ensured virginity of the females by separating the butterflies in the parental generation into sex-specific cages on the day of eclosion. All families were started within 5 days of each other using adults with ages ranging 1–3-days old. While initial adult age ranged from 1 to 3 days, the average age was consistent across the three DHSN treatments (ANOVA, $n = 18$, d.f. $= 2$, $F = 0.8266$, $P = 0.4565$). Each breeding pair was placed in a cylindrical hanging net cage of 30-cm diameter × 40-cm height, with food (banana slices), water, and a young corn plant on which to lay eggs. When corn plants were covered with eggs, they were placed in family-specific mesh sleeve cages for larval growth. Females were given new plants on which to lay eggs until they died. Pupae and prepupae were removed from the sleeve cages and placed in family-specific cylindrical hanging net cages for eclosion. The cages were checked daily for newly emerged butterflies. On the day of eclosion, DHSN was recorded for each offspring. DHSN was calculated for each individual by averaging the number of dorsal hindwing eyespots on the left and right wing, which allowed for intermediate values when the wings were asymmetric. Narrow-sense heritability (h^2) was calculated by regressing offspring on midparent values and correcting the heritability estimate for assortative mating. This corrected heritability value h_0^2 was calculated using correlations between phenotypic values of mated pairs (r) to calculate the correlation between breeding values of mates ($m = rh^2$), with $h_0^2 = h^2[(1\text{-}m)/(1 + mh^2)]$ as described in Falconer and Mackay (1996). Estimates were obtained for the pooled offspring data as well as for separate regressions of female and male offspring data on midparent values. Sex-specific heritabilities were calculated using the correction for unequal variances in the two sexes (for example, regression of daughter on father, adjusted regression $h' = b\sigma_m/\sigma_f$ with σ_m/σ_f the ratio of phenotypic standard deviations of males to females) (Falconer and Mackay 1996). Given the known effect of sex on DHSN (Westerman *et al.* 2014), we then also tested for an interaction of parental phenotype and offspring sex on offspring phenotype using a general linear model, with sex, parental phenotype, and sex*parental phenotype as fixed variables.

Sample collection and phenotype determination for genomic association study

To identify regions in the genome that are associated with DHSN variation, we collected and sequenced a total of 30 individuals from the previously described laboratory population. Fifteen individuals contained no eyespots (absence), and 15 contained two or more eyespots (presence) (Supplemental Material, Table S1). Both groups contained an assorted number of male and female individuals. Wings of the collected individuals were removed and the bodies were preserved in ethanol for DNA extraction.

RAD library preparation and sequencing

Genomic DNA of the preserved bodies was extracted using a DNeasy Blood & Tissue Kit (QIAGEN, Valencia, CA), with an additional RNase digestion to remove RNA from the extracted nucleic acid samples. The quality and concentration of the extracted DNA was verified using gel electrophoresis and a Qubit 2.0 fluorometer (Life Technologies). Extracted genomic DNA was used to prepare Illumina RAD-seq libraries based on previously described protocols (Baird *et al.* 2008; Etter *et al.* 2011). DNA was digested with the frequent cutting enzyme *Pst*I and ligated to P1 adapters containing a unique 5-bp barcode. Samples were pooled and sheared using a Covaris M220 (Covaris) instrument, and size selected for inserts between 300 and 500 bp in length. After end repair and

P2 adapter ligation, the library was amplified by PCR. The pooled library was then sequenced utilizing a single lane of an Illumina HiSeq2000 100-bp paired-end module.

Read quality and filtering

Following sequencing of a RAD-seq library composed of 30 *B. anynana* individuals, we obtained 127 million paired-end reads 100 bp in length. The raw RAD reads were demultiplexed using the *Stacks v1.42* (Catchen *et al.* 2011, 2013) *process_radtags* pipeline, and reads with low quality and/or ambiguous barcodes were discarded. Further, we removed Illumina adapter sequences from the reads and trimmed sequences to 80 bp in length, as suggested from the *FastQC* quality control tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) results. We retained a total of 111 million (86%) and an individual average of 3.7 million ± 1.2 million filtered paired-end reads.

Reference alignment

The 111 million retained read pairs were aligned to the *B. anynana* genome assembly (v1.2) with its corresponding annotation (Nowell *et al.* 2017). This reference assembly is highly fragmented and composed of 10,800 scaffolds with an N50 of 638.3 kb, a total assembly size of 475 Mb, and 22,642 annotated genes. Filtered reads were aligned to the reference genome using BWA v0.7.13 (Li and Durbin 2009) *mem* with default seed lengths, and mismatch and gap scores, but allowing for the marking of shorter split reads as secondary alignments for compatibility with PicardTools v1.123 (https://broadinstitute.github.io/picard/). The resulting alignments were directly converted to BAM files using SamTools v1.12 (H. Li *et al.* 2009) *view*. BAM files were then sorted with SamTools *index*, and filtered for duplicates using PicardTools v1.123 *MarkDuplicates* and processed with *AddOrReplaceReadGroups* for GATK compatibility. In total, we obtained 92.9% read alignment and 79.6% properly mapped read pairs. At each RAD locus, average per-individual sequencing coverage was 18.8X (± 4.3; median: 18.3X).

Variant calling and association mapping

After aligning reads to the reference genome, we used two parallel analytical procedures to guarantee our results were method agnostic. First, the processed and filtered alignments were genotyped using GATK v3.5 (McKenna *et al.* 2010) *UnifiedGenotyper*, with the default call confidence values and only outputting variant sites. The resulting 4,661,849 raw variant sites were then filtered using *VCFTools* v0.1.14 (Danecek *et al.* 2011) *recode* to obtain only calls with a minimum genotype quality of 30 and a minimum genotype depth of 5, present in > 50% of all individuals, a maximum allele number of 2, and a minimum allele frequency of 0.05. This resulted in a filtered VCF (variant call format) containing only 350,121 high-quality biallelic variants.

In parallel, the filtered mapped reads were genotyped using *Stacks* v1.42 (Catchen *et al.* 2011, 2013). Using the *ref_map.pl* wrapper script with default parameters, RAD loci were assembled from the mapped reads using *pstacks*. A catalog of all loci was generated with *cstacks* and samples were matched to this catalog using *sstacks*. The *populations* program was then rerun on this catalog to generate population genetic measures, enabling the calculation of F_{ST} statistics. For a variant to be included in the analysis, it had to be present in both study groups ($P = 2$) in > 75% of individuals ($r = 0.75$) and have a minimum allele frequency of ~0.05. Using *Stacks* we reconstructed 207,752 RAD loci and 673,340 raw SNPs. After filtering, 73,159 RAD loci and 238,786 SNPs were retained. Filtered variants obtained from both data sets were compared to retain only SNPs with support from both genotyping methods. A total of 216,338 SNPs were shared between the two data sets and were used for subsequent comparisons.

To identify areas of the genome associated with our desired phenotype, we scanned the genome using two methods: F_{ST} and association via a relatedness-corrected univariate linear mixed model. F_{ST} for all 216,000 shared variants was calculated using the *Stacks populations* module. Fisher's exact test *P*-value correction was applied to the resulting F_{ST} and analysis of molecular variance (AMOVA) F_{ST} values. Future references to F_{ST} in the manuscript specifically refer to the corrected AMOVA F_{ST} values from this analysis. During this *Stacks* run, we enabled the *–plink* flag to output our 216,000 variant sites into *Plink*'s *.map* and *.ped* format for use in the association analysis. The generated *.map* and *.ped* files were converted to binary *.bed* files using *Plink* v1.90b6.9 (Purcell *et al.* 2007; Clarke *et al.* 2011), in addition to adding phenotype information to all samples. From this, we generated a relatedness matrix of all individuals using *GEMMA* v0.98.1 (Zhou and Stephens 2012, 2014), which was used as one of the parameters of a *GEMMA* univariate linear mixed model. We used the *P*-values of the likelihood ratio tests as the basis of the phenotype-to-genotype association. To correct for multiple testing, we performed permutation over a null distribution generated by independently rerunning the *GEMMA* relatedness calculation and univariate linear mixed model 2500 times, randomly shuffling the phenotype assignments of the individuals in each iteration. For this distribution, composed of the combined *P*-values across all runs, association thresholds were determined by calculating the 95th and 99th quantiles of the distribution, obtaining 95% and 99% C.I.s, respectively.

Selection of candidate loci

To minimize the identification of false genotype-to-phenotype relationships, only areas of the genome displaying both association and F_{ST} outliers were used for further analysis. The use of both metrics simultaneously ensured that the relationships observed were method agnostic. This multimetric approach, including a combination of association and F_{ST} outliers, has been utilized repeatedly to identify genomic regions associated with domestication in both dogs and cats (Axelsson *et al.* 2013; Montague *et al.* 2014), variation in feather coloration in warblers (Brelsford *et al.* 2017), and

the architecture and modularity of wing pattern variation in *Heliconius* butterflies (Nadeau *et al.* 2014; Van Belleghem *et al.* 2017). The relationship between the two metrics was explored by plotting both association and F_{ST} outliers. We explored all annotated genes within a 500-kb window centered around outlier SNPs [as determined by linkage disequilibrium (LD) analysis, discussed in the *Results*], and especially under the strongest associated peak area. To identify direct effects of the outlier SNPs over the genes within these windows, we annotated the variants using *SnpEff* v4.3T (Cingolani *et al.* 2012), building a *de novo* database with the available *B. anynana* reference annotation.

Principal component analysis

To determine the baseline-level genome-wide diversity and divergence among individuals in the present population, we performed principal component analysis (PCA) on the obtained genotypes. Although our sampled individuals originated from a single, freely breeding population, performing this analysis corroborates that the observed genomic diversity lacks potential substructuring that could impact our outlier identification. To do this, we randomly selected a subset of 5000 filtered variants from the *Stacks* catalog and made them into a whitelist, as described by the *Stacks* manual and by Rochette and Catchen (2017). We then ran the *populations* module on this subset of variants with the addition of the *–genepop* export format flag. The resulting genepop file was processed using the *adegenet* v2.1.1 R package (Jombart 2008; Jombart and Ahmed 2011) by converting the genotype calls into a *genind* object, scaling missing data by mean allele frequency, and analyzing via PCA.

Ordering of the B. anynana scaffolds along the Heliconius melpomene genome

The current *B. anynana* reference assembly (Nowell *et al.* 2017) has an N50 of 638.3 kb and is composed of 10,800 unlinked scaffolds. To assess whether associated SNPs on separate *B. anynana* genome scaffolds could be part of the same block of association, we ordered the scaffolds of the *B. anynana* genome along the *H. melpomene* v2 genome assembly (Davey *et al.* 2016). Although *B. anynana* and *H. melpomene* diverged ~80 MYA (Espeland *et al.* 2018) and have different karyotypes ($n = 28$ in *B. anynana vs.* $n = 21$ in *H. melpomene*), the *H. melpomene* genome is the most closely related butterfly genome that has been assembled into highly contiguous chromosome-level scaffolds using pedigree-informed linkage maps. Aligning both genomes provides valuable information to interpret our association analysis. To construct this alignment, we used the alignment tool *promer* from the MUMmer v3.0 software suite (Kurtz *et al.* 2004). *Promer* was used with default settings to search for matches between sequences translated in all six possible reading frames between the *B. anynana* and *H. melpomene* genomes. The obtained alignments were subsequently filtered for a minimum alignment length of 200 bp and a minimum percent identity [%IDY = (matches × 100)/(length of aligned region)] of 90%. These filtered alignments were used to order the *B. anynana* scaffolds according to the order in which they aligned along the *H. melpomene* genome. If a scaffold aligned to multiple locations or chromosomes, priority was given to the position it matched with highest identity. For scaffolds that contained significant associations with hindwing eyespot number, we also retained alignments with a minimum %IDY of 70% and a minimum alignment length of 150 bp to investigate possible fine scale rearrangements between the *B. anynana* and *H. melpomene* genomes.

LD analysis

In addition to ordering the *B. anynana* scaffolds to the *H. melpomene* genome to assess the genomic linkage of SNPs, we calculated LD in our *B. anynana* study population. To calculate LD for genomic SNPs, we phased 216,000 SNPs that were genotyped in all samples using *beagle* v4.1 (Browning and Browning 2007). Estimates of LD were calculated from 100,000 randomly selected SNPs using the *VCFtools* v0.1.14 (Danecek *et al.* 2011) *–hap-r2* function, with a maximum LD window of 5 Mbp and minimum allele frequency cutoff of 0.10. Resulting LD comparisons for genomic SNPs were then plotted in R, where a Loess local regression was calculated and used to determine the genome-wide window size of LD decay. Subsequently, this LD window size was used for the investigation of genes near associated loci. An additional interscaffold LD calculation was also performed across all outlier SNPs using the *VCFtools–interchrom-hap-r2* function. The results from this calculation were plotted as a matrix heat map using both the values across all individual SNPs (Figure S1) and the average r^2 value for each pairwise scaffold comparison (Figure 2C).

Post hoc power analysis

We conducted *post hoc* power analysis to assess the power to detect loci significantly associated with DHSN in our genome-wide association analysis, using *GWAPower* (Feng *et al.* 2011). For this power analysis, we used the population-wide heritability of DHSN determined in our heritability analysis (described in the *Results* below), a sample size of 30, and the filtered number of SNPs (216,338). We then ran the analysis with an LD of 1 between SNP and variant, and a range of covariates (0, 10, 15) that explain 0% or 30% of the variance in our trait of interest (DHSN).

Data availability

The *B. anynana Pst*I RAD-seq data are available via the Genbank BioProject identifier PRJNA509697. All bioinformatic code used for analysis is available at Bitbucket (https://bitbucket.org/angelgr2/bany_dhen_arc2019/). Supplemental material, including genotype VCF files, relatedness matrix, and supplemental figures and tables, are available at figshare: https://doi.org/10.25386/genetics.11766624.

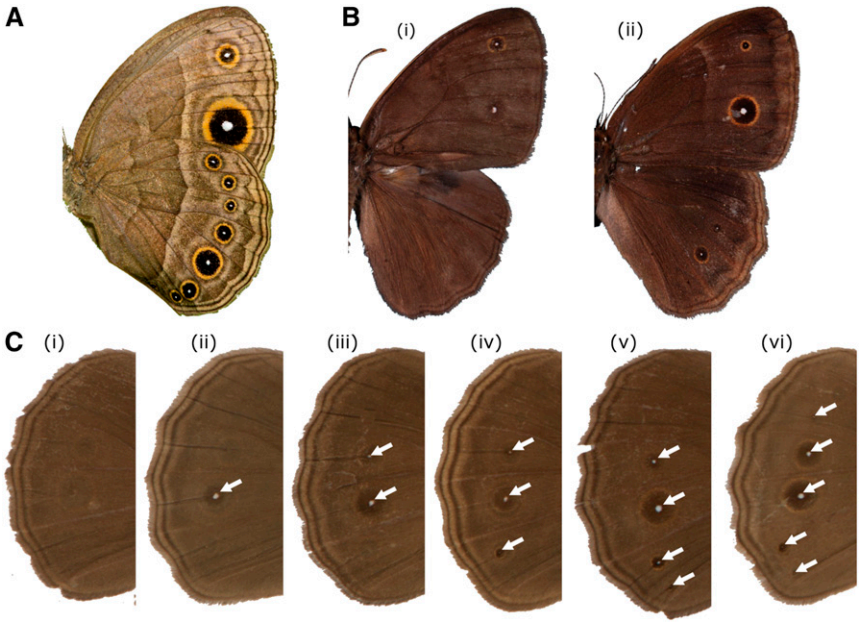

Figure 1 Eyespot pattern and number variation in *B. anynana*. (A) Eyespot pattern on the ventral side of wings: forewing displays two eyespots and hindwing displays seven eyespots. (B) Eyespot pattern on dorsal side of wings. Male (left) displaying two dorsal forewing eyespots and zero dorsal hindwing eyespots, and female displaying two dorsal forewing eyespots and three dorsal hindwing eyespots. (C) Dorsal hindwing spot number variation, ranging from zero to five UV-reflective spots, marked by white arrows (i–vi).

Results

DHSN variation has moderate to high heritability

Zero-spot females were only produced by 0 × 0 families and one 1 × 1 family, and were absent from all 2 × 2 DHSN families. However, zero-spot males were produced by all 0 × 0 families, all but one of the 1 × 1 families, and all but one of the 2 × 2 families. Two-spot females were produced by all but one (a 0 × 0) family, while two-spot males were produced by all 2 × 2 families, but only two 1 × 1 families and one 0 × 0 family (Table 1). These results demonstrate that alleles are sufficiently segregating in our experimental design to perform heritability estimates.

DHSN had a heritability of 0.4442 ± 0.264 for females, 0.5684 ± 0.306 for males, and 0.5029 ± 0.159 when the sexes were pooled. There was no significant interaction of parental phenotype and offspring sex on offspring phenotype (general linear model with sex, parental phenotype, and sex*parental phenotype as parameters, Akaike information criterion = 1498.204, and effect tests: sex χ^2 = 293.361, $P < 0.0001$; parental phenotype χ^2 = 271.56, $P < 0.0001$; and sex* parental phenotype χ^2 = 0.032, $P = 0.8576$).

Genome-wide variation and LD of the study population

To confirm the absence of substructure in our study population, we calculated measures of genetic variation and diversity between the samples displaying the presence of DHSN (pre) and samples with absent DHSN (abs). As expected, the two groups showed very little genome-wide genetic divergence, with a genome-wide F_{ST} equal to 0.0075. The absence of any population substructure between the two sampled phenotype groups was further demonstrated by complete overlap of the two groups in the PCA, as well as little contribution of the phenotype group to the observed variation in the first and second principal components (Figure 2A). Additionally, we observed very similar genome-wide nucleotide diversity in the group displaying the presence of DHSN (pre, π = 0.0090) when compared to the group with absent DHSN (abs, π = 0.0083). Hence, we do not observe any apparent demographic substructuring of the study population that could potentially bias our genetic association analysis. Further, potential relatedness of individuals in our sampling was controlled for by the addition of a relatedness matrix in the generation of the linear mixed model used for testing association.

After calculating genome-wide estimates of LD decay (Figure 2B), we observed a maximum smoothed r^2 value of 0.272 and a halving of r^2 within 470 kb (Figure 2B). This window size suggests that average linkage blocks are ~500 kb in length and that variants within this distance may be in strong LD. When calculating LD between different scaffolds, we observe that outlier SNPs within a scaffold show higher values of linkage than those between scaffolds (Figure 2C). These results suggest that our association peaks (further discussed below) are likely to be independent sections across the genome, and not a contiguous section of the chromosome broken down due to the draft status of the assembly. We have also plotted the distribution of observed *vs.* expected *P*-values as a quantile–quantile plot (Figure 2D). The tight correlation between the observed and expected distribution highlights that our association values do not seem to be inflated by structural attributes of the data, or unknown confounding factors. Additionally, the observed distribution does not appear to be indicative of one or very few major effect loci underlying DHSN, suggesting that the polygenic architecture of our investigated phenotype is likely biologically real.

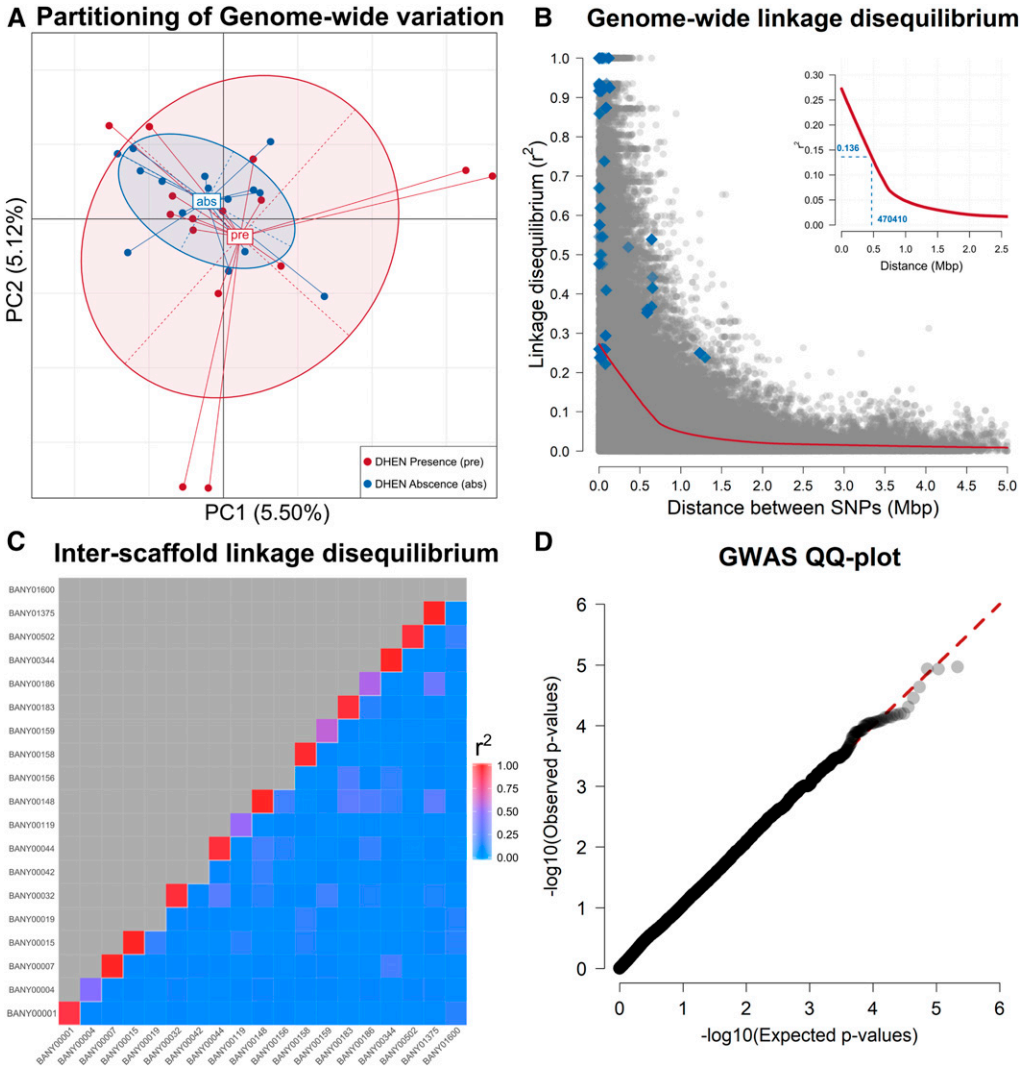

Figure 2 Suitability control for using the laboratory-reared *B. anynana* population for a genome-wide association study. (A) PC analysis of the allelic variation observed in 5000 randomly selected genome-wide SNPs in the study population across both phenotype groups, DHSN presence (*pre*, red) and DHSN absence (*abs*, blue). Ellipses display boundaries of the 95% C.I.s. Little contribution to variation in the PCs and overlap of the variation on both phenotype groups suggests lack of underlying demographic substructuring in the study population. (B) Genome-wide LD in the *B. anynana* study population. Gray dots represent LD values for the pairwise comparison across genome-wide SNPs. Blue diamonds show LD values for SNP pairwise comparisons across outlier SNPs. In red, Loess regression smoothed curve representing LD decay. Insert: zoomed-in LD decay curve, indicating distance at which LD is halved (470, 410 bp) and corresponding r^2 value (0.137). (C) Interscaffold LD for all outlier SNPs. r^2 values show the averaged r^2 for all SNPs in a per-scaffold pairwise comparison. Missing squares represent scaffolds with a single outlier SNP, for which a pairwise LD calculation is not possible. (D) QQ-plot showing relationship between observed and expected association values [−log10(*P*-value)]. Each dot represents the values for a single SNP. Dashed red line shows the 1:1 relationship for these two values. DHSN, dorsal hindwing spot number; LD, linkage disequilibrium; PC, principal component; QQ-plot, quantile–quantile plot.

Association mapping of DHSN variation

After mapping and genotyping RAD loci across the *B. anynana* reference genome, we identified a total of 216,338 SNPs shared between the two different genotyping strategies used (see *Materials and Methods*). After correcting for multiple testing using a permutation over a null distribution, we obtained two distinct association thresholds representing 95% and 99% C.I.s. The first had a significant threshold *P*-value of 0.0005 (association of 3.3047), while for the second the threshold had a *P*-value of 0.0003 (association of 3.5902). Using the 95% C.I., we identified 69 outlier SNPs on 19 different scaffolds. These SNPs, when accounting for the adjacent genomic regions in physical linkage, accounted for 21.2 Mb, or ∼4.46%, of the BANY.1.2 assembly. In the 99% C.I., we identified 33 outlier SNPs across nine different scaffolds, which accounted for 9.75 Mb, or 2.03%, of the total assembly.

Ordering these scaffolds along the contiguously assembled *H. melpomene* genome suggests that the 18 scaffolds with outlier SNPs (using the 95% C.I.) belong to 16 discrete regions in the *B. anynana* genome, whereas the smaller subset of 9 scaffolds that have outlier SNPs in the 99% C.I. belong to 7 different regions in the genome (Figure 4).

Post hoc power analysis

We estimated that our power was between 0.60 and 0.95, depending on the number of covariates (*i.e.*, amount of pleiotropy) for DHSN. Thus, we had sufficient power to detect genes of large effect, but not genes of small effect (assuming a conservative estimate of numbers of covariates). Given our DHSN heritability of 0.5029 when the sexes were pooled, our sample size of 30 for genome sequencing, and the filtered number of SNPs (216,338), we had a power of 0.60 when LD = 1 and there were no covariates. However, if we had 10 covariates that explained 30% of the variance in DHSN,

Table 1 DHSN is heritable

	0 × 0 DHSN families		1 × 1 DHSN families		2 × 2 DHSN families	
	Females	Males	Females	Males	Females	Males
0 DHSN	46	159	2	47	0	14
1 DHSN	55	15	45	36	19	38
2 DHSN	24	2	43	9	67	27
3 DHSN	2	0	4	0	6	0
Average DHSN	0.87	0.17	1.49	0.63	1.87	1.13

Summary DHSN data for offspring from six families each of 0 × 0, 1 × 1, and 2 × 2 DHSN crosses, separated by sex. Offspring with asymmetric DHSN are included in the average DHSN estimate. DHSN, dorsal hindwing spot number.

we would have a power of 0.95, and if we had 15 covariates that explained 30% of the variance in DHSN, our power would be roughly 0.71. Our power decreased by ~12% for every 0.1 decrease in LD between SNP and causative locus.

Candidate gene identification

Using the available Lepbase reference annotation for the BANY.1.2 assembly, we identified the neighboring annotated genes and relative positioning of the 69 outlier SNPs on the 16 genomic regions. We observed a total of 559 annotated genes within our 500-kb windows around the outlier SNPs (Table S2), of which 262 remained when compared to the 33 outlier SNPs in the 99% C.I. Only 21 of these 559 genes contained outlier SNPs in their introns or exons (Table 3 and Table S2). On average, our outlier SNPs were 185 kb (± 144 kb) from their nearest genes. This suggests that a significant portion of our outliers are intergenic and potentially associated with regulatory variants. Only a few SNPs mapped within a gene's coding sequence, of which seven resulted in synonymous and two in nonsynonymous mutations (Table 3 and Table S2). Thus, to identify the most promising candidate genes that may be linked to noncoding associated SNPs and regulating DHSN, we used our current knowledge of butterfly developmental genetics and identified 26 such genes (Figure 4 and Table 2). Among those, 6 candidates were previously associated with eyespot development—*i.e.*, *CDase*, *Dpp*, *Antp*, *Ubx*, *Sal3*, and *Cd63*—and the remaining 20 were linked to eyespot number variation for the first time here (Table 2). The latter included genes related to Notch, Toll, and Ras signaling pathways, wing morphogenesis, dorso/ventral pattern formation, pigment biosynthesis, and a number of molecules with regulatory and signaling functions (Table 2).

Discussion

The use of population genomic analyses, including association mapping, genome-wide association studies, and F_{ST} scans, has been extensively used in natural populations to determine the genomic signatures underlying a number of biological processes such as hybridization, speciation, selection, and ecological adaptation [reviewed in Narum *et al.* (2013), Rellstab *et al.* (2015), and Campbell *et al.* (2018)]. In our work, we applied a population genomics approach to characterize the genetic basis of DHSN variation in the butterfly *B. anynana* via the comparison of genomic diversity between individuals from a single laboratory-maintained, freely breeding population, that display moderate to high narrow-sense DHSN heritability. Like natural populations, we demonstrated that the laboratory population maintains homogenization of the genetic variation across the genome due to a lack of demographic substructuring, and thus provides a powerful system to detect associations between SNPs and the trait of interest (DHSN variation). We also provide evidence that the multiple genomic regions identified in this study are not inflated by structural attributes of either the data or the underlying reference assembly (Figure 2), and that our experimental design provides enough power to identify candidate loci of large effect.

The combined results of our heritability and genome-wide association studies suggest that variation in DHSN in *B. anynana* is a complex trait regulated by multiple loci. Our heritability analysis using families reared in controlled environments demonstrates that DHSN is moderately to highly heritable and has a genetic basis. We were then able to use a genomic approach to determine how many genomic regions and candidate genes are likely to be involved in determining DHSN. Combining a RAD-seq genotyping approach with genome-wide population differentiation and association mapping, we identified up to 16 discrete genomic intervals associated with DHSN variation in *B. anynana* (Figure 3 and Figure 4). Within the 16 genomic regions, we observed a total of 559 annotated genes (Table S2), which we narrowed down to 26 potential candidates (Table 2). Using a higher threshold for the top 1% of associations, a total of 12 candidate genes were identified. Below, we address the possible relationship of these candidates with eyespot development and DHSN variation.

Known eyespot developmental genes associated with DHSN variation

Among our identified candidates, we observed a total of six genes previously implicated in eyespot development in *B. anynana*: *CDase*, *Dpp*, *Antp*, *Ubx*, *Sal3*, and *Cd63* (Brunetti *et al.* 2001; Tong *et al.* 2014; Özsu and Monteiro 2017; Connahs *et al.* 2019; Matsuoka and Monteiro 2019) (Table 2). From these, only *dpp* and *Cd63* remained associated at the highest significant threshold (99% C.I.). Most of these genes were within or near the area of strongest association, and only *Cd63* had an associated SNP within an intron (Figure 4 and Table 3). Neutral ceramidase (*CDase*; BANY.1.2.g00030) is an enzyme involved in the metabolism of sphingolipids that is downregulated in eyespots relative to flanking wing tissue (Özsu and Monteiro 2017). Sphingolipids appear to function in a variety of cell signaling pathways but are still poorly understood (Hannun and Obeid 2017). Decapentaplegic (*Dpp*; BANY.1.2.g00653) is a known developmental gene expressed in dynamic patterns during eyespot center formation during the larval stage (Monteiro *et al.* 2006; Connahs

Table 2 DHSN candidate genes

BANY.1.2 gene ID	Gene name	Gene description	BANY.1.2 scaffold ID	Distance to nearest outlier SNP (bp)	Previously characterized eyespot gene	Function	Biological processes	Recovered in 99% C.I.?
BANY.1.2.g00030	*CDase*	Neutral ceramidase	BANY00001	8,479	Yes: Özsu and Monteiro (2017)	*N*-acylsphingosine amidohydrolase activity	Sphingolipid metabolic process; ceramidase activity	No
BANY.1.2.g00624	*gpr180*	Integral membrane G protein-coupled receptor	BANY00004	6,319	No		Signal transduction	No
BANY.1.2.g00636	*Dach1*	Dachshund homolog 1	BANY00004	267,821	No	DNA binding	Transcription factor complex; DNA-binding transcription repressor activity, RNA polymerase II-specific; cell population proliferation; Leg disc proximal–distal pattern formation	Yes
BANY.1.2.g00653	*dpp*	Protein decapentaplegic	BANY00004	28,669	Yes: Connahs *et al.* (2019)	Morphogene activity	Cell fate specification, anterior/posterior pattern formation, imaginal disc development, imaginal disc pattern formation; developmental pigmentation; dorsal/ventral axis specification	Yes
BANY.1.2.g00658	*dlg1*	Disks large 1 tumor suppressor protein	BANY00004	61,933	No, but suggested to regulate the Notch signaling pathway, which is associated with eyespot development: Reed and Serfas (2004)	Epidermal growth factor receptor binding	Anterior/posterior axis specification; morphogenesis of larval imaginal disc epithelium	Yes
BANY.1.2.g00681	*numb*	Protein numb	BANY00004	222,545	No, but suggested to regulate the Notch signaling pathway, which is associated with eyespot development: Reed and Serfas (2004)	Notch binding	Asymmetric cell division; regulation of Notch signaling pathway	Yes
BANY.1.2.g01110	*ZC3H10*	Zinc finger CCCH domain-containing protein 10	BANY00007	0 [coding sequence nonsynonymous (Ile142→Val)]	No	Metal ion binding, RNA binding	Post-transcriptional regulation of gene expression	No

(continued)

Table 2, *continued*

BANY.1.2 gene ID	Gene name	Gene description	BANY.1.2 scaffold ID	Distance to nearest outlier SNP (bp)	Previously characterized eyespot gene	Function	Biological processes	Recovered in 99% C.I.?
BANY.1.2.g02164	*Taf2*	Transcription initiation factor TFIID subunit 2	BANY00015	16,341	No, but suggested to regulate the Notch signaling pathway, which is associated with eyespot development: Reed and Serfas (2004)	Chromatin binding	DNA template transcription, initiation; positive regulation of transcription of Notch receptor target	No
BANY.1.2.g02571	*Antp*	Homeotic protein antennapedia	BANY00019	276,092	Yes: Saenko *et al.* (2011), Matsuoka and Monteiro (2019)	DNA binding	Anterior/posterior axis specification; specification of segmental identity	No
BANY.1.2.g02579	*Ubx*	Homeotic protein ultrabithorax	BANY00019	88,613	Yes: Weatherbee *et al.* (1999), Tomoyasu *et al.* (2005), Tong *et al.* (2014), Matsuoka and Monteiro (2019)	DNA binding	Multicellular organism development; regulation of transcription; anterior/posterior pattern specification	No
BANY.1.2.g02585	*abd-A*	Homeobox protein abdominal-A homolog	BANY00019	401,028	No	DNA binding	Multicellular organism development; regulation of transcription, anterior/posterior pattern specification	No
BANY.1.2.g04718	*TRAIP*	E3 ubiquitin-protein ligase TRAIP	BANY00042	4,838	No	—	Apoptotic process; cell population proliferation; signal transduction	Yes
BANY.1.2.g04910	*spz*	Protein spaetzle	BANY00044	18,317	No, but Toll has been implicated in eyespot development: Özsu and Monteiro (2017)	Morphogene activity, Toll binding	Regulation of Toll signaling pathway	Yes
BANY.1.2.g09546	*Sall3*	Sal-like protein 3; Spalt	BANY00119	135,802	Yes: Brunetti *et al.* (2001), Monteiro *et al.* (2006)	Nucleic acid binding	Limb morphogenesis	No
BANY.1.2.g10814	*rergl*	Ras-related and estrogen-regulated growth inhibitor-like protein	BANY00148	214,847	No, but Ras signaling and 20E signaling have been implicated in eyespot development: (Özsu and Monteiro 2017; Bhardwaj *et al.* 2018)	GTP binding	Signal transduction	No

(continued)

Table 2, *continued*

BANY.1.2 gene ID	Gene name	Gene description	BANY.1.2 scaffold ID	Distance to nearest outlier SNP (bp)	Previously characterized eyespot gene	Function	Biological processes	Recovered in 99% C.I.?
BANY.1.2.g10819	*Ero1L*	Ero1-like protein	BANY00148	156,666	No, but suggested to regulate the Notch signaling pathway, which is associated with eyespot development: Reed and Serfas (2004)	Oxidoreductase activity	Chaperone cofactor-dependent protein refolding; regulation of Notch signaling pathway	No
BANY.1.2.g11151	*mbo*	Nuclear pore complex protein Nup88	BANY00156	0 [coding sequence nonsynonymous (Leu584→Pro)]	No	Chromatin binding	Messenger RNA export from nucleus; protein export from nucleus	Yes
BANY.1.2.g11298	*Eif4g3*	Eukaryotic translation initiation factor 4 γ 3	BANY00158	39,941	No	RNA binding	Positive regulation of meiosis I; positive regulation of protein phosphorylation	Yes
BANY.1.2.g11326	*msh*	Muscle segmentation homeobox	BANY00159	246,092	No, but involved in dorsal/ventral pattern formation in *Drosophila* wing: Milán *et al.* (2001)	DNA binding	Dorsal/ventral pattern formation; ectoderm development; embryonic morphogenesis; imaginal disc-derived wing morphogenesis; wing and notum subfield formation; wing disc pattern formation	Yes
BANY.1.2.g12183	*hd*	Protein downstream neighbor of son homolog (Protein humpty dumpty)	BANY00183	3,204	No	—	Cell division; cell population proliferation; multicellular organism development; nuclear DNA replication	No
BANY.1.2.g12184	*Mi-2*	Chromodomain-helicase-DNA-binding protein Mi-2 homolog	BANY00183	0 (intronic, 4 SNPs)	No	ATP binding, chromatin binding, DNA binding	Chromosome decondensation; chromosome organization	No
BANY.1.2.g12275	*Cd63*	CD63 antigen	BANY00186	0 (intronic, 1 SNP)	Yes: Özsu and Monteiro (2017)	Protein-containing complex binding	Cell–matrix adhesion; cell migration; cell surface receptor signaling pathway; epithelial cell differentiation; pigment granule maturation	Yes

(continued)

Table 2, *continued*

BANY.1.2 gene ID	Gene name	Gene description	BANY.1.2 scaffold ID	Distance to nearest outlier SNP (bp)	Previously characterized eyespot gene	Function	Biological processes	Recovered in 99% C.I.?
BANY.1.2.g16720	*14-3-3* ε	14-3-3 protein ε	BANY00344	63,487	No, but this protein is a component of the Ras signaling pathway, which has been implicated in eyespot development: Özsu and Monteiro (2017). Also, wing disc dorso/ventral pattern formation in *Drosophila* Bejarano, *et al.* (2008)	Transcription factor binding; protein domain-specific binding	Imaginal disc development; wing disc dorsal/ventral pattern formation; regulation of growth; Ras signaling	Yes
BANY.1.2.g16724	*ct*	Homeobox protein cut	BANY00344	0 (coding sequence, 1 SNP)	No, but this protein is a regulator of Notch signaling, which has been associated with eyespot development: Reed and Serfas (2004)	DNA binding	Imaginal disc-derived wing margin morphogenesis; formation of a compartment boundary; negative regulation of Notch signaling pathway	Yes
BANY.1.2.g19639	*Roc1a*	RING-box protein 1A (Regulator of cullins 1a)	BANY00502	3,407	No, but a regulator of Wnt signaling, and Wg has been implicated functionally in eyespot size regulation: Özsu *et al.* (2017)	Zinc ion binding	Cell population proliferation; lipid storage; negative regulation of canonical Wnt signaling pathway	No
BANY.1.2.g25219	*AnxB9*	Annexin B9	BANY01600	1,408	No, but involved in wing disc dorso/ventral pattern formation: Bejarano, *et al.* (2008)	Actin binding	Wing disc dorsal/ventral pattern formation	No

BANY.1.2 gene ID, gene name, gene description, and BANY.1.2 scaffold ID all refer to the *B. anynana* v1.2 genome assembly and annotation (Nowell *et al.* 2017). Distance to nearest SNP refers to the distance (bp) of each of the candidate genes to one of our identified SNPs. Genes with SNPs inside their sequence have additional information describing the SNP annotation. Eyespot regulatory networks describes direct or indirect evidence on the role of each of our identified candidates, within the gene networks responsible for eyespot development, in addition to suggested functions within this network. Genes were identified based on two possible association thresholds, as described in the *Materials and Methods*. ID, identifier; RING, Really Interesting New Gene.

et al. 2019). *Dpp* is likely involved in the differentiation of the eyespot centers in *B. anynana* via a reaction–diffusion mechanism (Connahs *et al.* 2019). *Dpp* messenger RNA is expressed in regions around the developing eyespots in mid to late larval-stage wings (Connahs *et al.* 2019) in anticolocalized patterns to Armadillo, a transcription factor effector of the Wingless signaling pathway. Armadillo is expressed in the actual eyespot centers in late larval stages (Connahs *et al.* 2019), as is Antennapedia (*Antp*; BANY.1.2.g02571), a hox gene (Saenko *et al.* 2011; Tong *et al.* 2014; Özsu and Monteiro 2017). *Antp* is required for forewing eyespot development, and eyespot size and center differentiation in the hindwings of *B. anynana* (Matsuoka and Monteiro 2019). *Antp* is also involved in the differentiation of the thoracic limbs in *Bombyx* moths (Chen *et al.* 2013).

Ultrabithorax (*Ubx*; BANY.1.2.g02579) is another hox gene that gives insect hindwings a different identity from forewings (Weatherbee *et al.* 1999; Tong *et al.* 2014). In *B. anynana*, this gene is expressed across the whole hindwing, a conserved expression pattern observed across insects, but has additional, stronger expression in the eyespot centers of the hindwings only, something that is not seen in other butterflies with eyespots such as *Junonia coenia* (Tong *et al.* 2014). Overexpression of *Ubx* leads to

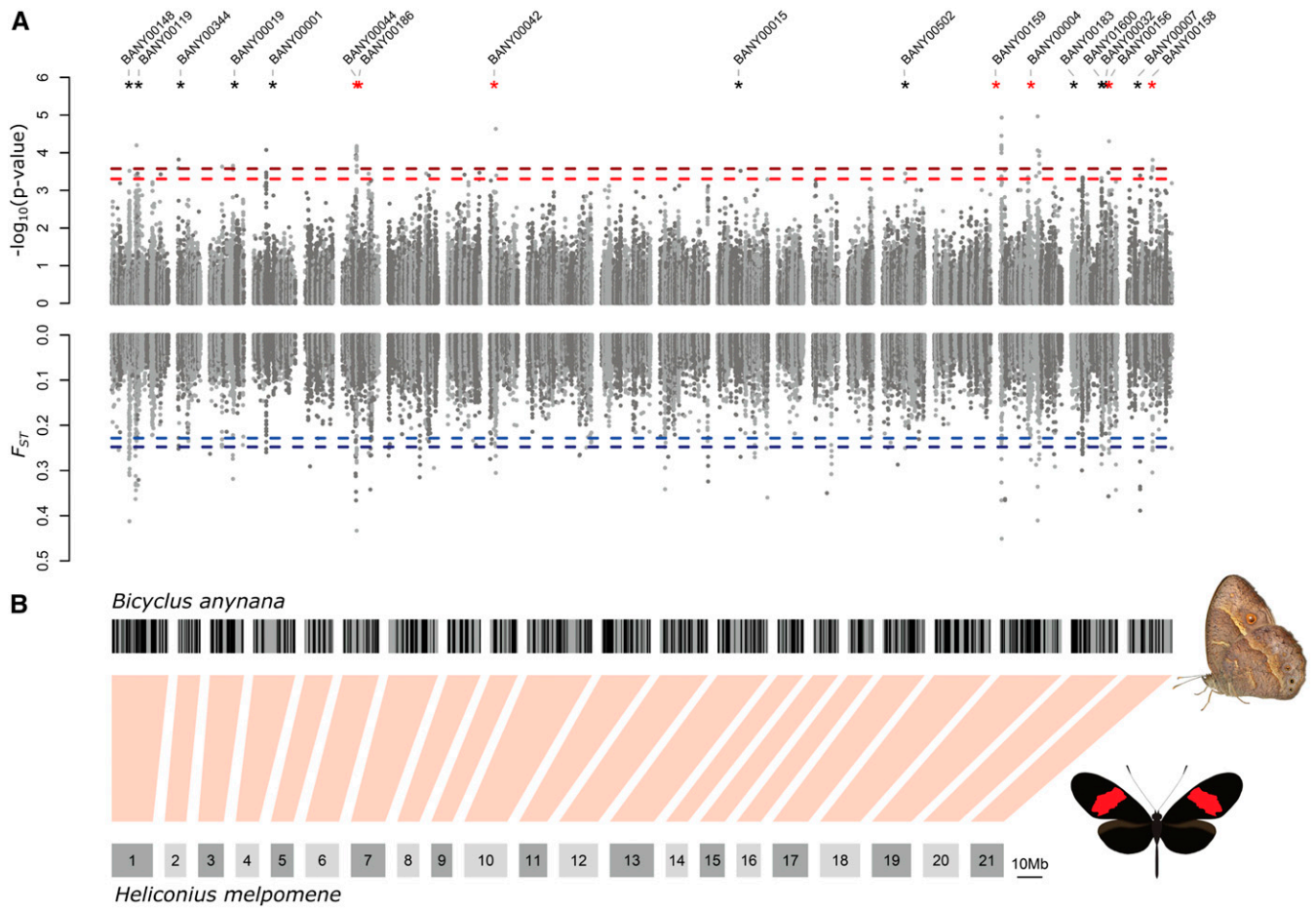

Figure 3 Genome-wide association with dorsal hindwing eyespot number. (A) Plots show genomic association to dorsal hindwing eyespot number (top) and F_{ST} between individuals with a different number of dorsal hindwing eyespots (bottom). Each dot represents a single SNP. Dashed lines represent the threshold for detecting a significant genome-wide association (top, in red) and F_{ST} (bottom, in blue) representing the top 5% and 1% of values, corrected for multiple testing. Scaffolds containing both significant association and F_{ST} outliers are marked with asterisks (black for 95% C.I. and red for 99% C.I.). (B) Genomic scaffolds from the *B. anynana* BANY.1.2 genome assembly are arranged along the 21 chromosomes of the *H. melpomene* v2 assembly. For ordering the *B. anynana* scaffolds along the *H. melpomene* genome, only matches with a minimum percentage of identity of 90% and a minimum alignment length of 200 bp were used. If scaffolds matched multiple *H. melpomene* chromosomes, the scaffold was positioned along the chromosome to which it had the most matches. Using this strategy, 76.7% of the *B. anynana* genome scaffolds were aligned to the *H. melpomene* genome.

eyespot size reductions in both fore- and hindwings of *B. anynana* (Tong *et al.* 2014), whereas an absence of *Ubx* in clones of cells in the hindwings of *J. coenia*—via a spontaneous unknown mutation (Weatherbee *et al.* 1999)—leads to eyespot enlargements (of Cu1 hindwing eyespots to sizes that match Cu1 forewing eyespot sizes). Furthermore, *Ubx* crispants in *B. anynana* lose M3 hindwing eyespots (Matsuoka and Monteiro 2019). These results suggest both a repressive as well as an activating role of *Ubx* on eyespots of the hindwing, dependent on eyespot position. Genetic variation at *Ubx* might contribute to eyespot number variation via a threshold-like mechanism acting on eyespot size. Sal-like protein 3 (SALL3), also known as *Spalt* (*Sal*; BANY.1.2.g09546), is a transcription factor that is expressed in eyespot and spot centers in a variety of butterflies including *B. anynana* (Brunetti *et al.* 2001; Oliver *et al.* 2012; Shirai *et al.* 2012), and was recently shown to be required for eyespot development in both *Vanessa cardui* and *J. coenia* butterflies (Zhang and Reed 2016). Lastly, the trespanin *Cd63* (*Cd63*; BANY.1.2.g12275) is a protein associated with exosomes, which in turn have been associated with wingless signaling in *Drosophila* (Beckett *et al.* 2013). This gene was the only candidate that had been previously characterized as upregulated in eyespots (Özsu and Monteiro 2017) to be found here to contain an association within its gene module, and more specifically a SNP in its second intron (Figure 4 and Table 3). Genetic variation either linked with the protein-coding sequences of these genes, or more likely with their regulatory regions, is likely affecting eyespot number variation in hindwings.

Novel candidate genes associated with DHSN variation

In addition to the above eyespot-associated genes, we also identified a group of 20 candidates that have not been

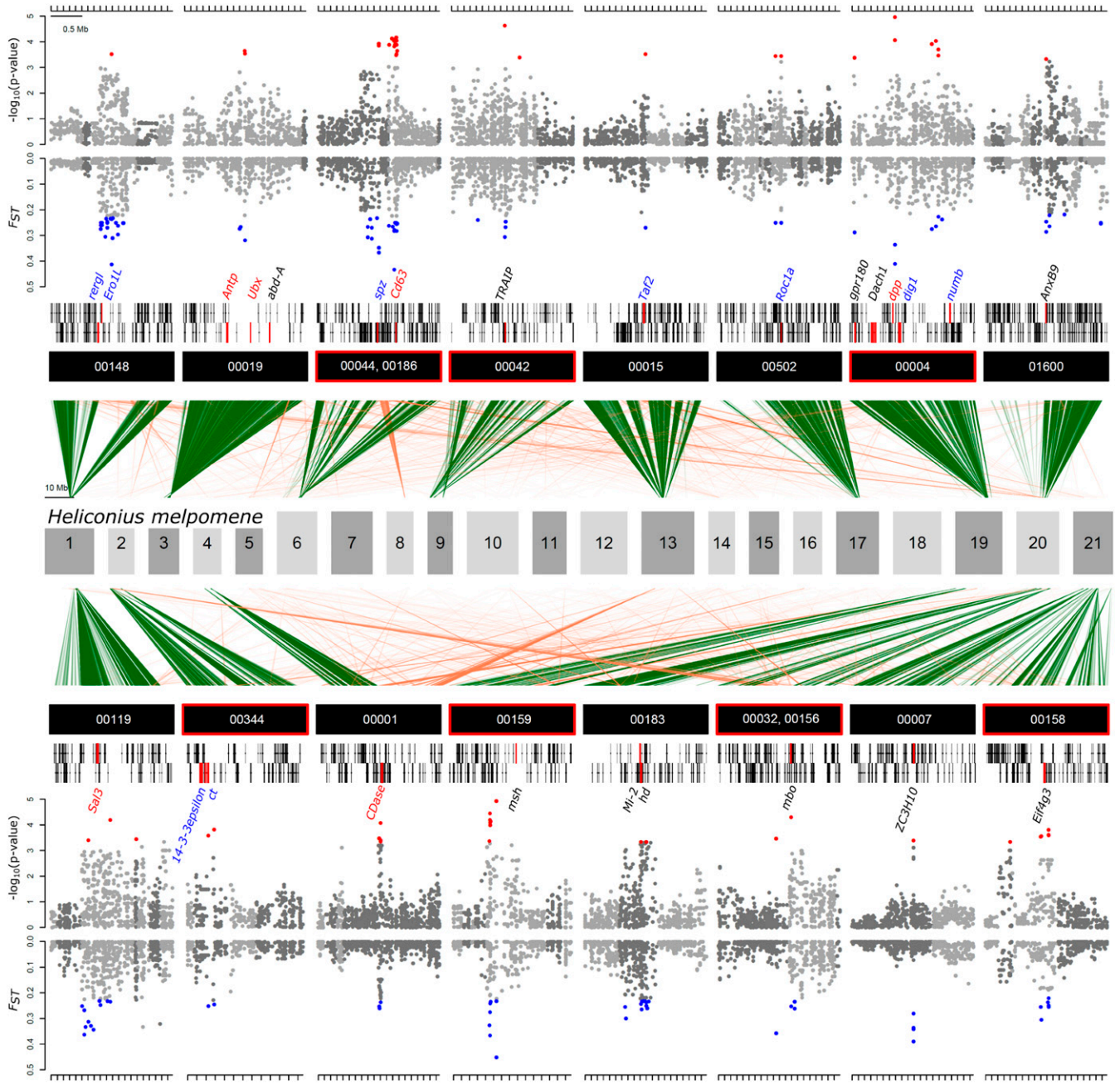

Figure 4 Zoomed-in image of putative genomic regions underlying dorsal hindwing eyespot number variation. Plots show genomic association to dorsal hindwing eyespot number (top) and F_{ST} (bottom) between individuals with different dorsal hindwing eyespot numbers for scaffolds with significant outliers (red for association and blue for F_{ST}). Each dot represents a single SNP. Light and dark gray dots correspond to alternating BANY.1.2 scaffolds mapping adjacently to the *H. melpomene* genome. Only names of scaffolds with significant outliers are listed. Scaffolds with SNPs in the top 1% of associations are highlighted with a red box. Green and orange lines show matches of the *B. anynana* scaffolds (minimum percentage of identity of 70% and a minimum alignment length of 150 bp) to the *H. melpomene* v2 assembly. Green lines represent the most frequent matches of the scaffold to an *H. melpomene* chromosome, whereas orange lines represent matches to a different *H. melpomene* chromosome. Vertical black rectangles represent gene models with candidate genes highlighted in red. Names of genes that overlap or are close to an outlier SNP, and that are reported in the literature to affect eyespot development, are indicated in red. Names of genes that are within the same pathway as genes involved in eyespot development are indicated in blue. Names of other candidates are indicated in black.

associated with eyespot development before. However, many of these genes are part of signaling pathways previously associated with eyespot development such as the Notch, Ras, and Toll signaling pathways (Table 2) (Özsu and Monteiro 2017), and others with roles in wing morphogenesis, dorso/ventral pattern formation, and pigment biosynthesis.

A total of five genes (*dlg1*, *numb*, *Taf2*, *Ero1L*, and *ct*) implicated in our study (Table 2) are known to interact with the *Notch* eyespot-associated gene (Reed and Serfas 2004).

Table 3 Genes containing outlier SNPs

BANY.1.2 gene ID	Gene name	Gene description	BANY.1.2 scaffold ID	Number of variants	Annotated variants	Function	Biological process	Recovered in 99% C.I.
BANY.1.2.g00028	*STRADA*	STE20-related kinase adapter protein α	BANY00001	1	Intronic (intron 1)	ATP binding; protein kinase	MAPK cascade	No
BANY.1.2.g00029	*TKFC*	Triokinase/FMN cyclase	BANY00001	1	Intronic (intron 1)	ATP binding; metal ion binding	Innate immune response; MDA-5 signaling pathway	No
BANY.1.2.g00671	*CG4341*	Protein *O*-mannosyl-transferase	BANY00004	1	Intronic (intron 1)	Mannosyltrans-ferase activity	—	Yes
BANY.1.2.g01110 ❋	*ZC3H10*	Zinc finger CCCH domain-containing protein 10	BANY00007	2	Synonymous (exon 2); Nonsynony-mous (exon 2: Ile142→Val)	Metal ion bind-ing, RNA binding	Post-transcriptional regulation of gene expression	No
BANY.1.2.g02168	*kh*	Kynurenine 3-monooxygenase	BANY00015	2	Two intronic (intron 4)	FAD binding; NAD(P)H oxidase	NAD metabolism	No
BANY.1.2.g04716	Unknown	Unknown	BANY00042	1	Synonymous (exon 2)	—	—	Yes
BANY.1.2.g04911	Unknown	Unknown	BANY00044	3	Three intronic (intron 1)	—	—	Yes
BANY.1.2.g11151 ❋	*mbo*	Nuclear pore complex protein *Nup88*	BANY00156	1	Nonsynony-mous (exon 14: Leu584→ Pro) (1)	Chromatin binding	Messenger RNA export from nucleus; protein export from nucleus	Yes
BANY.1.2.g11296	*APBB1IP*	APBB1-interacting protein 1	BANY00158	1	Intronic (intron 2)	—	Signal transduction	Yes
BANY.1.2.g11300	*Tret1*	Facilitated trehalose transporter *Tret1*	BANY00158	4	Three intronic (intron 9); two synony-mous (exon 9)	Trehalose trans-membrane transporter activity	Trehalose transport	Yes
BANY.1.2.g11312	*idhb-1*	Isocitric dehydroge-nase subunit β	BANY00159	3	Intronic (intron 1); two synonymous (exon 5)	Isocitrate dehy-drogenase activity; magnesium ion binding; NAD binding	Tricarboxylic acid cycle	Yes
BANY.1.2.g12184 ❋	*Mi-2*	Chromodomain-helicase-DNA-binding protein *Mi-2* homolog	BANY00183	7	4 intronic (intron 8); splice region (intron 8); synonymous (exon 8); synonymous (exon 15)	ATP binding, chromatin binding, DNA binding	Chromosome decondensation; chromosome organization	No
BANY.1.2.g12186	Unknown	Unknown	BANY00183	3	Three intronic (intron 4)	—	—	No
BANY.1.2.g12187	*spt16*	FACT complex subunit *spt16*	BANY00183	1	Synonymous (exon 12)	Chromatin binding	Transcription elon-gation; DNA replication	No
BANY.1.2.g12189	*TRIM37*	E3 ubiquitin-protein ligase	BANY00183	6	Five intronic (intron 7); synonymous (exon 7)	Chromatin binding	Regulation of RNA polymerase binding	No
BANY.1.2.g12272	*PPIL2*	RING-type E3 ubiquitin-protein ligase	BANY00186	1	Intronic (intron 2)	Ubiquitin protein ligase activity	Protein folding; protein localization	Yes

(continued)

Table 3, *continued*

BANY.1.2 gene ID	Gene name	Gene description	BANY.1.2 scaffold ID	Number of variants	Annotated variants	Function	Biological process	Recovered in 99% C.I.
BANY.1.2.g12275	*Cd63*	CD63 antigen	BANY00186	1	Intronic (intron 2)	Protein-containing complex binding	Cell–matrix adhesion; cell migration; cell surface receptor signaling pathway; epithelial cell differentiation; pigment granule maturation	Yes
BANY.1.2.g12279 *	*Cd53*	Leukocyte surface antigen CD53	BANY00186	1	Synonymous (exon 2)	—	Cell surface receptor signaling pathway	Yes
BANY.1.2.g16724 *	*ct*	Homeobox protein cut	BANY00344	1	Synonymous (exon 2)	DNA binding	Imaginal disc-derived wing margin morphogenesis; formation of a compartment boundary; negative regulation of Notch signaling pathway	Yes
BANY.1.2.g19628	*LanB1*	Laminin subunit β-1	BANY00502	1	Intronic (intron 8)	—	Substrate adhesion; cell migration	No
BANY.1.2.g19640	Unknown	Unknown	BANY00502	2	2 Intronic (Intron 7)	—	—	No

BANY.1.2 gene ID, gene name, gene description, BANY.1.2 scaffold ID, and function/biological process all refer to the *B. anynana* v1.2 genome assembly and annotation (Nowell *et al.* 2017). Genes were identified based on two possible association thresholds, as described in the *Materials and Methods*. The number of variants refers to the number of outlier SNPs present inside the gene. Each individual variant is annotated, and its position within the gene's sequence and impact on amino acid changes are described. Asterisks in gene IDs represent genes that overlap with our dorsal hindwing spot number candidates in Table 2. FACT, facilitates chromatin transcription; ID, identifier; RING, Really Interesting New Gene.

Disks large one tumor suppressor (*dlg1*; BANY.1.2.g00658), numb (*numb*; BANY.1.2.g00681), Ero1-like protein (*Ero1L*; BANY.1.2.g10819), and ct (*ct*; BANY.1.2.g16724) are known to interact with and regulate the Notch signaling pathway in *Drosophila melanogaster* (Cheah *et al.* 2000; Tien *et al.* 2008; Q. Li *et al.* 2009). The gene *ct* is the only one of these genes with an associated SNP causing a synonymous mutation. The existence and role of these interactions are unknown in *B. anynana*, as is the role of the Notch receptor itself. However, the Notch receptor has a dynamic pattern of expression (Reed and Serfas 2004) that is very similar to that of *Distal-less*, a gene that has recently been implicated in setting up the eyespot centers, likely via a reaction–diffusion mechanism (Connahs *et al.* 2019). Genetic variation at these three genes could be interacting with the eyespot differentiation process through unknown mechanisms.

We also identified new members of the Ras and Toll signaling pathways that have previously been associated with eyespot development (Özsu and Monteiro 2017). Protein14-3-3 ε (BANY.1.2.g16720) is a member of the Ras signaling pathway (Chang and Rubin 1997), and Ras-related and estrogen-regulated growth inhibitor-like protein (*rergl*; BANY.1.2.g10814) is a Ras-related superfamily gene with unique characteristics (Finlin *et al.* 2001). Protein spaetzle (*spz*; BANY.1.2.g04910) is a ligand that enables the activation of the Toll pathway in *Drosophila* (Yamamoto-Hino and Goto 2016). The role of *spz* is currently unknown in the context of eyespot development but this ligand could be an interesting target for further study. Our data suggest that genetic variation at these loci might also be regulating hindwing eyespot number variation.

Interestingly, some of our candidate genes involve *Drosophila* embryonic, wing-specific, and dorso/ventral patterning genes. The latter function is in line with eyespot number variation being observed on the dorsal, but not on the ventral, surface of the hindwing. Muscle segmentation homeobox (*msh*; BANY.1.2.g11326), Really Interesting New Gene-box protein 1A (*Roc1a*; BANY.1.2.g19639), and Annexin B9 (*AnxB9*; BANY.1.2.g25219) are all involved in embryonic development or wing morphogenesis in *Drosophila*. *Msh* specifies dorsal cell fate in the *Drosophila* wing (Milán *et al.* 2001) and *Roc1a* negatively regulates canonical Wnt signaling (Roberts *et al.* 2012), a key pathway for eyespot development (Özsu *et al.* 2017). Lastly, *AnxB9* has been proposed to be involved in wing morphogenesis and wing dorsal pattern identity in *Drosophila* (Bejarano *et al.* 2008).

Finally, we have identified several candidate genes with roles in pre- and post-transcriptional regulation (*Dach1*, *ZC3H10*, *abd-A*, *mbo*, *Eif4g3*, and *Mi-2*), and signal transduction

(*gpr180* and *TRAIP*), that might play a role in DHSN variation. Of particular interest is Dachshund homolog 1 (*Dach1*; BANY.1.2.g00636), a gene that is involved in limb development and that is activated by Distal-less in *Drosophila* (Estella *et al.* 2012). Another is *abd-A* (abdominal-A; BANY.1.2.g02585), one of only two genes providing butterfly hindwing chromatin regulatory identity (Lewis and Reed 2019). Other genes (*ZC3H10*, *mbo*, and *Mi*-2) have SNPs within introns or SNPs leading to nonsynonymous mutations that can possibly provide them with new binding affinity (Table 3). One of these intronic variants is annotated as putatively affecting a splice region within an intron in Chromodomain-helicase-DNA-binding protein *Mi-2* homolog (*Mi-2*; BANY.1.2.g12184). Moreover, we observed nonsynonymous mutations in two different genes: nuclear pore complex protein Nup88 (*mbo*; BANY.1.2.g11151) with a leucine to proline substitution at position 584, and Zinc finger CCCH domain-containing protein 10 (*ZC3H10*; BANY.1.2.g01110) with an isoleucine to valine substitution at position 142 (Table 3 and Table S2). Future functional studies will be required to further implicate these genes in eyespot development and eyespot number regulation.

Effects of noncoding mutations on the evolution of spot number variation

After identifying 16 regions of the *B. anynana* genome associated with DHSN variation and characterizing the relationship of identified SNPs with nearby genes, we observed that the majority of SNPs fall outside coding sequences. Only two genes, *mbo* and *ZC3H10*, contain variants annotated as nonsynonymous mutations of unknown effect on the resulting protein, while a third, *Mi-2*, possesses an annotated splice region variant. Overall, DNA variation at noncoding loci suggests a complex *cis*-regulatory landscape controlling DHSN throughout the mediated expression of the nearby genes described above. Indeed, *cis*-regulatory elements are thought to have profound implications in the evolution of morphological diversity (Carroll 2008). In particular, they have been associated with variation in pigmentation patterns in a wide variety of animal systems, including the evolution of eggspot pigmentation patterns in cichlids (Santos *et al.* 2014), wing pigmentation patterns in *Drosophila* (Werner *et al.* 2010; Koshikawa *et al.* 2015), divergent pigmentation patterns in capuchino seedeater finches (Campagna *et al.* 2017), and variation in red, black, and yellow color patterns in *Heliconius* butterflies due to regulatory changes in the *optix*, *WntA*, *cortex*, and *aristaless* genes (Reed *et al.* 2011; Supple *et al.* 2013; Martin and Reed 2014; Van Belleghem *et al.* 2017; Westerman *et al.* 2018) among others. In the case of eyespot number variation, regulatory mutations around the genes identified here might disrupt the reaction–diffusion mechanism of eyespot center differentiation (Connahs *et al.* 2019), or later processes of eyespot center signaling, that eventually translate to the presence or absence of an eyespot in particular wing sectors.

The genetic variation uncovered in this work affects eyespot number variation on the dorsal surface but not on the ventral surface of the wing. Thus, our work suggests that the genetic variants identified with our analysis affect eyespot number in a surface-specific manner. This surface-specific regulation is potentially mediated via *apA*, a previously identified dorsal eyespot repressor (Prakash and Monteiro 2018). The polygenic nature of our results suggests that genetic variation at the loci identified above (*e.g.*, *Antp*, *Ubx*, *dpp*, etc.), rather than at the *apA* locus itself, regulates dorsal eyespot number. We can speculate that changes in gene expression at the identified loci might impact the expression of *apA* in specific wing sectors on the dorsal surface, ultimately controlling eyespot differentiation in those regions.

Finally, the use of a genome-wide sequencing strategy allowed us to discover a series of independent loci that appear to contribute to DHSN in *B. anynana*. These loci, predominantly composed of (or linked to) polymorphisms in noncoding DNA, suggest that changes in DHSN mostly occur in regions that may regulate the expression of previously known eyespot-associated genes. This study strongly suggests a polygenic architecture of hindwing eyespot number variation that might be most likely mediated by epistatic interactions among a set of developmental genes. Thus, while our work has enriched the list of genes involved in eyespot number variation, to be tested by future works, it also supports the hypothesis that variation at multiple genes, rather than at a single master regulator or input–output gene (such as *shavenbaby* or *achaete-scute*), is involved in regulating the number of serial homologs. This highlights a more complex, but still poorly understood, genetic architecture for serial homolog number regulation.

Acknowledgments

We thank Elizabeth Schyling for help rearing families for heritability analysis; Robert Rak and Chris Bolick for rearing corn plants for the larvae; and the University of Puerto Rico, Río Piedras Campus (UPR-RP) Sequencing and Genomics Facility, and the UPR-RP High Performance Computing Facility, for additional support in library preparation and computational analysis. This work was funded by a National Science Foundation (NSF) Doctoral Dissertation Improvement grant to E.L.W. and A.M. (IOS-110382); an NSF Puerto Rico Louis Stokes Alliance for Minority Participation Bridge to the Doctorate Program (NSF grant award HRD-1139888) to A.G.R.-C.; NSF grant IOS-1656389 to R.P.; and a Puerto Rico IDeA Networks of Biomedical Research Excellence grant P20 GM-103475 from the National Institute for General Medical Sciences, a component of the National Institutes of Health, and Ministry of Education, Singapore grants (MOE R-154-000-602-112 and MOE2015-T2-2-159) to A.M.

Literature Cited

Axelsson, E., A. Ratnakumar, M.-L. Arendt, K. Maqbool, M. T. Webster *et al.*, 2013 The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495: 360–364. https://doi.org/10.1038/nature11837

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3: e3376. https://doi.org/10.1371/journal.pone.0003376

Beckett, K., S. Monier, L. Palmer, C. Alexandre, H. Green *et al.*, 2013 Drosophila S2 cells secrete wingless on exosome-like vesicles but the wingless gradient forms independently of exosomes. Traffic 14: 82–96. https://doi.org/10.1111/tra.12016

Bejarano, F., C. M. Luque, H. Herranz, G. Sorrosal, N. Rafel *et al.*, 2008 A gain-of-function suppressor screen for genes involved in dorsal–ventral boundary formation in the Drosophila wing. Genetics 178: 307–323. https://doi.org/10.1534/genetics.107.081869

Beldade, P., and P. M. Brakefield, 2003 Concerted evolution and developmental integration in modular butterfly wing patterns. Evol. Dev. 5: 169–179. https://doi.org/10.1046/j.1525-142X.2003.03025.x

Beldade, P., V. French, and P. M. Brakefield, 2008 Developmental and genetic mechanisms for evolutionary diversification of serial repeats: eyespot size in Bicyclus anynana butterflies. J. Exp. Zool. B Mol. Dev. Evol. 310: 191–201. https://doi.org/10.1002/jez.b.21173

Bhardwaj, S., K. L. Prudic, A. Bear, M. Dasgupta, B. R. Wasik *et al.*, 2018 Sex Differences in 20-hydroxyecdysone hormone levels control sexual dimorphism in *Bicyclus anynana* wing patterns. Mol. Biol. Evol. 35: 465–472. https://doi.org/10.1093/molbev/msx301

Brakefield, P. M., and A. J. van Noordwijk, 1985 The genetics of spot pattern characters in the meadow brown butterfly Maniola jurtina (Lepidoptera: Satyrinae). Heredity 54: 275–284. https://doi.org/10.1038/hdy.1985.37

Brakefield, P. M., and V. French, 1993 Butterfly wing patterns. Acta Biotheor. 41: 447–468. https://doi.org/10.1007/BF00709376

Brelsford, A., D. P. L. Toews, and D. E. Irwin, 2017 Admixture mapping in a hybrid zone reveals loci associated with avian feather coloration. Proc. Biol. Sci. 284: 20171106. https://doi.org/10.1098/rspb.2017.1106

Browning, S. R., and B. L. Browning, 2007 Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81: 1084–1097. https://doi.org/10.1086/521987

Brunetti, C. R., J. E. Selegue, A. Monteiro, V. French, P. M. Brakefield *et al.*, 2001 The generation and diversification of butterfly eyespot color patterns. Curr. Biol. 11: 1578–1585. https://doi.org/10.1016/S0960-9822(01)00502-4

Campagna, L., M. Repenning, L. F. Silveira, C. S. Fontana, P. L. Tubaro *et al.*, 2017 Repeated divergent selection on pigmentation genes in a rapid finch radiation. Sci. Adv. 3: e1602404. https://doi.org/10.1126/sciadv.1602404

Campbell, C. R., J. W. Poelstra, and A. D. Yoder, 2018 What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. Biol. J. Linn. Soc. Lond. 124: 561–583. https://doi.org/10.1093/biolinnean/bly063

Carroll, S. B., 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134: 25–36. https://doi.org/10.1016/j.cell.2008.06.030

Catchen, J. M., A. Amores, P. A. Hohenlohe, W. A. Cresko, J. H. Postlethwait *et al.*, 2011 Stacks: building and genotyping loci de novo from short-read sequences. G3 (Bethesda) 1: 171–182. https://doi.org/10.1534/g3.111.000240

Catchen, J. M., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko, 2013 Stacks: an analysis tool set for population genomics. Mol. Ecol. 22: 3124–3140. https://doi.org/10.1111/mec.12354

Chang, H. C., and G. M. Rubin, 1997 14–3-3 epsilon positively regulates Ras-mediated signaling in Drosophila. Genes Dev. 11: 1132–1139. https://doi.org/10.1101/gad.11.9.1132

Cheah, P. Y., W. Chia, and X. Yang, 2000 Jumeaux, a novel Drosophila winged-helix family protein, is required for generating asymmetric sibling neuronal cell fates. Development 127: 3325–3335.

Chen, P., X. L. Tong, D. D. Li, M. Y. Fu, S. Z. He *et al.*, 2013 Antennapedia is involved in the development of thoracic legs and segmentation in the silkworm, Bombyx mori. Heredity 111: 182–188. https://doi.org/10.1038/hdy.2013.36

Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen *et al.*, 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6: 80–92. https://doi.org/10.4161/fly.19695

Clarke, G. M., C. A. Anderson, F. H. Pettersson, L. R. Cardon, A. P. Morris *et al.*, 2011 Basic statistical analysis in genetic case-control studies. Nat. Protoc. 6: 121–133. https://doi.org/10.1038/nprot.2010.182

Connahs H., S. Tlili, J. van Creij, T. Y. J. Loo, T. Das Banerjee *et al.*, 2019 Activation of butterfly eyespots by Distal-less is consistent with a reaction-diffusion process. Development 146: dev169367. https://doi.org/https://doi.org/10.1242/dev.169367

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Davey J. W., M. Chouteau, S. L. Barker, L. Maroja, S. W. Baxter, *et al.*, 2016 Major improvements to the Heliconius melpomene genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. G3 (Bethesda) 6: 695–708 (erratum: G3 (Bethesda) 6: 1489). https://doi.org/10.1534/g3.115.023655

Espeland, M., J. Breinholt, K. R. Willmott, A. D. Warren, R. Vila *et al.*, 2018 A comprehensive and dated phylogenomic analysis of butterflies. Curr. Biol. 28: 770–778.e5. https://doi.org/10.1016/j.cub.2018.01.061

Estella, C., R. Voutev, and R. S. Mann, 2012 A dynamic network of morphogens and transcription factors patterns the fly leg, pp. 173–198 in *Transcriptional Switches During Development*, edited by S. Plaza and F. Payre. Academic Press, New York.

Etter, P. D., S. L. Bassham, P. A. Hohenlohe, E. Johnson, and W. A. Cresko, 2011 SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol. Biol. 772: 157–178. https://doi.org/10.1007/978-1-61779-228-1_9

Falconer, D. S., and T. F. C. Mackay, 1996 Heritability, pp. 160–183 in *Introduction to Quantitative Genetics*, Longman, Essex, England.

Feng, S., S. Wang, C.-C. Chen, and L. Lan, 2011 GWAPower: a statistical power calculation software for genome-wide association studies with quantitative traits. BMC Genet. 12: 12. https://doi.org/10.1186/1471-2156-12-12

Finlin, B. S., C.-L. Gau, G. A. Murphy, H. Shao, T. Kimel *et al.*, 2001 RERG is a novel ras-related, estrogen-regulated and growth-inhibitory gene in breast cancer. J. Biol. Chem. 276: 42259–42267. https://doi.org/10.1074/jbc.M105888200

Galant, R., and S. B. Carroll, 2002 Evolution of a transcriptional repression domain in an insect Hox protein. Nature 415: 910–913. https://doi.org/10.1038/nature717

García-Bellido, A., and J. F. de Celis, 2009 The complex tale of the achaete-scute complex: a paradigmatic case in the analysis of

gene organization and function during development. Genetics 182: 631–639. https://doi.org/10.1534/genetics.109.104083
Hannun, Y. A., and L. M. Obeid, 2017 Sphingolipids and their metabolism in physiology and disease. Nat. Rev. Mol. Cell Biol. 19: 175–191 (erratum: Nat. Rev. Mol. Cell Biol. 19: 673). https://doi.org/10.1038/nrm.2017.107
Holloway, G. J., P. M. Brakefield, and S. Kofman, 1993 The genetics of wing pattern elements in the polyphenic butterfly, Bicyclus anynana. Heredity 70: 179–186. https://doi.org/10.1038/hdy.1993.27
Jombart, T., 2008 adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403–1405. https://doi.org/10.1093/bioinformatics/btn129
Jombart, T., and I. Ahmed, 2011 Adegenet 1.3–1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27: 3070–3071. https://doi.org/10.1093/bioinformatics/btr521
Koshikawa, S., M. W. Giorgianni, K. Vaccaro, V. A. Kassner, J. H. Yoder *et al.*, 2015 Gain of cis -regulatory activities underlies novel domains of wingless gene expression in Drosophila. Proc. Natl. Acad. Sci. USA 112: 7524–7529. https://doi.org/10.1073/pnas.1509022112
Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway *et al.*, 2004 Versatile and open software for comparing large genomes. Genome Biol. 5: R12. https://doi.org/10.1186/gb-2004-5-2-r12
Larouche, O., M. L. Zelditch, and R. Cloutier, 2017 Fin modules: an evolutionary perspective on appendage disparity in basal vertebrates. BMC Biol. 15: 32. https://doi.org/10.1186/s12915-017-0370-x
Lewis, J. J., and R. D. Reed, 2019 Genome-wide regulatory adaptation shapes population-level genomic landscapes in Heliconius. *Mol. Biol. Evol.* 36: 159–173. https://doi.org/10.1093/molbev/msy209
Li, H., and R. M. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
Li, H., B. Handsaker, A. Wysoker, T. J. Fennell, J. Ruan *et al.*, 2009 The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
Li, Q., L. Shen, T. Xin, W. Xiang, W. Chen *et al.*, 2009 Role of Scrib and Dlg in anterior-posterior patterning of the follicular epithelium during Drosophila oogenesis. BMC Dev. Biol. 9: 60. https://doi.org/10.1186/1471-213X-9-60
Marcellini, S., and P. Simpson, 2006 Two or four bristles: functional evolution of an enhancer of scute in Drosophilidae. *PLoS Biol.* 4: e386. https://doi.org/10.1371/journal.pbio.0040386
Martin, A., and R. D. Reed, 2014 Wnt signaling underlies evolution and development of the butterfly wing pattern symmetry systems. Dev. Biol. 395: 367–378. https://doi.org/10.1016/j.ydbio.2014.08.031
Matsuoka, Y., and A. Monteiro, 2019 Hox genes are essential for the development of novel serial homologous eyespots on the wings of Bicyclus anynana butterflies. bioRxiv. Available at: https://www.biorxiv.org/content/10.1101/814848v1. Accessed: January 7, 2020. https://doi.org/10.1101/814848
McGregor, A. P., V. Orgogozo, I. Delon, J. Zanet, D. G. Srinivasan *et al.*, 2007 Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature 448: 587–590. https://doi.org/10.1038/nature05988
McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303. https://doi.org/10.1101/gr.107524.110
Milán M., U. Weihe, S. Tiong, W. Bender, and S. M. Cohen, 2001 msh specifies dorsal cell fate in the Drosophila wing. Development 128: 3263–3268.
Montague, M. J., G. Li, B. Gandolfi, R. Khan, B. L. Aken *et al.*, 2014 Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. Proc. Natl. Acad. Sci. USA 111: 17230–17235. https://doi.org/10.1073/pnas.1410083111
Monteiro, A. F., P. M. Brakefield, and V. French, 1994 The evolutionary genetics and developmental basis of wing pattern variation in the butterfly Bicyclus anynana. Evolution 48: 1147–1157. https://doi.org/10.2307/2410374
Monteiro, A., P. M. Brakefield, and V. French, 1997 Butterfly eyespots: the genetics and development of the color rings. Evolution 51: 1207–1216.
Monteiro, A. A., J. Prijs, M. Bax, T. Hakkaart, and P. M. Brakefield, 2003 Mutants highlight the modular control of butterfly eyespot patterns. Evol. Dev. 5: 180–187. https://doi.org/10.1046/j.1525-142X.2003.03029.x
Monteiro, A., G. Glaser, S. Stockslager, N. Glansdorp, and D. M. Ramos, 2006 Comparative insights into questions of lepidopteran wing pattern homology. BMC Dev. Biol. 6: 52. https://doi.org/10.1186/1471-213X-6-52
Monteiro, A., B. Chen, L. C. Scott, L. Vedder, H. J. Prijs *et al.*, 2007 The combined effect of two mutations that alter serially homologous color pattern elements on the fore and hindwings of a butterfly. BMC Genet. 8: 22. https://doi.org/10.1186/1471-2156-8-22
Monteiro, A., B. Chen, D. M. Ramos, J. C. Oliver, X. Tong *et al.*, 2013 Distal-less regulates eyespot patterns and melanization in Bicyclus butterflies. J. Exp. Zool. Part B Mol. Dev. Evol. 320: 321–331. https://doi.org/10.1002/jez.b.22503
Nadeau, N. J., M. Ruiz, P. Salazar, B. A. Counterman, J. A. Medina *et al.*, 2014 Population genomics of parallel hybrid zones in the mimetic butterflies, H. melpomene and H. erato. Genome Res. 24: 1316–1333. https://doi.org/10.1101/gr.169292.113
Narum, S. R., C. A. Buerkle, J. W. Davey, M. R. Miller, and P. A. Hohenlohe, 2013 Genotyping-by-sequencing in ecological and conservation genomics. Mol. Ecol. 22: 2841–2847. https://doi.org/10.1111/mec.12350
Nijhout, H. F., 1991 *The development and evolution of butterfly wing patterns*. Smithsonian Institution Press, Washington, D. C.
Nowell, R. W., B. Elsworth, V. Oostra, B. J. Zwaan, C. W. Wheat *et al.*, 2017 A high-coverage draft genome of the mycalesine butterfly Bicyclus anynana. Gigascience 6: 1–7. https://doi.org/10.1093/gigascience/gix035
Ohde T., T. Yaginuma, and T. Niimi, 2013 Insect morphological diversification through the modification of wing serial homologs. Science 340: 495–498. https://doi.org/10.1126/science.1234219
Oliver, J. C., X. Tong, L. F. Gall, W. H. Piel, and A. Monteiro, 2012 A single origin for nymphalid butterfly eyespots followed by widespread loss of associated gene expression. PLoS Genet. 8: e1002893. https://doi.org/10.1371/journal.pgen.1002893
Oliver, J. C., J. M. Beaulieu, L. F. Gall, W. H. Piel, and A. Monteiro, 2014 Nymphalid eyespot serial homologues originate as a few individualized modules. Proc. Biol. Sci. 281: 20133262. https://doi.org/10.1098/rspb.2013.3262
Owen, D., 1993 Spot variation in Maniola jurtina (L.) (Lepidoptera: satyridae) in southern Portugal and a comparison with the canary islands. Biol. J. Linn. Soc. Lond. 49: 355–365. https://doi.org/10.1111/j.1095-8312.1993.tb00911.x
Özsu, N., and A. Monteiro, 2017 Wound healing, calcium signaling, and other novel pathways are associated with the formation of butterfly eyespots. BMC Genomics 18: 788. https://doi.org/10.1186/s12864-017-4175-7
Özsu, N., Q. Y. Chan, B. Chen, M. Das Gupta, and A. Monteiro, 2017 Wingless is a positive regulator of eyespot color patterns in Bicyclus anynana butterflies. Dev. Biol. 429: 177–185. https://doi.org/10.1016/j.ydbio.2017.06.030
Payre, F., A. Vincent, and S. Carreno, 1999 ovo/svb integrates Wingless and DER pathways to control epidermis differentiation. Nature 400: 271–275. https://doi.org/10.1038/22330

Prakash, A., and A. Monteiro, 2018 Apterous A specifies dorsal wing patterns and sexual traits in butterflies. Proc. R. Soc. Lond. B Biol. Sci. 285: 20172685. https://doi.org/10.1098/rspb.2017.2685

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira *et al.*, 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81: 559–575. https://doi.org/10.1086/519795

Reed, R. D., and M. S. Serfas, 2004 Butterfly wing pattern evolution is associated with changes in a notch/distal-less temporal pattern formation process. Curr. Biol. 14: 1159–1166. https://doi.org/10.1016/j.cub.2004.06.046

Reed R. D., R. Papa, A. Martin, H. M. Hines, B. A. Counterman, *et al.*, 2011 optix drives the repeated convergent evolution of butterfly wing pattern mimicry. Science 333: 1137–1141. https://doi.org/10.1126/science.1208227

Rellstab, C., F. Gugerli, A. J. Eckert, A. M. Hancock, and R. Holderegger, 2015 A practical guide to environmental association analysis in landscape genomics. Mol. Ecol. 24: 4348–4370. https://doi.org/10.1111/mec.13322

Roberts, D. M., M. I. Pronobis, K. M. Alexandre, G. C. Rogers, J. S. Poulton *et al.*, 2012 Defining components of the ß-catenin destruction complex and exploring its regulation and mechanisms of action during development. *PLoS One* 7: e31284. https://doi.org/10.1371/journal.pone.0031284

Rochette, N. C., and J. M. Catchen, 2017 Deriving genotypes from RAD-seq short-read data using Stacks. Nat. Protoc. 12: 2640–2659. https://doi.org/10.1038/nprot.2017.123

Ronshaugen, M., N. McGinnis, and W. McGinnis, 2002 Hox protein mutation and macroevolution of the insect body plan. Nature 415: 914–917. https://doi.org/10.1038/nature716

Ruvinsky I., and J. J. Gibson-Brown, 2000 Genetic and developmental bases of serial homology in vertebrate limb evolution. Development 127: 5233–5244.

Saccheri, I. J., and M. W. Bruford, 1993 DNA fingerprinting in a butterfly, Bicyclus anynana (Satyridae). J. Hered. 84: 195–200. https://doi.org/10.1093/oxfordjournals.jhered.a111316

Saccheri, I. J., P. M. Brakefield, and R. A. Nichols, 1996 Severe inbreeding depression and rapid fitness rebound in the butterfly Bicyclus anynana (Satyridae). Evolution 50: 2000–2013. https://doi.org/10.1111/j.1558–5646.1996.tb03587.x

Saenko, S. V., M. S. Marialva, and P. Beldade, 2011 Involvement of the conserved Hox gene Antennapedia in the development and evolution of a novel trait. Evodevo 2: 9. https://doi.org/10.1186/2041-9139-2-9

Santos, M. E., I. Braasch, N. Boileau, B. S. Meyer, L. Sauteur *et al.*, 2014 The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat. Commun. 5: 5149. https://doi.org/10.1038/ncomms6149

Schachat, S. R., J. C. Oliver, and A. Monteiro, 2015 Nymphalid eyespots are co-opted to novel wing locations following a similar pattern in independent lineages. BMC Evol. Biol. 15: 20. https://doi.org/10.1186/s12862-015-0300-x

Shirai, L. T., S. V. Saenko, R. A. Keller, M. A. Jerónimo, P. M. Brakefield *et al.*, 2012 Evolutionary history of the recruitment of conserved developmental genes in association to the formation and diversification of a novel trait. BMC Evol. Biol. 12: 21. https://doi.org/10.1186/1471-2148-12-21

Stern, D. L., and V. Orgogozo, 2008 The loci of evolution: how predictable is genetic evolution? Evolution 62: 2155–2177. https://doi.org/10.1111/j.1558–5646.2008.00450.x

Supple, M. A., H. M. Hines, K. K. Dasmahapatra, J. J. Lewis, D. M. Nielsen *et al.*, 2013 Genomic architecture of adaptive color pattern divergence and convergence in Heliconius butterflies. Genome Res. 23: 1248–1257. https://doi.org/10.1101/gr.150615.112

Tien, A.-C., A. Rajan, K. L. Schulze, H. D. Ryoo, M. Acar *et al.*, 2008 Ero1L, a thiol oxidase, is required for Notch signaling through cysteine bridge formation of the Lin12-Notch repeats in Drosophila melanogaster. J. Cell Biol. 182: 1113–1125. https://doi.org/10.1083/jcb.200805001

Tokita, C. K., J. C. Oliver, and A. Monteiro, 2013 A survey of eyespot sexual dimorphism across nymphalid butterflies. Int. J. Evol. Biol. 2013: 1–6. https://doi.org/10.1155/2013/926702

Tomoyasu, Y., S. R. Wheeler, and R. E. Denell, 2005 Ultrabithorax is required for membranous wing identity in the beetle Tribolium castaneum. Nature 433: 643–647. https://doi.org/10.1038/nature03272

Tong, X., S. Hrycaj, O. Podlaha, A. Popadic, and A. Monteiro, 2014 Over-expression of Ultrabithorax alters embryonic body plan and wing patterns in the butterfly Bicyclus anynana. Dev. Biol. 394: 357–366. https://doi.org/10.1016/j.ydbio.2014.08.020

Van Belleghem S. M., P. Rastas, A. Papanicolaou, S. H. Martin, C. F. Arias, *et al.*, 2017 Complex modular architecture around a simple toolkit of wing pattern genes. Nat. Ecol. Evol. 1: 52. https://doi.org/10.1038/s41559-016-0052

van Bergen, E., P. M. Brakefield, S. Heuskin, B. J. Zwaan, and C. M. Nieberding, 2013 The scent of inbreeding: a male sex pheromone betrays inbred males. Proc. Biol. Sci. 280: 20130102. https://doi.org/10.1098/rspb.2013.0102

Weatherbee, S. D., H. Frederik Nijhout, L. W. Grunert, G. Halder, R. Galant *et al.*, 1999 Ultrabithorax function in butterfly wings and the evolution of insect wing patterns. Curr. Biol. 9: 109–115. https://doi.org/10.1016/S0960-9822(99)80064-5

Werner, T., S. Koshikawa, T. M. Williams, and S. B. Carroll, 2010 Generation of a novel wing colour pattern by the Wingless morphogen. Nature 464: 1143–1148. https://doi.org/10.1038/nature08896

Westerman, E. L., N. Chirathivat, E. Schyling, and A. Monteiro, 2014 Mate preference for a phenotypically plastic trait is learned, and may facilitate preference-phenotype matching. Evolution 68: 1661–1670. https://doi.org/10.1111/evo.12381

Westerman, E. L., N. W. VanKuren, D. Massardo, A. Tenger-Trolander, W. Zhang *et al.*, 2018 Aristaless controls butterfly wing color variation used in mimicry and mate choice. Curr. Biol. 28: 3469–3474.e4. https://doi.org/10.1016/j.cub.2018.08.051

Yamamoto-Hino, M., and S. Goto, 2016 Spätzle-processing enzyme-independent activation of the Toll pathway in *Drosophila* innate immunity. Cell Struct. Funct. 41: 55–60. https://doi.org/10.1247/csf.16002

Zhang, L., and R. D. Reed, 2016 Genome editing in butterflies reveals that spalt promotes and Distal-less represses eyespot colour patterns. Nat. Commun. 7: 11769. https://doi.org/10.1038/ncomms11769

Zhou, X., and M. Stephens, 2012 Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44: 821–824. https://doi.org/10.1038/ng.2310

Zhou, X., and M. Stephens, 2014 Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat. Methods 11: 407–409. https://doi.org/10.1038/nmeth.2848

Communicating editor: B. Lazzaro