Multiple loci control eyespot number variation on the hindwings of *Bicyclus anynana* butterflies

3

4 Angel G. Rivera-Colón^{*,†}, Erica L. Westerman[‡], Steven M. Van Belleghem[†], Antónia
5 Monteiro^{§,**}, and Riccardo Papa[†]

6

Affiliations

8 ^{*}Department of Animal Biology, University of Illinois, Urbana-Champaign

9 [†]University of Puerto Rico-Rio Piedras Campus

10 [‡]University of Arkansas, Fayetteville

11 [§]National University of Singapore

12 ^{**}Yale-NUS College

13

Data accessibility

15 The *Bicyclus anynana* PstI RAD-tag sequencing data is available via the Genbank
16 BioProject PRJNA509697. Genotype VCF files will be made available through
17 Figshare upon acceptance.

18 Running Title

19 Genetics of hindwing eyespot number variation

20

21 Key words

22 Serial homology, genetics, *apterous*, eyespot number, *Bicyclus anynana*, genetic

23 architecture

24

25 Co-Corresponding Authors

26 antonia.monteiro@nus.edu.edu and rpapa.lab@gmail.com

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. Other hypotheses implicate genetic variation in upstream or downstream loci of master control genes. Here, we use the butterfly *Bicyclus anynana*, a species which 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 number of offspring on mid-parent 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 (RAD-seq) from multiple individuals varying in number of eyespots sampled across a freely breeding lab population. We found that dorsal hindwing eyespot number has a moderately high heritability of approximately 0.50. In addition, multiple loci near previously identified genes involved in eyespot development 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.

Introduction

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 co-option 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). Regulatory targets of abdominal *hox* genes 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 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 overlaps some, but not all bristles, is also known to impact bristle number in lab mutants (Garcia-Bellido and de Celis 2009). Finally, *shavenbaby* and *scute* genes are also known as input-output genes due to their central “middle of the hour-glass” 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 trichome 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. More systems, thus, need to be investigated for a more thorough understanding of the genetic basis underlying 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) (Fig. 1). Dorsal hindwing eyespot number 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). Lab 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 eyespot (P-) 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 CRISPR-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 surface of wings.

The genetic architecture of eyespot number variation in any butterfly species remains unknown. Here, we examine the genetic basis of dorsal hindwing eyespot number (DHEN) 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 mid-parent 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 lab population.

Materials and Methods

Study organism

Bicyclus anynana is a Nymphalid butterfly common to sub-tropical Africa for which a colony has been maintained in the laboratory since 1988. All *Bicyclus 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 dorsal hindwing eyespot number 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 populations 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 post establishment (Saccheri and Bruford 1993; Saccheri *et al.* 1996). The colony was kept in controlled conditions of 12 hours light/dark cycles, 80% relative humidity and a temperature of 27°C. Larvae were fed on corn plants and adult butterflies on mashed banana, as described in previous publications (Westerman *et al.* 2014).

Heritability of dorsal hindwing eyespot number

We examined the number of dorsal hindwing eyespots (DHEN) on all offspring from 18 separately reared families whose parents differed in eyespot number: six families where both parents had DHEN of zero (OF x OM); six where both parents

187 had DHEN of one (1F x 1M); and six where both parents had DHEN of two (2F x
188 2M). All generations were reared in the conditions described above. We ensured
189 virginity of the females by separating the butterflies in the parental generation into
190 sex-specific cages on the day of eclosion. All families were started within 5 days of
191 each other using adults ranging from 1-3 days old. While initial adult age ranged
192 from 1-3 days, the average age was consistent across the three DHEN treatments
193 (ANOVA, $n=18$, $DF=2$, $F=0.8266$, $p=0.4565$). Each breeding pair was placed in a
194 cylindrical hanging net cage of 30 cm diameter X 40 cm height, with food (banana
195 slices), water and a young corn plant on which to lay eggs. When corn plants were
196 covered with eggs, they were placed in family-specific mesh sleeve cages for larval
197 growth. Females were given new plants on which to lay eggs until they died. Pupae
198 and pre-pupae were removed from the sleeve cages and placed in family-specific
199 cylindrical hanging net cages for eclosion. The cages were checked daily for newly
200 emerged butterflies. On the day of eclosion, DHEN was recorded for each
201 offspring. DHEN was calculated for each individual by averaging the number of
202 dorsal hindwing eyespots on the left and right wing, which allowed for
203 intermediate values when the wings were asymmetric. Narrow-sense heritability
204 (h^2) was calculated by regressing offspring on midparent values and correcting the
205 heritability estimate for assortative mating. This corrected heritability value h_0^2 was
206 calculated using correlations between phenotypic values of mated pairs (r) to
207 calculate the correlation between breeding values of mates ($m = rh^2$), with $h_0^2 =$
208 $h^2[(1-m)/(1+mh^2)]$ as described in Falconer & Mackay (1996). Estimates were
209 obtained for the pooled offspring data as well as for separate regressions of female
210 and male offspring data on mid-parent 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 DHEN (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 DHEN variation, we collected and sequenced a total of 30 individuals from the previously described laboratory population. Fifteen individuals contained no eyespots (absence) and 15 containing two or more eyespots (presence) (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 DNeasy Blood & Tissue Kit (Qiagen), with an additional RNase digestion for removing RNA from the extracted nucleic acid samples. The quality and concentration of the extracted DNA was verified using gel electrophoresis and Qubit 2.0 fluorometer (Life Technologies). Extracted genomic DNA was used for preparing Illumina RAD sequencing libraries based on previously described protocols (Baird *et al.* 2008;

Etter *et al.* 2011). DNA was digested with the frequent cutting enzyme *PstI* and ligated to P1 adapters containing a unique barcode 5 bp in length. Samples were pooled and sheared using a Covaris M220 (Covaris Inc.) instrument and size selected for 300-500 bp inserts on average. 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 *Stacks v1.42* (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; Catchen et al., 2011) *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 $\pm$ 1.2 million filtered paired-end reads.

Reference alignment

The 111 million retained read pairs were aligned to the most recent *B. anynana* genome assembly (v1.2)shiv 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, 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/>). Resulting alignments were directly converted to BAM files using SamTools v1.12 (Li *et al.* 2009a) *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 a 92.9 % read alignment and 79.6 % properly mapped read pairs. At each RAD locus, average per-individual sequencing coverage was 18.8X ($\pm$ 4.3; median: 18.3).

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, a minimum genotype depth of 5, present in over 50 % of all individuals, a max allele number of 2, and minimum allele frequency of 0.05. This results in a filtered VCF containing only 350,121 high-quality biallelic variants.

282 In parallel, the filtered mapped reads were genotyped using *Stacks* v1.42 (Catchen
283 *et al.* 2011, 2013). Using the *ref_map.pl* wrapper script with default parameters,
284 RAD loci were assembled from the mapped reads using *pstacks*. A catalog of all
285 loci was generated with *cstacks* and samples were matched to this catalog using
286 *sstacks*. The *populations* program was then run on this catalog to generate
287 population genetic measures, enabling the calculation of F_{ST} statistics. For a
288 variant to be included in the analysis, it had to be present in both study groups
289 ($p=2$) in over 75% percent of individuals ($r=0.75$) and have a minimum allele
290 frequency above 0.05. Using *Stacks* we reconstructed 207,752 RAD loci and
291 673,340 raw SNPs. After filtering, 73,159 RAD loci and 238,786 SNPs were
292 retained. Filtered variants obtained from both datasets were compared to retain
293 only SNPs with support from both genotyping methods. A total of 216,338 SNPs
294 were shared between the two datasets and were used for subsequent comparisons.
295

296 To identify areas of the genome associated with our desired phenotype, we scanned
297 the genome using two methods: F_{ST} and association via a relatedness-corrected
298 linear mixed model. F_{ST} for all 216 thousand shared variants was calculated using
299 the *Stacks populations* module. A Fisher's Exact Test p-value correction was
300 applied to the resulting F_{ST} and Analysis of Molecular Variance (AMOVA) F_{ST}
301 values as a multiple testing correction. Future references to F_{ST} in the manuscript
302 specifically refer to the corrected AMOVA F_{ST} values from this analysis. During this
303 *Stacks* run, we enabled the *--plink* flag to output our 216 thousand variant sites
304 into *Plink's .map* and *.ped* format for use in the association analysis. The generated
305 *.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.

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 are method agnostic. This multi-metric 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 a correlation and a quantile-quantile plot (Q-Q plot) of the F_{ST} and association for all variants (Fig. S1). Significant peaks were defined as areas of the genome with SNPs containing association and F_{ST} 6 standard deviations above the genome-wide mean. A second threshold was established for variants with *p*-value association and F_{ST} at least 7 standard deviations above the mean. We explored all annotated genes within a 500 kb window centered around outlier SNPs, 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 *SnEff*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 a 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 & 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 *adeget* 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 analyzed with 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 *Heliconius melpomene* v2 genome assembly (Davey *et al.* 2016). Although *B. anynana* and *H. melpomene* diverged about 80 My ago (Espeland *et al.* 2018) and have a different karyotype (n=28 in *B. anynana* versus n=21 in *H. melpomene*), the *H. melpomene* genome is the most closely related butterfly genome that has been assembled into highly contiguous chromosomal 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* genome. The obtained alignments were subsequently filtered for a minimum alignment length of 200 bp and a minimum percent identity (%IDY = (matches x 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* genome.

Linkage disequilibrium analysis

In addition to ordering the *B. anynana* scaffolds to the *H. melpomene* genome for assessing the genomic linkage of SNPs, we calculated linkage disequilibrium in our *B. anynana* study population. To calculate linkage disequilibrium for genomic SNPs, we phased 213,000 SNPs that were genotyped in all samples using *beagle* v4.1 (Browning and Browning 2007). Estimates of linkage disequilibrium were calculated from 100,000 randomly selected SNPs, using the *VCFtools* v0.1.14 (Danecek *et al.* 2011) *-hap-r2* function, with a max 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 linkage disequilibrium decay. Subsequently, this LD window size was used for the investigation of genes near associated loci.

Results

Dorsal hindwing spot number variation has moderate to high heritability

Zero spot females were only produced by 0x0 families and one 1x1 family, and were absent from any 2x2 DHS families. Zero spot males, however, were produced by all 0x0 families, all but one of the 1x1 families and all but one of the 2x2 families. Two spot females were produced by all but one (a 0x0) family, while 2 spot males were produced by all 2x2 families, but only two 1x1 families and one 0x0 family

(Table 1). These results demonstrate that alleles are sufficiently segregating in our experimental design to perform heritability estimates.

DHEN has 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, AIC (Akaike Information Criterion) = 1498.204, effect tests: sex $\chi^2=293.361$, $p<0.0001$; parental phenotype $\chi^2=271.56$, $p<0.0001$; sex* parental phenotype $\chi^2=0.032$, $p=0.8576$).

Genome-wide variation and linkage disequilibrium 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 presence of DHEN (pre) and samples with absent DHEN (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 phenotype group to the observed variation in first and second Principal Components (Fig. 2A). Additionally, we observed very similar genome-wide nucleotide diversity in the group displaying presence of DHEN (pre, $\pi = 0.0090$) when compared to the group

with absent DHEN (abs, $\pi = 0.0083$). 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 linkage disequilibrium decay (Fig. 2B), we observed a max smoothed r^2 value of 0.272, and a halving of r^2 within 470 kb (Fig. 2B). This window size suggests that average linkage blocks are around 500 Kb in length, and that variants within this distance may be in strong linkage disequilibrium.

Association mapping of dorsal hindwing spot number 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 methods). The distributions and correlations between the distributions of F_{ST} and association are strongly positively correlated across all SNPs with a slight bias of association values toward more significant values for intermediate p -values, however, below our outlier thresholds (Figure 2C, Figure S1). We observed 99 SNPs on 25 different scaffolds with both F_{ST} and association with DHEN variation at least 6 standard deviations above the mean F_{ST} and association estimates. These SNPs span 28.5 Mb (6.01 %) of the reference genome assembly (Fig. 3). When using a more stringent threshold at 7 standard deviations, we observe a total of 28 outlier SNPs in 8 scaffolds, spanning 7.78 Mb (1.64%) of

the genome. Ordering these scaffolds along the contiguously assembled *Heliconius melpomene* genome suggests that the 25 scaffolds with outlier SNPs above 6 standard deviations from the mean F_{ST} and association belong to 21 different regions in the genome, whereas the smaller subset of 8 scaffolds that have outlier SNPs above 7 standard deviations belong to 6 different regions in the genome (Fig. 4).

Candidate gene identification

Using the available Lepbase reference annotation for the *B. anynana* v1.2 assembly we identified the neighboring annotated genes and relative positioning of the 99 outlier SNPs on the 21 genomic regions. We observed a total of 742 annotated genes within our 500 kb windows around the outlier SNPs (Supplemental Table S2), of which 251 remained with the 28 outlier SNPs that are above the higher 7 standard deviation threshold. Only 25 of these 742 genes contained outlier SNPs in their introns or exons (Table 3, supplemental Table S2). On average, our outlier SNPs are 189 kb (± 144 kb) from the nearest gene. This suggest 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 9 resulted in synonymous and 2 in non-synonymous mutations (Table 3, Supplemental Table S2). Thus, in order to identify the most promising candidate genes that may be linked to the non-coding associated SNPs and regulating DHEN, we used our current knowledge of butterfly developmental genetics and identified 35 such genes (Table 2, Figure 4). Among those, 7 candidates were previously associated with eyespot development, i.e., *CDase*, *Dpp*,

Antp, *Ubx*, *Sal3*, *Cd63*, and *Syntaxin 7*, and the remaining 28 are linked to eyespot number variation here for the first time (Table 2). The latter include genes related to Notch, Toll, and Ras signaling pathways, wing morphogenesis, dorso/ventral pattern formation, and pigment biosynthesis and a number of molecules with regulatory and signaling functions (Table 2).

Discussion

The use of population genomic analyses, including association mapping, GWAS, and F_{ST} scans, has been extensively used in natural populations to identify the genomic signature underlying a number of biological processes such as hybridization, speciation, selection and ecological adaptation (Reviewed in Campbell, Poelstra, & Yoder, 2018; Narum, Buerkle, Davey, Miller, & Hohenlohe, 2013; Rellstab, Gugerli, Eckert, Hancock, & Holderegger, 2015). In our work, we applied a population genomics approach to characterize the genetic basis of dorsal hindwing eyespot number (DHEN) 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 DHEN 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 (DHEN variation).

494

495 The combined results of our heritability and genome-wide association study
 496 suggest that variation in dorsal hindwing eyespot number in *B. anynana* is a
 497 complex trait regulated by multiple loci. Combining a RAD sequencing genotyping
 498 approach with genome-wide population differentiation and association mapping,
 499 we identified up to 21 and 6 genomic regions associated with DHEN variation in
 500 *B. anynana*, at different levels of stringency (Figure 3 and 4). Within the 21
 501 genomic regions, we observed a total of 742 annotated genes (Supplemental Table
 502 2), which we narrowed down to 35 potential candidates (Table 2). Within the
 503 stronger associated 6 genomic regions (Figure 4), we reduced the list of potential
 504 candidate genes to a total of 11 (Table 2). Below we address the possible
 505 relationship of these candidates with eyespot development and DHEN variation.

506

507 **Known eyespot developmental genes associated to DHEN variation**

508 Among our identified candidates, we observed a total of seven genes previously
 509 implicated in eyespot development in *B. anynana*: *CDase*, *Dpp*, *Antp*, *Ubx*, *Sal3*,
 510 *Cd63*, and *Syntaxin 7* (Brunetti *et al.* 2001; Tong *et al.* 2014; Özsu and Monteiro
 511 2017; Connahs *et al.* 2019) (Table 2). From these, only *dpp* and *Cd63* remained
 512 associated at the highest significant threshold of 7 SD. Most of these genes are
 513 within or near the area of strongest association, and only *Cd63* has an associated
 514 SNP within an intron (Table 3 and Figure 4). Neutral Ceramidase (*CDase*;
 515 BANY.1.2.g00030) is an enzyme involved in the metabolism of sphingolipids
 516 which is down-regulated in eyespots relative to flanking wing tissue (Özsu and
 517 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 *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* mRNA is expressed in regions around the
developing eyespots in mid to late larval wings (Connahs *et al.* 2019) in anti-
colocalized 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 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 hindwing only, something that is not seen
in other butterflies with eyespots such as *Junonia coenia* (Tong *et al.* 2014). Over-
expression of *Ubx* led to eyespot size reductions in both fore and hindwings of *B.*
anynana (Tong *et al.* 2014) whereas absence of *Ubx* in clones of cells in the
hindwing of *J. coenia* (via a spontaneous unknown mutation) (Weatherbee *et al.*
1999) led to eyespot enlargements (of Cu1 hindwing eyespots to sizes that match
Cu1 forewing eyespot sizes). These results suggest a repressing role of *Ubx* for
eyespot on the hindwing. 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 *V. cardui* and *J. coenia* butterflies (Zhang and Reed 2016). 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 is the only candidate previously characterized as up-regulated in eyespots (Özsu and Monteiro 2017) to contain an association within its gene module, and more specifically a SNP in its 2nd intron. Lastly, Syntaxin 7 (*Stx7*; BANY.1.2.g20777) has a similar function to *Syntaxin 1a*, a gene that is down-regulated in eyespots (Özsu and Monteiro 2017). Genetic variation either linked with the protein coding sequence of these genes, or more likely with their regulatory region is likely affecting eyespot number variation in hindwings.

Novel candidate genes associated to DHEN variation

In addition to the above eyespot-associated genes, we also identified a group of 28 candidates that have not been 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.

565 A total of five genes (*dlg1*, *numb*, *Taf2*, *Ero1L*, and *ct*) implicated in our study
566 (Table 2) are known to interact with the *Notch* eyespot-associated gene (Reed and
567 Serfas 2004). Disks large 1 tumor suppressor, (*dlg1*; BANY.1.2.g00658), *numb*
568 (*numb*; BANY.1.2.g00681), Ero1-like protein, (*Ero1L*; BANY.1.2.g10819) and *ct*
569 (*ct*; BANY.1.2.g16724) are known to interact and regulate the *Notch* (*N*) signaling
570 pathway in *D. melanogaster* (Cheah *et al.* 2000; Tien *et al.* 2008; Li *et al.* 2009b).
571 The gene *ct* is the only one of these genes with an associated SNP, causing a
572 synonymous mutation. The existence and role of these interactions are unknown
573 in *B. anynana*, as is the role of the *Notch* receptor itself. However, the *Notch*
574 receptor has a dynamic pattern of expression (Reed and Serfas 2004) that is very
575 similar to that of *Distal-less*, a gene that has recently been implicated in setting up
576 the eyespot centers likely via a reaction-diffusion mechanism (Connahs *et al.*
577 2019). Genetic variation at these three genes could be interacting with the eyespot
578 differentiation process through unknown mechanisms.

579 We also identified new members of the Ras and Toll signaling pathways that have
580 previously been associated with eyespot development (Özsu and Monteiro 2017).
581 14-3-3 epsilon (BANY.1.2.g16720) is a member of the Ras signaling pathway
582 (Chang and Rubin 1997) and Ras-related and estrogen-regulated growth
583 inhibitor-like protein (*rergl*; BANY.1.2.g10814), is a Ras-related superfamily gene
584 with unique characteristics (Finlin *et al.* 2001). Protein spaetzle (*spz*;
585 BANY.1.2.g04910) is a ligand that enables the activation of the Toll pathway in
586 *Drosophila* (Yamamoto-Hino and Goto 2016). The role of *spz* is currently
587 unknown in the context of eyespot development but this ligand could be an

interesting target of further study. Our data suggests 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. cAMP-dependent protein kinase catalytic subunit 3 (*Pka-C3*; BANY.1.2.g02300), Transmembrane emp24 domain-containing protein *baiser* (*bai*; BANY.1.2.g02301), Muscle segmentation homeobox (*msh*; BANY.1.2.g11326) and RING-box protein 1A (*Roc1a*; BANY.1.2.g19639) are all involved in embryonic development or wing morphogenesis in *Drosophila*. *Pka-C3* is specifically expressed in wing disc and leg discs of third instar larvae (Meléndez *et al.* 1995). *bai* is involved in the regulation of a receptor of *dpp* in *Drosophila* embryos (Bartoszewski *et al.* 2004) and in *Wnt* protein secretion (Li *et al.* 2015), both key signaling pathways implicated in eyespot size regulation (Özsu and Monteiro 2017; Connahs *et al.* 2019). *Msh* specifies dorsal cell fate in the *Drosophila* wing (Milán *et al.* 2001), and *Roc1a* negatively regulates the 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).

Among our candidates we also have two transporters of pigment precursors into pigment cells responsible for a fruit fly's eye color: White (*w*; BANY.1.2.g21276) and scarlet (*st*; BANY.1.2.g21277). In our study *white* is one of the few genes that has an associated SNP in the 7th intron. In *Drosophila* these two genes are

612 required, together with *Brown*, for normal pigmentation of the compound eye
613 (Merzendorfer 2014). The two transporters white and scarlet have been suggested
614 to import precursors of ommochrome pigments into butterfly scales, which are
615 necessary for yellow, red and brown colorations (Hines *et al.* 2012). It is unclear,
616 however, how these genes might play a role in the regulation of eyespot number
617 variation.

618 Finally, we have identified several candidate genes with roles in pre- and post-
619 transcriptional regulation (*Dach1*, *ZC3H10*, *abd-A*, *hhex*, *tll1*, *mbo*, *Eif4g3*, and
620 *Mi-2*) and signal transduction (*gpr180*, *PLEKHG4B*, and *TRAIP*) that might play
621 a role in DHEN variation. Of particular interest is Dachshund homolog 1 (*Dach1*;
622 BANY.1.2.g00636), a gene that is involved in limb development and that is
623 activated by Distal-less in *Drosophila* (Estella *et al.* 2012). Another is *abd-A*
624 (abdominal-A; BANY.1.2.g02585), one of only two genes providing butterfly's
625 hindwing chromatin regulatory identity (Lewis and Reed 2019). Other genes
626 (*ZC3H10*, *tll1*, *mbo*, and *Mi-2*) have SNPs within introns or SNPs leading to non-
627 synonymous mutations that can possibly provide them new binding affinity (Table
628 3). One of these intronic variants is annotated as putatively affecting a splice region
629 within an intron in Chromodomain-helicase-DNA-binding protein *Mi-2* homolog
630 (*Mi-2*; BANY.1.2.g12184). Moreover, we observed non-synonymous mutations in
631 two different genes: Nuclear pore complex protein Nup88 (*mbo*;
632 BANY.1.2.g11151) with a Leucine to Proline substitution at position 584, and Zinc
633 finger CCCH domain-containing protein 10 (*ZC3H10*; BANY.1.2.g01110) with a
634 Isoleucine to Valine substitution at position 142 (Table 3; Supplemental Table S2).

Future functional studies will be required to further implicate these genes in eyespot development and eyespot number regulation.

Effects of non-coding mutations in the evolution of eyespot number variation

After identifying 21 regions of the *B. anynana* genome associated with DHEN variation and characterizing the relationship of identified SNPs with nearby genes, we observed that the majority of SNPs (92.7%) fall outside coding sequences. Only two genes, *mbo* and *ZC3H10* contain variants annotated as non-synonymous mutations of unknown effect on the resulting protein, while a third, *Mi-2* possesses an annotated splice region variant. Overall, DNA variation at non-coding loci suggests a complex *cis*-regulatory landscape controlling DHEN 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). Particularly, 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 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 proposes 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 DHEN in *B. anynana*. These loci, predominantly composed of (or linked to) polymorphisms in non-coding DNA, suggest that changes in DHEN are mostly occurring in regions that regulate the expression of previously known eyespot-associated genes. This study identifies 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,

it also confirms that variation at multiple genes, rather than at a single top master regulator or input-output gene (such as *shavenbaby* or *achaete-scute*) is involved in regulating number of serial homologs. This highlights a more complex, but still poorly understood, genetic architecture for serial homolog number regulation.

Acknowledgements

We thank Elizabeth Schyling for help rearing families for heritability analysis, and Robert Rak and Chris Bolick for rearing corn plants for the larvae. We also thank the 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 NSF IOS-110382 Doctoral Dissertation Improvement Grant to ELW and AM, a NSF PR-LSAMP Bridge to the Doctorate Program (NSF Grant Award HRD1139888) to AGRC, NSF grant IOS-1656389 to RP, NSF grant OIA 1736026 to BC and RP, and a Puerto Rico INBRE Grant P20 GM103475 from the National Institute for General Medical Sciences (NIGMS), a component of the National Institutes of Health (NIH) and Ministry of Education, and Singapore grants (MOE R-154-000-602-112 and MOE2015-T2-2-159) to AM.

701 Literature Cited

- 702 Axelsson E., A. Ratnakumar, M.-L. Arendt, K. Maqbool, M. T. Webster, *et al.*, 2013 The genomic
703 signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495: 360–
704 364. <https://doi.org/10.1038/nature11837>
- 705 Baird N. a., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, *et al.*, 2008 Rapid SNP Discovery
706 and Genetic Mapping Using Sequenced RAD Markers, (J. C. Fay, Ed.). *PLoS One* 3: e3376.
707 <https://doi.org/10.1371/journal.pone.0003376>
- 708 Bartoszewski S., S. Luschig, I. Desjeux, J. Grosshans, and C. Nüsslein-Volhard, 2004 *Drosophila*
709 p24 homologues *eclair* and *baiser* are necessary for the activity of the maternally expressed
710 *Tkv* receptor during early embryogenesis. *Mech. Dev.* 121: 1259–1273.
711 <https://doi.org/10.1016/j.mod.2004.05.006>
- 712 Beckett K., S. Monier, L. Palmer, C. Alexandre, H. Green, *et al.*, 2013 *Drosophila* S2 Cells Secrete
713 Wingless on Exosome-Like Vesicles but the Wingless Gradient Forms Independently of
714 Exosomes. *Traffic* 14: 82–96. <https://doi.org/10.1111/tra.12016>
- 715 Bejarano F., C. M. Luque, H. Herranz, G. Sorrosal, N. Rafel, *et al.*, 2008 A Gain-of-Function
716 Suppressor Screen for Genes Involved in Dorsal–Ventral Boundary Formation in the
717 *Drosophila* Wing. *Genetics* 178: 307–323. <https://doi.org/10.1534/genetics.107.081869>
- 718 Beldade P., and P. M. Brakefield, 2003 Concerted evolution and developmental integration in
719 modular butterfly wing patterns. *Evol. Dev.* 5: 169–179. <https://doi.org/10.1046/j.1525-142X.2003.03025.x>
- 720 Beldade P., V. French, and P. M. Brakefield, 2008 Developmental and genetic mechanisms for
721 evolutionary diversification of serial repeats: eyespot size in *Bicyclus anynana* butterflies. *J.*
722 *Exp. Zool. Part B Mol. Dev. Evol.* 310B: 191–201. <https://doi.org/10.1002/jez.b.21173>
- 723 Belleghe S. M. Van, P. Rastas, A. Papanicolaou, S. H. Martin, C. F. Arias, *et al.*, 2017 Complex
724 modular architecture around a simple toolkit of wing pattern genes. *Nat. Ecol. Evol.* 1: 0052.
725 <https://doi.org/10.1038/s41559-016-0052>
- 726 Bergen E. van, P. M. Brakefield, S. Heuskin, B. J. Zwaan, and C. M. Nieberding, 2013 The scent of
727 inbreeding: a male sex pheromone betrays inbred males. *Proc. R. Soc. B Biol. Sci.* 280:
728 20130102–20130102. <https://doi.org/10.1098/rspb.2013.0102>
- 729 Brakefield P. M., and A. J. van Noordwijk, 1985 The genetics of spot pattern characters in the
730 meadow brown butterfly *Maniola jurtina* (Lepidoptera: Satyrinae). *Heredity* (Edinb). 54:
731 275–284. <https://doi.org/10.1038/hdy.1985.37>
- 732 Brakefield P. M., and V. French, 1993 Butterfly wing patterns. *Acta Biotheor.* 41: 447–468.
733 <https://doi.org/10.1007/BF00709376>
- 734 Brelsford A., D. P. L. Toews, and D. E. Irwin, 2017 Admixture mapping in a hybrid zone reveals
735 loci associated with avian feather coloration. *Proc. R. Soc. B* 284: 20171106.
736 <https://doi.org/10.1098/RSPB.2017.1106>
- 737 Browning S. R., and B. L. Browning, 2007 Rapid and Accurate Haplotype Phasing and Missing-
738 Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype
739 Clustering. *Am. J. Hum. Genet.* 81: 1084–1097. <https://doi.org/10.1086/521987>
- 740 Brunetti C. R., J. E. Selegue, A. Monteiro, V. French, P. M. Brakefield, *et al.*, 2001 The generation
741 and diversification of butterfly eyespot color patterns. *Curr. Biol.* 11: 1578–1585.
742 [https://doi.org/10.1016/S0960-9822\(01\)00502-4](https://doi.org/10.1016/S0960-9822(01)00502-4)
- 743 Campagna L., M. Repenning, L. F. Silveira, C. S. Fontana, P. L. Tubaro, *et al.*, 2017 Repeated
744 divergent selection on pigmentation genes in a rapid finch radiation. *Sci. Adv.* 3: e1602404.
745 <https://doi.org/10.1126/sciadv.1602404>
- 746 Campbell C. R., J. W. Poelstra, and A. D. Yoder, 2018 What is Speciation Genomics? The roles of
747 ecology, gene flow, and genomic architecture in the formation of species. *Biol. J. Linn. Soc.*
748 124: 561–583. <https://doi.org/10.1093/biolinnean/bly063>
- 749 Carroll S. B., 2008 Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of
750 Morphological Evolution. *Cell* 134: 25–36. <https://doi.org/10.1016/j.cell.2008.06.030>
- 751 Catchen J. M., A. Amores, P. A. Hohenlohe, W. A. Cresko, J. H. Postlethwait, *et al.*, 2011 Stacks:
752 Building and Genotyping Loci De Novo From Short-Read Sequences. *G3 Genes, Genomes,*
753 *Genet.* 1: 171–182. <https://doi.org/10.1534/g3.111.000240>

755 Catchen J. M., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko, 2013 Stacks: an
756 analysis tool set for population genomics. *Mol. Ecol.* 22: 3124–3140.
757 <https://doi.org/10.1111/mec.12354>
758 Chang H. C., and G. M. Rubin, 1997 14-3-3 epsilon positively regulates Ras-mediated signaling in
759 *Drosophila*. *Genes Dev.* 11: 1132–9. <https://doi.org/10.1101/gad.11.9.1132>
760 Cheah P. Y., W. Chia, and X. Yang, 2000 Jumeaux, a novel *Drosophila* winged-helix family
761 protein, is required for generating asymmetric sibling neuronal cell fates. *Development* 127:
762 3325–3335.
763 Chen P., X. L. Tong, D. D. Li, M. Y. Fu, S. Z. He, *et al.*, 2013 Antennapedia is involved in the
764 development of thoracic legs and segmentation in the silkworm, *Bombyx mori*. *Heredity*
765 (Edinb). 111: 182–188. <https://doi.org/10.1038/hdy.2013.36>
766 Cingolani P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, *et al.*, 2012 A program for annotating and
767 predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)*. 6: 80–92.
768 <https://doi.org/10.4161/fly.19695>
769 Clarke G. M., C. A. Anderson, F. H. Pettersson, L. R. Cardon, A. P. Morris, *et al.*, 2011 Basic
770 statistical analysis in genetic case-control studies. *Nat. Protoc.* 6: 121–133.
771 <https://doi.org/10.1038/nprot.2010.182>
772 Connahs H., S. Tlili, J. van Creijl, T. Y. J. Loo, T. Das Banerjee, *et al.*, 2019 Distal-less activates
773 butterfly eyespots consistent with a reaction diffusion process. *Development* dev.169367.
774 <https://doi.org/10.1242/dev.169367>
775 Danecek P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, *et al.*, 2011 The variant call format and
776 VCFtools. *Bioinformatics* 27: 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
777 Davey J. W., M. Chouteau, S. L. Barker, L. Maroja, S. W. Baxter, *et al.*, 2016 Major Improvements
778 to the *Heliconius melpomene* Genome Assembly Used to Confirm 10 Chromosome Fusion
779 Events in 6 Million Years of Butterfly Evolution. *G3 Genes|Genomes|Genetics* 6: 695–708.
780 <https://doi.org/10.1534/g3.115.023655>
781 Espeland M., J. Breinholt, K. R. Willmott, A. D. Warren, R. Vila, *et al.*, 2018 A Comprehensive
782 and Dated Phylogenomic Analysis of Butterflies. *Curr. Biol.* 28: 770–778.e5.
783 <https://doi.org/10.1016/j.cub.2018.01.061>
784 Estella C., R. Voutev, and R. S. Mann, 2012 Chapter seven - A Dynamic Network of Morphogens
785 and Transcription Factors Patterns the Fly Leg, pp. 173–198 in *Transcriptional Switches*
786 *During Development*, edited by Plaza S., Payre F. B. T.-C. T. in D. B. Academic Press.
787 Etter P. D., S. L. Bassham, P. A. Hohenlohe, E. a. Johnson, and W. A. Cresko, 2011 SNP Discovery
788 and Genotyping for Evolutionary Genetics Using RAD Sequencing. *Methods Mol Biol* 772:
789 157–178. <https://doi.org/10.1007/978-1-61779-228-1>
790 Falconer D. S., and T. F. C. Mackay, 1996 Continuous Variation, pp. 100–107 in *Introduction to*
791 *Quantitative Genetics*, Longman, Essex, England.
792 Finlin B. S., C.-L. Gau, G. A. Murphy, H. Shao, T. Kimel, *et al.*, 2001 RERG Is a Novel ras -related,
793 Estrogen-regulated and Growth-inhibitory Gene in Breast Cancer. *J. Biol. Chem.* 276:
794 42259–42267. <https://doi.org/10.1074/jbc.M105888200>
795 Galant R., and S. B. Carroll, 2002 Evolution of a transcriptional repression domain in an insect
796 Hox protein. *Nature* 415: 910–913. <https://doi.org/10.1038/nature717>
797 Garcia-Bellido A., and J. F. de Celis, 2009 The Complex Tale of the achaete-scute Complex: A
798 Paradigmatic Case in the Analysis of Gene Organization and Function During Development.
799 *Genetics* 182: 631–639. <https://doi.org/10.1534/genetics.109.104083>
800 Hannun Y. A., and L. M. Obeid, 2017 Sphingolipids and their metabolism in physiology and
801 disease. *Nat. Rev. Mol. Cell Biol.* 19: 175–191. <https://doi.org/10.1038/nrm.2017.107>
802 Hines H. M., R. Papa, M. Ruiz, A. Papanicolaou, C. Wang, *et al.*, 2012 Transcriptome analysis
803 reveals novel patterning and pigmentation genes underlying *Heliconius* butterfly wing
804 pattern variation. *BMC Genomics* 13: 288. <https://doi.org/10.1186/1471-2164-13-288>
805 Holloway G. J., P. M. Brakefield, and S. Kofman, 1993 The genetics of wing pattern elements in
806 the polyphenic butterfly, *Bicyclus anynana*. *Heredity (Edinb)*. 70: 179–186.
807 <https://doi.org/10.1038/hdy.1993.27>
808 Jombart T., 2008 adegenet: a R package for the multivariate analysis of genetic markers.
809 *Bioinforma.* 24: 1403–5. <https://doi.org/10.1093/bioinformatics/btn129>
810 Jombart T., and I. Ahmed, 2011 adegenet 1.3-1: new tools for the analysis of genome-wide SNP
811 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.* 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*, (P. Wittkopp, Ed.). *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.*, 2009a 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.*, 2009b 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>
- Li X., Y. Wu, C. Shen, T. Y. Belenkaya, L. Ray, *et al.*, 2015 *Drosophila* p24 and Sec22 regulate Wingless trafficking in the early secretory pathway. *Biochem. Biophys. Res. Commun.* 463: 483–489. <https://doi.org/10.1016/j.bbrc.2015.04.151>
- Marcellini S., and P. Simpson, 2006 Two or Four Bristles: Functional Evolution of an Enhancer of scute in *Drosophilidae*, (M. Noor, Ed.). *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>
- 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>
- Meléndez A., W. Li, and D. Kalderon, 1995 Activity, expression and function of a second *Drosophila* protein kinase A catalytic subunit gene. *Genetics* 141: 1507 LP – 1520.
- Merzendorfer H., 2014 ABC Transporters and Their Role in Protecting Insects from Pesticides and Their Metabolites, pp. 1–72 in.
- 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 LP – 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.* 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* (N. Y). 48: 1147. <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* (N. Y).
- Monteiro 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>

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*. <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* (80-.). 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. <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. R. Soc. B Biol. Sci.* 281: 20133262–20133262. <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.* 49: 355–365. <https://doi.org/10.1006/bijl.1993.1041>

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/https://doi.org/10.1016/j.ydbio.2017.06.030>

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>

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. London B Biol. Sci.* 285. <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* (80-.). 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 β catenin Destruction Complex and Exploring Its Regulation and Mechanisms of Action during Development, (M. Katoh, Ed.). *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. <https://doi.org/10.0.4.14/nature716>

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

926 Saccheri I. J., and M. W. Bruford, 1993 DNA Fingerprinting in a Butterfly, *Bicyclus anynana*
927 (Satyridae). *J. Hered.* 84: 195–200.
928 <https://doi.org/10.1093/oxfordjournals.jhered.a111316>
929 Saccheri I. J., P. M. Brakefield, and R. A. Nichols, 1996 evere inbreeding depression and rapid
930 fitness rebound in the butterfly *Bicyclus anynana* (Satyridae). *Evolution* (N. Y). 50: 2000–
931 2013. <https://doi.org/10.1111/j.1558-5646.1996.tb03587.x>
932 Saenko S. V., M. S. Marialva, and P. Beldade, 2011 Involvement of the conserved Hox gene
933 *Antennapedia* in the development and evolution of a novel trait. *Evodevo* 2: 9.
934 <https://doi.org/10.1186/2041-9139-2-9>
935 Santos M. E., I. Braasch, N. Boileau, B. S. Meyer, L. Sauter, *et al.*, 2014 The evolution of cichlid
936 fish egg-spots is linked with a cis-regulatory change. *Nat. Commun.* 5: 5149.
937 <https://doi.org/10.1038/ncomms6149>
938 Schachat S. R., J. C. Oliver, and A. Monteiro, 2015 Nymphalid eyespots are co-opted to novel
939 wing locations following a similar pattern in independent lineages. *BMC Evol. Biol.* 15: 20.
940 <https://doi.org/10.1186/s12862-015-0300-x>
941 Shirai L. T., S. V. Saenko, R. A. Keller, M. A. Jerónimo, P. M. Brakefield, *et al.*, 2012 Evolutionary
942 history of the recruitment of conserved developmental genes in association to the formation
943 and diversification of a novel trait. *BMC Evol. Biol.* 12: 21. [https://doi.org/10.1186/1471-](https://doi.org/10.1186/1471-2148-12-21)
944 2148-12-21
945 Stern D. L., and V. Orgogozo, 2008 The Loci of Evolution: How Predictable is Genetic Evolution?
946 *Evolution* (N. Y). 62: 2155–2177. <https://doi.org/10.1111/j.1558-5646.2008.00450.x>
947 Supple M. A., H. M. Hines, K. K. Dasmahapatra, J. J. Lewis, D. M. Nielsen, *et al.*, 2013 Genomic
948 architecture of adaptive color pattern divergence and convergence in *Heliconius* butterflies.
949 *Genome Res.* 23: 1248–1257. <https://doi.org/10.1101/gr.150615.112>
950 Tien A.-C., A. Rajan, K. L. Schulze, H. D. Ryoo, M. Acar, *et al.*, 2008 *Ero1L*, a thiol oxidase, is
951 required for Notch signaling through cysteine bridge formation of the *Lin12*-Notch repeats
952 in *Drosophila melanogaster*. *J. Cell Biol.* 182: 1113–1125.
953 <https://doi.org/10.1083/jcb.200805001>
954 Tokita C. K., J. C. Oliver, and A. Monteiro, 2013 A Survey of Eyespot Sexual Dimorphism across
955 Nymphalid Butterflies. *Int. J. Evol. Biol.* 2013: 1–6. <https://doi.org/10.1155/2013/926702>
956 Tomoyasu Y., S. R. Wheeler, and R. E. Denell, 2005 Ultrabithorax is required for membranous
957 wing identity in the beetle *Tribolium castaneum*. *Nature* 433: 643–647.
958 <https://doi.org/10.1038/nature03272>
959 Tong X., S. Hrycaj, O. Podlaha, A. Popadic, and A. Monteiro, 2014 Over-expression of
960 Ultrabithorax alters embryonic body plan and wing patterns in the butterfly *Bicyclus*
961 *anynana*. *Dev. Biol.* 394: 357–366. <https://doi.org/10.1016/j.ydbio.2014.08.020>
962 Weatherbee S. D., H. Frederik Nijhout, L. W. Grunert, G. Halder, R. Galant, *et al.*, 1999
963 Ultrabithorax function in butterfly wings and the evolution of insect wing patterns. *Curr.*
964 *Biol.* 9: 109–115. [https://doi.org/10.1016/S0960-9822\(99\)80064-5](https://doi.org/10.1016/S0960-9822(99)80064-5)
965 Werner T., S. Koshikawa, T. M. Williams, and S. B. Carroll, 2010 Generation of a novel wing
966 colour pattern by the Wingless morphogen. *Nature* 464: 1143–1148.
967 <https://doi.org/10.1038/nature08896>
968 Westerman E. L., N. Chirathivat, E. Schyling, and A. Monteiro, 2014 Mate preference for a
969 phenotypically plastic trait is learned, and may facilitate preference-phenotype matching.
970 *Evolution* (N. Y). 68: 1661–1670. <https://doi.org/10.1111/evo.12381>
971 Westerman E. L., N. W. VanKuren, D. Massardo, A. Tenger-Trolander, W. Zhang, *et al.*, 2018
972 *Aristaless* Controls Butterfly Wing Color Variation Used in Mimicry and Mate Choice. *Curr.*
973 *Biol.* 28: 3469–3474.e4. <https://doi.org/10.1016/j.cub.2018.08.051>
974 Yamamoto-Hino M., and S. Goto, 2016 Spätzle-Processing Enzyme-independent Activation of the
975 Toll Pathway in *Drosophila* Innate Immunity. *Cell Struct. Funct.* 41: 55–60.
976 <https://doi.org/10.1247/csf.16002>
977 Zhang L., and R. D. Reed, 2016 Genome editing in butterflies reveals that spalt promotes and
978 Distal-less represses eyespot colour patterns. *Nat. Commun.* 7: 11769.
979 <https://doi.org/10.1038/ncomms11769>
980 Zhou X., and M. Stephens, 2012 Genome-wide efficient mixed-model analysis for association
981 studies. *Nat. Genet.* 44: 821–824. <https://doi.org/10.1038/ng.2310>
982 Zhou X., and M. Stephens, 2014 Efficient multivariate linear mixed model algorithms for

983 genome-wide association studies. Nat. Methods 11: 407–409.
984 <https://doi.org/10.1038/nmeth.2848>
985

Table legends

Table 1. Dorsal hindwing eyespot number (DHEN) is heritable.

Summary of DHEN data for offspring from 6 families each of 0x0, 1x1, and 2x2 DHEN crosses, separated by sex. Offspring with asymmetric DHEN are included in the average DHEN estimate.

Table 2. DHEN candidate genes. 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 of each of the candidate genes to one of our identified SNPs in bp. 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 methods.

Table 3. Genes containing outlier SNPs. BANY.1.2 gene ID, gene name, gene

description and 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 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 acids changes are described. Asterisks in gene IDs represents a gene that overlaps with our DHEN candidates in Table 2.

Figure legends

Figure 1. Eyespot pattern and number variation in *Bicyclus anynana*.

(A.) Eyespot pattern on the ventral side of wings: forewing displays two eyespots; hindwing displays seven eyespots. **(B.)** Eyespots 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 eyespot number (DHEN) variation, ranging from zero to five UV-reflective spots, marked by white arrows (i–vi).

Figure 2. Suitability control for using the lab reared *B. anynana*

population for GWAS. (A.) Principal Component Analysis (PCA) of the allelic variation observed in 5000 randomly-selected genome-wide SNPs in the study population across both phenotype groups, DHE presence (*pre*, red) and DHE absence (*abs*, blue). Ellipses display boundaries of the 95% confidence interval. Little contribution to variation in the principal components and overlap of the variation on both phenotype groups suggests lack of underlying demographic substructuring in the study population. **(B.)** Genome-wide linkage disequilibrium

(LD) in the *B. anynana* study population. Grey dots represent LD values for a SNP pairwise comparison. 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 (470,410bp) and corresponding r^2 value (0.137). **(C.)** Correlation between association ($-\log_{10}(p\text{-value})$) and F_{ST} between study groups. Dots denote the values for association and F_{ST} for all SNPs in the dataset. Black dots represent non-outlier SNPs. Dark and light red dots represent outlier SNPs above the 6 and 7 standard deviation above the mean cutoff, respectively (dark and light blue dashed line).

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) above 6 and 7 standard deviations (SD) above the genome-wide mean. Scaffolds containing both significant association and F_{ST} outliers are marked with asterisks (black for SD > 6 and red for SD > 7). **(B.)** Genomic scaffolds from the *Bicyclus anynana* BANY.1.2 genome are arranged along the 21 chromosomes of the *Heliconius 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.

Figure 4. Zoom-in on 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 above the SD 7 threshold 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 a *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.

Table 1: DHEN is heritable. Summary DHEN data for offspring from 6 families each of 0x0, 1x1, and 2x2 DHEN crosses, separated by sex. Offspring with asymmetric DHEN are included in the average DHEN estimate.

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

Table 2. DHEN candidate genes. 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 of each of the candidate genes to one of our identified SNPs in bp. 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 methods.

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	SD threshold
BANY.1.2.g00030	<i>CDase</i>	Neutral ceramidase	BANY00001	8479	Yes (Özsu & Monteiro, 2017)	N-acylsphingosine amidohydrolase activity	sphingolipid metabolic process; ceramidase activity	SD6
BANY.1.2.g00624	<i>gpr180</i>	Integral membrane G protein coupled receptor	BANY00001	6319	No		Signal transduction	SD6
BANY.1.2.g00636	<i>Dach1</i>	Dachshund homolog 1	BANY00004	267821	No		DNA binding	SD7
BANY.1.2.g00653	<i>dpp</i>	Protein decapentaplegic	BANY00004	28669	Yes (Connahs <i>et al.</i> , 2019)		morphogene activity	SD7
BANY.1.2.g00658	<i>dlg1</i>	Disks large 1 tumor suppressor protein	BANY00004	61933	No but suggested to regulates the Notch signaling pathway which is associated with eyespot development (Reed and Serfas 2004)	epidermal growth factor receptor binding	transcription factor complex; DNA-binding transcription repressor activity, RNA polymerase II-specific; cell population proliferation; Leg disc proximal-distal pattern formation	SD7
BANY.1.2.g00681	<i>numb</i>	Protein numb	BANY00004	222545	No but suggested to regulates the Notch signaling pathway which is associated with eyespot development (Reed and Serfas 2004)	Notch binding	cell fate specification, anterior/posterior pattern formation, imaginal disc development, imaginal disc pattern formation; developmental pigmentation; dorsal/ventral axis specification	SD7
BANY.1.2.g01110	<i>ZC3H10</i>	Zinc finger CCCH domain-containing protein 10	BANY00007	0 (coding sequence non-synonymous [Ile142 -> Val])	No	metal ion binding, RNA binding	asymmetric cell division; regulation of Notch signaling pathway	SD6
BANY.1.2.g02164	<i>Taf2</i>	Transcription initiation factor TFIID subunit 2	BANY00015	16341	No but suggested to regulates the Notch signaling pathway which is associated with eyespot development (Reed and Serfas 2004)	chromatin binding	posttranscriptional regulation of gene expression	SD6
BANY.1.2.g02292	<i>PLEKHG4B</i>	Pleckstrin homology domain-containing family G member 4B	BANY00017	78770	No	Rho guanine-nucleotide exchange factor activity	DNA-templated transcription, initiation; positive regulation of transcription of Notch receptor target	SD6
BANY.1.2.g02300	<i>Pka-C3</i>	cAMP-dependent protein kinase catalytic subunit 3	BANY00017	174351	No, but imaginal disc derived wing morphogenesis (Meléndez <i>et al.</i> 1995)	ATP binding, c-AMP dependent protein kinase activity	regulation of Rho protein signal transduction	SD6
BANY.1.2.g02301	<i>bai</i>	Transmembrane emp24 domain-containing protein bai	BANY00017	197270	No but suggested involved in Wnt protein secretion, and Wg has been implicated functionally in eyespot size regulation (Özsu <i>et al.</i> 2017). Also dorso/ventral pattern formation in <i>Drosophila</i> (Bartoszewski <i>et al.</i> 2004)	transporter activity	imaginal disc-derived wing morphogenesis	SD6
BANY.1.2.g02571	<i>Antp</i>	Homeotic protein antennapedia	BANY00019	276092	Yes (Saenko, Marialva, & Beldade, 2011)	DNA binding	dorsal/ventral pattern formation; Wnt protein secretion	SD6
BANY.1.2.g02579	<i>Ubx</i>	Homeotic protein ultrabithorax	BANY00019	88613	Yes (Weatherbee 1999; Tomoyasu <i>et al.</i> 2005; Tong <i>et al.</i> 2014)	DNA binding	anterior/posterior axis specification; specification of segmental identity	SD6
BANY.1.2.g02585	<i>abd-A</i>	Homeobox protein abdominal-A homolog	BANY00019	401028	No	DNA binding	multicellular organism development; regulation of transcription; anterior/posterior pattern specification	SD6
BANY.1.2.g04718	<i>TRAIP</i>	E3 ubiquitin-protein ligase TRAIP	BANY00042	4838	No		multicellular organism development; regulation of transcription, anterior/posterior pattern specification	SD6
BANY.1.2.g04910	<i>spz</i>	Protein spaetzle	BANY00044	18317	No but receptor Toll has been implicated in eyespot development (Özsu & Monteiro, 2017)	morphogene activity, Toll binding	apoptotic process; cell population proliferation; signal transduction	SD7
BANY.1.2.g05426	<i>hhex</i>	Hematopoietically-expressed homeobox protein HHEX homolog	BANY00052	17500	No	DNA binding transcription factor activity	regulation of Toll signaling pathway	SD7
BANY.1.2.g08887	<i>tlf1</i>	Toll-like protein 1	BANY00105	0 (Intronic 1 SNP)	No, but dorsal/ventral pattern formation in zebrafish (Wagner <i>et al.</i> 2002)	calcium ion binding	multicellular organism development; regulation of transcription	SD6
BANY.1.2.g09546	<i>Sal13</i>	Sal-like protein 3; Spalt	BANY00119	135802	Yes (Brunetti <i>et al.</i> 2001; Monteiro <i>et al.</i> 2006)	nucleic acid binding	cell differentiation; multicellular organism development	SD6
BANY.1.2.g10814	<i>rergrl</i>	Ras-related and estrogen-regulated growth inhibitor-like protein	BANY00148	214847	No but Ras signaling and 20E signaling have been implicated in eyespot development (Özsu & Monteiro, 2017; Bhardwaj <i>et al.</i> 2018)	GTP binding	limb morphogenesis	SD6
BANY.1.2.g10819	<i>Ero1L</i>	Ero1-like protein	BANY00148	156666	No but suggested to regulates the Notch signaling pathway which is associated with eyespot development (Reed and Serfas 2004)	oxidoreductase activity	signal transduction	SD6
BANY.1.2.g11151	<i>mbo</i>	Nuclear pore complex protein Nup88	BANY00156	0 (coding sequence non-synonymous [Leu584 -> Pro])	No	chromatine binding	chaperone cofactor-dependent protein refolding; regulation of Notch signaling pathway	SD7
BANY.1.2.g11298	<i>Eif4g3</i>	Eukaryotic translation initiation factor 4 gamma 3	BANY00158	39941	No	RNA binding	mRNA export from nucleus; protein export from nucleus	SD7
BANY.1.2.g11326	<i>msh</i>	Muscle segmentation homeobox	BANY00159	246092	No but dorsal/ventral pattern formation in <i>Drosophila</i> wing (Milán <i>et al.</i> 2001)	DNA binding	positive regulation of meiosis I; positive regulation of protein phosphorylation;	SD7
BANY.1.2.g12184	<i>Mi-2</i>	Chromodomain-helicase-DNA-binding protein Mi-2 homolog	BANY00183	0 (Intronic 4 SNPs)	No	ATP binding, chromatin binding, DNA binding	dorsal/ventral pattern formation; ectoderm development; embryonic morphogenesis; imaginal disc-derived wing morphogenesis; wing and notum subfield formation; wing disc pattern formation	SD6
BANY.1.2.g12183	<i>hd</i>	Protein downstream neighbor of son homolog (Protein humpty dumpty)	BANY00183	3204	No		chromosome decondensation; chromosome organization	SD6
BANY.1.2.g12275	<i>Cd63</i>	CD63 antigen	BANY00186	0 (Intronic 1 SNP)	Yes (Özsu & Monteiro, 2017)	protein-containing complex binding	cell division; cell population proliferation; multicellular organism development; nuclear DNA replication	SD7
BANY.1.2.g13873	<i>ZNF845</i>	Zinc finger protein 845	BANY00235	10819	No		cell-matrix adhesion; cell migration; cell surface receptor signaling pathway; epithelial cell differentiation; pigment granule maturation	SD6
BANY.1.2.g16720	<i>14-3-3epsilon</i>	14-3-3 protein epsilon	BANY00344	63487	No but this protein is a component of the Ras signaling pathway, which has been implicated in eyespot development (Özsu & Monteiro, 2017). Also, wing disc dorso/ventral pattern formation in <i>Drosophila</i> (Bejarano, <i>et al.</i> 2008)	transcription factor binding; protein domain specific binding	DNA-binding transcription factor activity	SD6
BANY.1.2.g16724	<i>ct</i>	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 development; wing disc dorsal/ventral pattern formation; regulation of growth; Ras signaling	SD6
BANY.1.2.g19639	<i>Roc1a</i>	RING-box protein 1A (Regulator of cullins 1a)	BANY00502	3407	No but a regulator of Wnt signaling, and Wg has been implicated functionally in eyespot size regulation (Özsu <i>et al.</i> 2017)	zinc ion binding;	imaginal disc-derived wing margin morphogenesis; formation of a compartment boundary; negative regulation of Notch signaling pathway	SD6
BANY.1.2.g20777	<i>Sxt7</i>	Syntaxin-7	BANY00593	61290	Yes (Özsu & Monteiro, 2017)	syntaxin binding	cell population proliferation; lipid storage; negative regulation of canonical Wnt signaling pathway	SD6
BANY.1.2.g21276	<i>w</i>	Protein white	BANY00641	0 (Intronic 1 SNP)	No but D. melanogaster ommochrome transporters (Merzendorfer 2014)	pigment binding	intracellular protein transport	SD6
BANY.1.2.g21277	<i>st</i>	Protein scarlet	BANY00641	37688	No but D. melanogaster ommochrome transporters (Merzendorfer 2014)	pigment binding	ommochrome biosynthetic process; integral component of pigment granule membrane	SD6
BANY.1.2.g25219	<i>AnxB9</i>	Annexin B9	BANY01600	1408	No, but wing disc dorso/ventral pattern formation (Bejarano, <i>et al.</i> 2008)	actin binding	ommochrome biosynthetic process; transmembrane transport; integral component of pigment granule membrane	SD6

Table 3. Genes containing outlier SNPs. BANY.1.2 gene ID, gene name, gene description and 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 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 acids changes are described. Asterisks in gene IDs represents a gene that overlaps with our DHEN candidates in Table 2.

BANY.1.2 gene ID	Gene name	Gene description	BANY.1.2 scaffold ID	Number of Variants	Annotated variants	Function	Biological Process	SD Threshold
BANY.1.2.g00028	<i>STRADA</i>	STE20-related kinase adapter protein alpha	BANY00001	1	Intronic (Intron 1)	ATP binding; protein kinase	MAPK cascade	SD6
BANY.1.2.g00029	<i>TKFC</i>	Triokinase/FMN cyclase	BANY00001	1	Intronic (Intron 1)	ATP binding; metal ion binding	innate immune response; MDA-5 signalling pathway	SD6
BANY.1.2.g00671	<i>CG4341</i>	Protein O-mannosyl-transferase	BANY00004	1	Intronic (Intron 1)	mannosyltransferase activity	--	SD7
BANY.1.2.g01110 *	<i>ZC3H10</i>	Zinc finger CCCH domain-containing protein 10	BANY00007	2	Synonymous (Exon 2); Non-Synonymous (Exon 2: Ile142 -> Val)	metal ion binding, RNA binding	posttranscriptional regulation of gene expression	SD6
BANY.1.2.g02168	<i>kh</i>	Kynurenine 3-monooxygenase	BANY00015	2	2 Intronic (Intron 4)	FAD binding; NAD(P)H oxidase	NAD metabolism	SD6
BANY.1.2.g04716	Unknown	Unknown	BANY00042	1	Synonymous (Exon 2)	--	--	SD7
BANY.1.2.g04911	Unknown	Unknown	BANY00044	3	3 Intronic (Intron 1)	--	--	SD7
BANY.1.2.g08887 *	<i>tlf1</i>	Tolloid-like protein 1	BANY00105	1	Intronic (Intron 4)	calcium ion binding; zinc ion binding	cell differentiation	SD6
BANY.1.2.g11151 *	<i>mbo</i>	Nuclear pore complex protein <i>Nup88</i>	BANY00156	1	Non-Synonymous (Exon 14: Leu584 -> Pro) (1)	chromatine binding	mRNA export from nucleus; protein export from nucleus	SD7
BANY.1.2.g11296	<i>APBB1P</i>	APBB1-interacting protein 1	BANY00158	1	Intronic (Intron 2)	--	signal transduction	SD7
BANY.1.2.g11300	<i>Tret1</i>	Facilitated trehalose transporter <i>Tret1</i>	BANY00158	4	3 Intronic (Intron 9); 2 Synonymous (Exon 9)	trehalose transmembrane transporter activity	trehalose transport	SD7
BANY.1.2.g11312	<i>idhb-1</i>	Isocitric dehydrogenase subunit beta	BANY00159	3	Intronic (Intron 1); 2 Synonymous (Exon 5)	isocitrate dehydrogenase activity; magnesium ion binding; NAD binding	tricarboxylic acid cycle	SD7
BANY.1.2.g12184 *	<i>Mi-2</i>	Chromodomain-helicase-DNA-binding protein <i>Mi-2</i> 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	SD6
BANY.1.2.g12186	Unknown	Unknown	BANY00183	3	3 Intronic (Intron 4)	--	--	SD6
BANY.1.2.g12187	<i>spt16</i>	FACT complex subunit <i>spt16</i>	BANY00183	1	Synonymous (Exon 12)	chromatine binding	transcription elongation; DNA replication	SD6
BANY.1.2.g12189	<i>TRIM37</i>	E3 ubiquitin-protein ligase	BANY00183	6	5 Intronic (Intron 7); Synonymous (Exon 7)	chromatine binding	regulation of RNA polymerase binding	SD6
BANY.1.2.g12272	<i>PPIL2</i>	RING-type E3 ubiquitin-protein ligase	BANY00186	1	Intronic (Intron 2)	ubiquitin protein ligase activity	protein folding; protein localization	SD7
BANY.1.2.g12275	<i>Cd63</i>	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	SD7
BANY.1.2.g12279 *	<i>Cd53</i>	Leukocyte surface antigen CD53	BANY00186	1	Synonymous (Exon 2)	--	cell surface receptor signaling pathway	SD7
BANY.1.2.g13875	<i>Fibp</i>	aFGF intracellular-binding protein	BANY00235	1	Intronic (Intron 2)	fibroblast growth factor binding	--	SD6
BANY.1.2.g16724 *	<i>ct</i>	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	SD6
BANY.1.2.g19628	<i>LanB1</i>	Laminin subunit beta-1	BANY00502	1	Intronic (Intron 8)	--	substrate adhesion; cell migration	SD6
BANY.1.2.g19640	Unknown	Unknown	BANY00502	2	2 Intronic (Intron 7)	--	--	SD6
BANY.1.2.g20774	<i>PAN3</i>	PAN2-PAN3 deadenylation complex subunit	BANY00593	2	2 Intronic (Intron 4)	ATP binding; RNA binding	RNA processing	SD6
BANY.1.2.g21276 *	<i>w</i>	Protein white	BANY00641	1	Intronic (Intron 7)	pigment binding	ommochrome biosynthetic process; integral component of pigment granule membrane	SD6

A

bioRxiv preprint doi: <https://doi.org/10.1101/063101>; this version posted May 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B

(i)

(ii)

C

(i)

(ii)

(iii)

(iv)

(v)

(vi)

Bicyclus anynana

Heliconius melpomene

Heliconius melpomene