Convergent molecular evolution associated with repeated transitions to gregarious larval behaviour in Heliconiini

Authors: Francesco Cicconardi[1*], Callum F. McLellan[1*], Alice Seguret[1], W. Owen McMillan[2], Stephen H. Montgomery[1,2]

Affiliations:

[1] School of Biological Sciences, University of Bristol, Bristol, United Kingdom;

[2] Smithsonian Tropical Research Institute, Gamboa, Panama

* *contributed equally*

Correspondence:

SHM: s.montgomery@bristol.ac.uk

CFM: cm15245@bristol.ac.uk

FC: f.cicconardi@bristol.ac.uk

Lead author:

SHM: s.montgomery@bristol.ac.uk

ORCiD:

CFM: 0000-0001-6257-4569

FC: 0000-0001-6509-6179

AS: 0000-0001-5159-5815

WOM: 0000-0003-2805-2745

SHM: 0000-0002-5474-5695

Abstract

Collective behaviour forms the basis for many anti-predator strategies. Within Lepidoptera, larval gregariousness has evolved convergently across many phylogenetically disparate lineages. While the selection pressures shaping variation in larval social behaviours are well investigated, much less is known about the mechanisms that control social attraction and behavioural coordination. Similarly, little is known about how secondary selection pressures associated with social living shape genome evolution. Here, using genomic data for over 60 species from an adaptive radiation of Neotropical butterflies, the Heliconiini, in which gregarious behaviour has evolved repeatedly, we explore the molecular basis of repeated convergent shifts towards gregarious larvae. We focus on three main areas of genomic evolution: differential selection on homologous genes, accelerated rates of evolution on non-coding regions of key genes, and differential gene expression in the brains of solitary and gregarious larvae. We identify strong signatures of convergent molecular evolution, on both coding and non-coding loci, in Heliconiini lineages which evolved gregarious behaviour. Molecular convergence is also detected at the transcriptomic level in larval brains, suggesting convergent shifts in gene regulation in neural tissue. Among loci showing strong signals of convergent evolution in gregarious lineages, we identify several strong candidates linked to neural activity, feeding behaviour, and immune pathways. Our results suggest sociality profoundly changes the selection pressures acting on multiple physiological, immunological and behavioural traits.

Introduction

The convergent evolution of similar phenotypes has been a focus of biological research for over a century. Modern sequencing tools have recently allowed the exploration of the genetic bases behind convergent evolution, greatly improving our understanding of the phylogenetic relationships between complex traits such as behaviours[1,2]. Molecular changes associated with the evolution of one such behaviour, sociality, have received a great deal of attention. Across taxa, individuals living as part of a group are likely to face similar trade-offs, which potentially involve similar molecular functions[2,3-5]. These may include both a molecular response to group living, and the molecular control of the behaviour itself.

Social behaviour can impact a range of selection regimes, which likely shape patterns of genetic divergence and gene expression between solitary and social species. Aggregating introduces stresses like crowding and an increased risk of starvation and pathogen infection[6-10], driving changes in gene expression to mitigate these costs[10,11]. Group-living animals may be driven to develop more elaborate neural circuitry to process frequent, and potentially complex, social encounters and sensory information in the form of communication signals from group members[12]. Some molecular functions may also be inherently better suited to producing the behavioural phenotype[2-4], such as those which can be altered without negative pleiotropic effects, or only require slight modifications to produce the new phenotype[3,4]. As such, convergent behavioural evolution may involve similar mechanisms, particularly among close relatives, as the sensory cues available, sensory systems and downstream neural pathways may have similar biases and variation, and these molecular pathways may require relatively fewer modifications to produce a complex trait such as social behaviour.

To examine the extent that convergent behavioural changes involve similar molecular mechanisms we focus research on larval gregariousness in Lepidoptera. These larvae vary widely in the size, longevity and nature of their social groups, and exhibit trail-following and synchronised defensive actions that require group coordination[13-16]. Allied to these behavioural traits is huge diversity in morphology, hostplant ecology, life history and development. This diversity positions larval Lepidoptera as an informative window into the convergent evolution of group living. Gregariousness behaviour in larvae has evolved convergently across many phylogenetically disparate lineages of Lepidoptera[17-19]. Yet, to date, research into the evolution of social behaviour in larval Lepidoptera has focused almost completely on ultimate causes and ecological contexts[20-23], with little consideration for the genetics driving the behaviour. This previous work highlights the phenotypic and environmental conditions that are important for gregariousness to evolve[19,23,24], yet almost

nothing is known of the genomic alterations required to produce the behaviour, or the effect that the behaviour-induced selection regimes have on the genome.

Here, we focus on the Heliconiini, a diverse tribe of Neotropical butterflies across which larval gregariousness has evolved independently on at least seven separate occasions[24,25], as a potentially powerful system in which to investigate the molecular basis of social behaviour. Using genomic data for 53 species in the Heliconiini tribe and additional closely related taxa[26], we conduct the first phylogeny-wide study testing whether convergence in social behaviours is underpinned by a molecular convergence and shared selection regimes. We examine shifts in selection pressure on orthologous genes across Heliconiini lineages with gregarious and solitary larvae, as well as testing for accelerated rates of evolution in non-coding regions associated with key loci. Finally, we compare gene expression profiles of three Heliconiini species with gregarious larvae, from three separate clades, in relation to three related species with solitary larvae, creating three monophyletic pairs of social and solitary species.

Methods

i) Detecting signatures of selection and convergence associated with the gregarious phenotype

To detect a signal of association between the gregarious phenotype with a shift in selective pressure (d*N*/d*S* or ω) across protein-coding genes (PCGs), we used the 3393 single-copy orthologous groups (scOG) list across the Heliconiinae phylogeny from Cicconardi *et al.*[26], and a set of computational approaches designed to explore the selective landscape operating across the genome.

The first method, Branch-site Unrestricted Statistical Test for Episodic Diversification for PHenotype (BUSTED-PH), is a method to test for evidence of episodic diversifying selection associated with a specific feature/phenotype/trait[27,28]. The method first tests whether both background branches and test branches experience episodic positive selection, then checks if there is a difference between the two partitioned ω. We specifically tested whether there is a difference between lineages associated with the gregarious phenotype *vs* all solitary lineages. The resulting *p* values were corrected for multiple testing with the Benjamini-Hochberg method ("BH" aka "FDR") as implemented in the R package P.ADJUST {STATS}, using an alpha threshold of 0.05. Because the within group test of episodic selection gives more stratified results rather than just differentially selected or not, we divided scOGs into two groups, *associated with gregariousness* if: there is a difference of selection between gregarious and non-gregarious (Adjusted $p < 0.05$), but not if episodic

positive selection occurred in both gregarious and non-gregarious ($p < 0.01$); and *not associated with gregariousness* for the other cases. In other words, we exclude genes under selection in both gregarious and non-gregarious lineages. The two groups were plotted comparing their reciprocal gregarious ω vs. solitary ω, and their relative scaling coefficients and intercepts were analysed with SMATR[29].

We also tested whether scOGs experienced relaxation or intensification by comparing gregarious vs solitary lineages with RELAX[30], as implemented in the HYPHY batch language[28]. More specifically, RELAX tests whether selection pushes all ω categories away from neutrality, *i.e.* intensification ($1 > \omega > 1$), or towards neutrality, *i.e.* relaxation ($\omega = 1$). A k parameter is computed, which is lower than one ($k < 1$) in cases where the model supports the relaxed scenario or higher than one ($k > 1$) where intensification is better supported. Also here, p values were corrected for multiple testing with the Benjamini-Hochberg method, and p adjusted < 0.05 were considered as significant.

Finally, we used CSUBST[31] to derive observed (O_C^N and O_C^S) and expected (E_C^N and E_C^S) numbers of non-synonymous and synonymous convergence and evaluate their rates (dN_C and dS_C), ω_C metric, in branch combinations in a phylogenetic tree. We ran the analysis twice, once accepting all convergent events that occurred only in gregarious lineages with a threshold of $\omega_C > 1.0$ and $dN_C > 1.0$, as an overall signal of convergence, and again with a more stringent threshold of $\omega_C > 5.0$ and $dN_C > 2.0$ for an *arity* ≥ 2 (degree of combinatorial substitutions or the number of branches to be compared) to be considered convergence. Other statistical tests such as the one-sided Wilcoxon rank-sum tests were used to test for differences in the distributions, and the Fisher-exact tests to test whether the overlap between gene sets deviates from a random distribution. Both statistical tests were run as implemented in R and SCIPY.STATS.

ii) Annotating conserved non-exonic elements

To identify conserved non-exonic elements (CNEEs), we used the 63-way whole genome alignment generated by Cicconardi *et al.*[26], together with the model of neutral evolution computed with PHAST V1.4 package[32,33]. Because the majority of gregarious lineages belong to the genus *Heliconius,* we recomputed the conserved and accelerated models using PHASTCONS with the *H. melpomene* genome as reference; combining the conserved and non-conserved models with PHYLOBOOTS and using the averaged models to predict conserved elements in PHASTCONS, using default parameters. After the initial estimation of conserved elements, we merged regions within 5 bp of each other into a single conserved element. We then excluded regions: *i*) shorter than 50 bp; or *ii*) with data for less than 50 species for that region; and *iii*) with 50% of gaps in the consensus sequence; following the protocol used in Cicconardi *et al.*[26]. We used PHYLOACC-GT[34], which computes the

183 maximum a posteriori (MAP) Z matrix (matrix of latent states), and two Bayes factors to test
184 for acceleration at the *Heliconius* stem. A Bayes Factor 1 (BF1), defined as the Bayes factor
185 that compares a null model (no acceleration allowed on any branch) to the test branch model
186 (acceleration allowed only on *Heliconius* branches) (M1); and a Bayes Factor 2 (BF2)
187 defined as the Bayes factor comparing the test model to the full model (acceleration allowed
188 on any branch) (M2). From all accelerated CNEEs (aCNEEs) under the M1 model we
189 defined as *strict* those which affect three different clades of gregarious species (*e.g.*: Doris,
190 Sara/Sapho and *Dione julia*). We also used PHYLOP from PHAST V1.4 package[32,33] [--mode
191 CONACC --wig-scores --method LRT] to generate the acceleration/conservation scores.
192
193 ***iii) Enrichment among CNEEs***
194 To test for gene ontology (GO) terms of functional elements enriched in Gregarious-
195 accelerated CNEEs, we used the approaches as implemented in Cicconardi *et al*.[26], *i*) two
196 permutation approaches to account for possible biases towards particular gene functions; *ii*)
197 a genomic fraction approach as implemented in GREAT[35]. To each gene in the reference
198 genome annotation (*H. melpomene*) was assigned a regulatory domain as the 5 + 1 kb with
199 extension strategy, similarly to GREAT[35]. In brief, to each coding locus a regulatory domain is
200 assigned consisting of a basal domain that extends 5 kb upstream and 1 kb downstream
201 from its transcription starting site (TSS), which may overlap between other basal domains of
202 flanking loci, plus a further extension up to the basal domain of the nearest upstream and
203 downstream locus, up to 1 Mb.
204 The "simple permutation test" was computed calculating the expected probability of
205 overlap between the regulatory domains of genes of a specific GO term and 10,000 random
206 aCNEEs datasets, shuffling the CNEEs and randomly selecting the same amount of
207 accelerated CNEEs among all CNEEs, using a binomial test to generate a *p* value. For the
208 second test "Genomic region test", similarly to GREAT, the binomial test was executed over
209 the total fraction of genomic regions associated with a given GO term. A third method was
210 implemented "Reshuffling permutation set" in which BEDTOOLS SHUFFLE [-excl gff3 -chrom -
211 chromFirst -noOverlapping] was adopted to shuffle all detected aCNEEs across the entire
212 genome avoiding overlaps with annotated genes and comparing the overlap with putative
213 regulatory elements, in a binomial test framework. Multiple test correction was done with the
214 Benjamini–Hochberg FDR (FDR_bh) and Bonferroni as implemented in python library
215 STATSMODELS.STATS.MULTITEST. With both approaches we tested all biological processes
216 where at least two aCNEEs were overlapping with a regulatory domain of the given GO
217 term. We used REVIGO[36] to summarize results.
218 In the same fashion we also looked for genomic regions, gene annotations and gene
219 sets (BUSTED-PH) enriched for aCNEEs. To do this, we generated non-overlapping 100 kb

220 sliding windows and computed the probability of observing aCNEEs based on the binomial
221 distribution over the 10,000 permutations of randomly selected accelerated CNEEs among
222 all CNEEs, where the number of trials is the number of aCNEEs in the window and the
223 probability of success is the average of the permutations; the excess of aCNEEs per
224 regulatory domain using the distribution over the 10,000 permutations as implemented in
225 Cicconardi *et al.*[26]; and the excess of aCNEEs to the gene of differentially selected scOGs
226 (BUSTED-PH) over 10,000 permutations considering as background all the genomic
227 regions/regulatory domains of all the tested scOGs. The *p* values of those tests where then
228 corrected as before.
229
230 ***iv) Larval rearing and sample collection***
231 Six species from the Heliconiini tribe were used: *Dione juno, Heliconius doris* and *H. sara*,
232 which are all gregarious as larvae, and *Agraulis vanillae, H. hecale* and *H. erato*, which are
233 solitary. These six species belong to three distinct monophyletic pairs of gregarious and
234 solitary species and were used to contrast gregarious lineages to solitary close relative
235 species. Although behavioural variation exists among gregarious species, we selected these
236 species for comparison because all three remain gregarious up to, and often including,
237 pupation[37-39]. Stock populations of each species were established from wild caught females
238 collected around Gamboa, Panama, in 2013. All larvae are therefore first or second
239 generation, insectary reared individuals, with the exception of the locally less abundant *D.*
240 *juno*, which was opportunistically sampled from a single brood found as eggs on a *Passiflora*
241 *vitifolia* in the outdoor host plant stocks. Stock populations were held in ambient, tropical
242 conditions outside in ~ 2 × 2 × 3 m cages, were fed on artificial feeders, and for *Heliconius*
243 *spp* with ~20% sugar/bee pollen solution with supplemented floral resources (*Lantana sp.*
244 and *Psiguria sp.*). Each species had access to their preferred host plant (*D. juno* and *H.*
245 *hecale*: *Passiflora vitifolia*; *H. doris*: *P. ambigua*; *H. sara*: *P. auriculata*; *A. vanillae* and *H.*
246 *erato*: *P. biflora*). Eggs were collected daily, and larvae were subsequently reared on shoots
247 in pop-up cages at ambient conditions. All species were reared concurrently so were
248 exposed to common environmental conditions. Larvae were sampled in the early 5th instar,
249 before the prepupal instar, or "wandering stage", where behaviour often changes drastically
250 as larvae cease foraging and search for a pupation site (*e.g.*: Smallegange *et al.*)[40]. We used
251 five separate individual samples per species, giving a total of 30 samples. For the gregarious
252 species, *H. doris* and *H. sara* samples were taken from individuals from four separate
253 broods, whereas all five *D. juno* samples were taken from the same brood. Solitary species
254 were selected from the same generation, but as eggs were derived from stock cages with
255 multiple fertilised females, the kinship between individuals was not known. The brain and

connecting suboesophageal ganglion were dissected out in RNA*later* (which stabilises RNA, preventing its breakdown; ThermoFisher Scientific), and stored at -20°C.

v) RNA extraction and sequencing

We extracted total RNA from each sample consisting of an intact, attached brain, and suboesophageal ganglion following the instructions of the RNeasy micro kit according to the manufacturer's protocol (Qiagen). RNA concentration, purity and integrity were assessed on an Agilent 4150 TapeStation (Bristol Genomics Facility, UK). To improve 28S RNA peak visibility, samples were not subjected to heat denaturation treatment, following the suggested approach for assessing insect RNA[41]. All samples had an RNA Integrity Number (RIN) of >= 9, indicating that the RNA was of suitable quality for sequencing[42]. Samples were sent to Novogene (UK) for library preparation (with poly A enrichment) and transcriptome sequencing. Libraries were sequenced on an Illumina NovaSeq 6000 platform, producing approximately 150 bp paired-end reads. We quality checked raw reads using FASTQC v. 0.11.9[43] and trimmed them using TRIMMOMATIC v. 0.39[44] to remove adapter contamination, low quality reads, and reads shorter than 75 bp, using default command line settings.

vi) Differential gene expression analysis

Annotated genomes for all six Heliconiini species were obtained from Cicconardi *et al.*[26]. We used STAR v. 2.7.11a[45] to map trimmed reads to genomes and HTSEQ v. 2.0.3[46] to count the mapped reads per gene. A table of Heliconiini orthologous groups (OGs) was obtained from Cicconardi *et al.*[26] and filtered to retain only scOGs containing genes for all of our target species (total retained genes = 6885). We then matched the read counts of genes to their corresponding OGs in order to compare OG expression across the study species.

All analyses of differential gene expression were performed in R v. 4.1.2[47] using LIMMA v. 3.5.3[48]. Raw read counts were first normalised based on gene length, library size and GC content using CQN v. 1.4.0[49], to account for potential bias due to differences in homologous gene lengths between species[50].

We compared gene expression between species in three separate pairs based on larval social behaviour and clade, where each pair consisted of a gregarious and solitary species belonging to the same clade: i) *D. juno* vs *A. vanillae* (diverged ~14 mya; ii) *H. doris* vs *H. hecale* (diverged ~9 mya), and iii) *H. sara* vs *H. erato* (diverged ~5 mya)[26]. Additionally, we performed a full comparison of all three solitary species *vs* all three gregarious species. Differentially expressed genes (DEGs) were selected using adjusted *p* value threshold of 0.05. Of these, we selected only those with a minimum $\log_2$ fold change ($\log_2$FC) of ±1 to account for any differences in brain tissue scaling (overall size, region volumes etc.) between species[51].

vii) Functional enrichment and overlap of DEGs

A list of GO terms corresponding to our OGs was obtained from Cicconardi *et al.*[26], providing information on the biological function towards which each DEG is likely to contribute[52]. These GO terms were used for functional enrichment analyses, using the R package topGO v. 2.46.0[53] on each species pair, with the aim of highlighting the functions that are overrepresented in our gene lists of interest, both down and up regulated.

Finally, we examined the overlap of DEGs between the three species-pair comparisons to identify convergent shifts in expression. We used 10,000 iteration permutation tests to assess the probability that the number of shared DEGs between the two- and three-way species-pair comparisons is greater than expected by chance. In detail, n genes were sampled without replacement from the full list of OGs, where *n* = the number of observed DEGs from a given species pair, to create a list of random DEGs. The number of shared genes between both (or all three) randomised lists was recorded, and the process repeated over 10,000 iterations to give a distribution of 10,000 sets of shared genes. We then tested the observed count of shared DEGs against this distribution for the probability of it occurring by chance.

Results and Discussion

i) Convergent shifts in selection regimes shapes coding genes in gregarious species.

Using a set of single-copy orthologous groups (scOGs) obtained from 53 Heliconiinae species (Figure 1)[26], we first explored the strength of convergent evolution among protein coding genes (PCGs) between gregarious species. To do so, we adopted three analytical approaches: *i*) using BUSTED-PH[27,28] we tested for evidence of different selective regimes between solitary and gregarious lineages; *ii*) using RELAX[30] we tested whether these shifts in selection regime can be explained with a relaxed selection pressure (a shift of rate-classes towards neutrality; $k < 1$), or with an intensification of selection pressure (a shift of rate-classes towards purifying and positive selection; $k > 1$); finally, *iii*) we scanned scOGs for signs of convergent substitution with CSUBST, a method for detecting the Combinatorial SUBSTitutions of codon sequences[31]. The CSUBST algorithm scans all combinations of target lineages from pairwise (*arity* = 2) to multiple (*arity* > 2) comparisons to identify genes that are under convergent positive selection ($\omega_C > 1.0$).

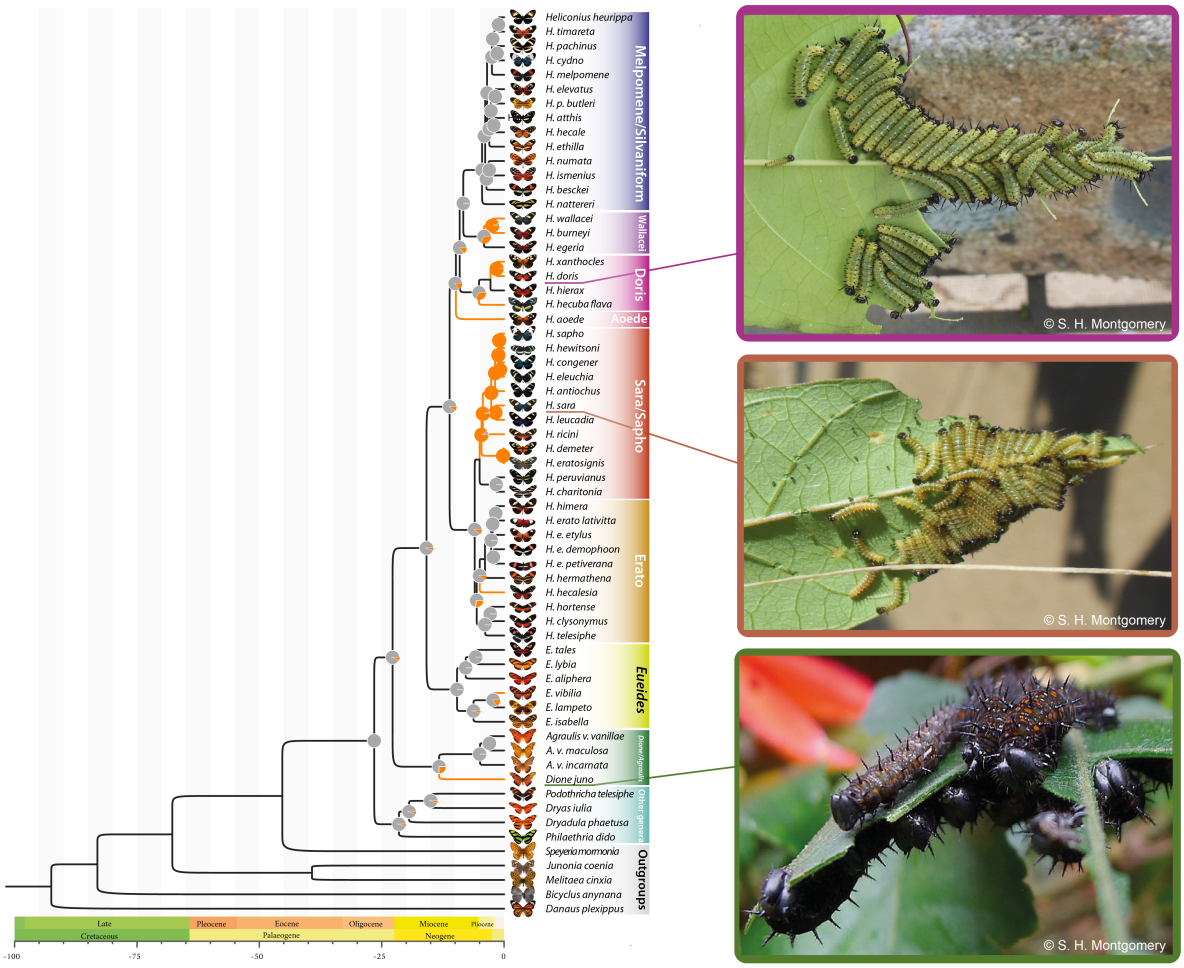

Figure 1. Heliconiini phylogeny and ancestral state reconstruction of gregarious behaviour. (A) Dated phylogeny of Heliconiini species from Cicconardi *et al.*[26]. Orange branches indicate lineages having gregarious behaviour, while pie charts indicate the maximum likelihood estimate of the most recent common ancestor to have gregarious larvae from McLellan and Montgomery[24]. (B) Photographs illustrating independently evolved larval social groups in (from top to bottom) *Heliconius doris*, *Heliconius sara*, and *Dione juno*.

Our BUSTED-PH[27,28] analysis detected 203 scOGs (6%) to be under distinct selection regimes in gregarious lineages (adjusted p values < 0.05; Figure 2A,F; Table S1). While the median strength of selection acting on scOGs in solitary (ω_s) and gregarious lineages (ω_g) is only marginally different overall (0.072 for not differentially selected genes (nDSGs) vs 0.066 for DSGs; One-sided Wilcoxon rank-sum test p adj = 0.047), we detect a much higher median ω_g for those genes with distinct selection regimes between solitary and gregarious lineages (0.12 vs 0.16; One-sided Wilcoxon rank-sum test p adj 2.2×10^{-16}; Figure 2A). This is associated with a proportion of nucleotides evolving under strong purifying selection in the solitary lineages, which shift towards neutrality or positive selection in gregarious lineages (Figure 2A, insert). Our RELAX analysis supports a role for increased neutrality in a class of scOGs in gregarious lineages, with 36% (1224 scOGs) showing a signature or relaxation

350 (Table S1). Among this class of genes, no gene ontology (GO) terms are significantly
351 enriched but several are associated with the response to viruses, including a *Lachesin*-
352 homolog gene and other ubiquitine-related genes (*e.g. sina*, an E3 ubiquitin-protein ligase).
353 In other insects, viruses have been demonstrated to interact with *lachesin* proteins during
354 their attachment to neurons[54], and ubiquitination is known to play an important role in viral
355 infection[55]. In *Bombyx mori* a *sina*-homolog is downregulated to inhibit cellular interactions
356 with viral proteins (*e.g. BmNPV*)[56]. A smaller proportion of scOGs also show evidence of
357 intensified positive selection in gregarious lineages (~4%, 122 scOGs; p adj < 0.01; Figure
358 2C, F). Among these genes, an odorant receptor (OR43; ortholog of the *B. mori* OR49[57]) is
359 known to bind to the antimicrobial and larval food attractant compound citral[58,59].
360 Our CSUBST[31] analysis next sought to identify convergent substitutions in gregarious
361 lineages. We first assessed the degree of convergent selection in genes previously identified
362 as evolving under differential selection (BUSTED-PH), relaxation or intensification (RELAX). To
363 do so, we selected a relatively low threshold for convergence of selection (ω_C > 1.0) and rate
364 of non-synonymous convergence (dN_C > 1.0). Genes under differential selection showed a
365 strong increase in both ω_C and the rate of non-synonymous convergent substitutions
366 (medians: 325.9 ω_C and 11.6 dN_C) compared with other scOGs, where no differential
367 selection was detected (Medians: 26.3 ω_C and 7.7 dN_C; One-sided Wilcoxon rank-sum test p
368 adj value < 2.7×10^{-5}), with no significant difference in the rate of synonymous convergent
369 substitutions (dS_C), which can function as negative control (Figure 2B). For scOGs under
370 relaxed and intensified selection there is less enrichment of for signals of convergence, but
371 in gregarious lineages, scOGs under intensified positive selection show a stronger
372 enrichment of ω_C compared with scOGs under relaxed selection (ω_C medians: 57.9 *vs* 22.8;
373 One-sided Wilcoxon rank-sum test p adj < 3.0×10^{-3}). This pattern suggests that the
374 intensified positive selection in gregarious lineages is partly driven by convergent
375 substitutions. Finally, selecting a more stringent threshold of ω_C > 5.0 and dN_C > 2.0 for an
376 *arity* ≥ 2, we were able to identify 192 scOGs under convergent positive selection (ω_C^+)
377 (Figure 2E, F). Most of this convergence occurred between pairs of gregarious lineages
378 (*arity* = 2) (Figure 2E). The most frequent lineages under ω_C^+ were *Heliconius aoede* and
379 *Dione juno* (32 scOGs with median ω_C of 9.9), which split from their MRCA more than 25
380 mya (Figure 1). Among these genes we find *Herc4*, a ubiquitin ligase4, *Notch,* and
381 *Neurotrophin 1* (*NT1*), which are all connected with Toll-related receptors involved in host
382 defense[60], and a (cytosine-5) tRNA methyltransferase (*Mt2*), which in *Drosophila* is
383 implicated in the innate immune response and lipid homeostasis[61]. *Heliconius aoede* also
384 has the highest number of scOGs under ω_C^+, with over 100 genes, followed by *D. juno* and
385 *H. ricini* (Figure 2e). The scOGs implicated in convergence among the highest number of

386 lineages are a *Rcc1*-homolog and *kazachoc* (*kcc*) (Figure 2E). *Rcc1* has roles in cell
387 proliferation, cell survival, apoptosis, epigenetic regulation, and neuronal specification[62,63].
388 *Kazachoc* encodes a potassium:chloride symporter expressed in the insect central nervous
389 system and is implicated in modulating neural firing rates[64].
390 Exploring overlap between genes highlighted by the independent analyses above, we
391 find significantly more overlap than expected by chance in four of the six pairwise datasets
392 (100,000 permutations; p values < 0.008). The most interesting overlap between genes
393 under relaxation, differentially selected, and under ω_C^+ include 14 scOGs. Among them are
394 *CycB3*, *Glg1*, *Vps51*, *Fancl*, which are implicated in the regulation of cell cycle, Golgi
395 apparatus, and response to DNA damage stimuli[65-68], and *norpA*, which in *Drosophila* plays a
396 prominent role in the transduction of light and environmental stimuli[69]. Overall, our analysis
397 of protein coding genes suggests a shift towards gregariousness that affects the selection
398 regimes shaping multiple biological processes, with a strong signal of convergence among a
399 small subset of genes.
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422

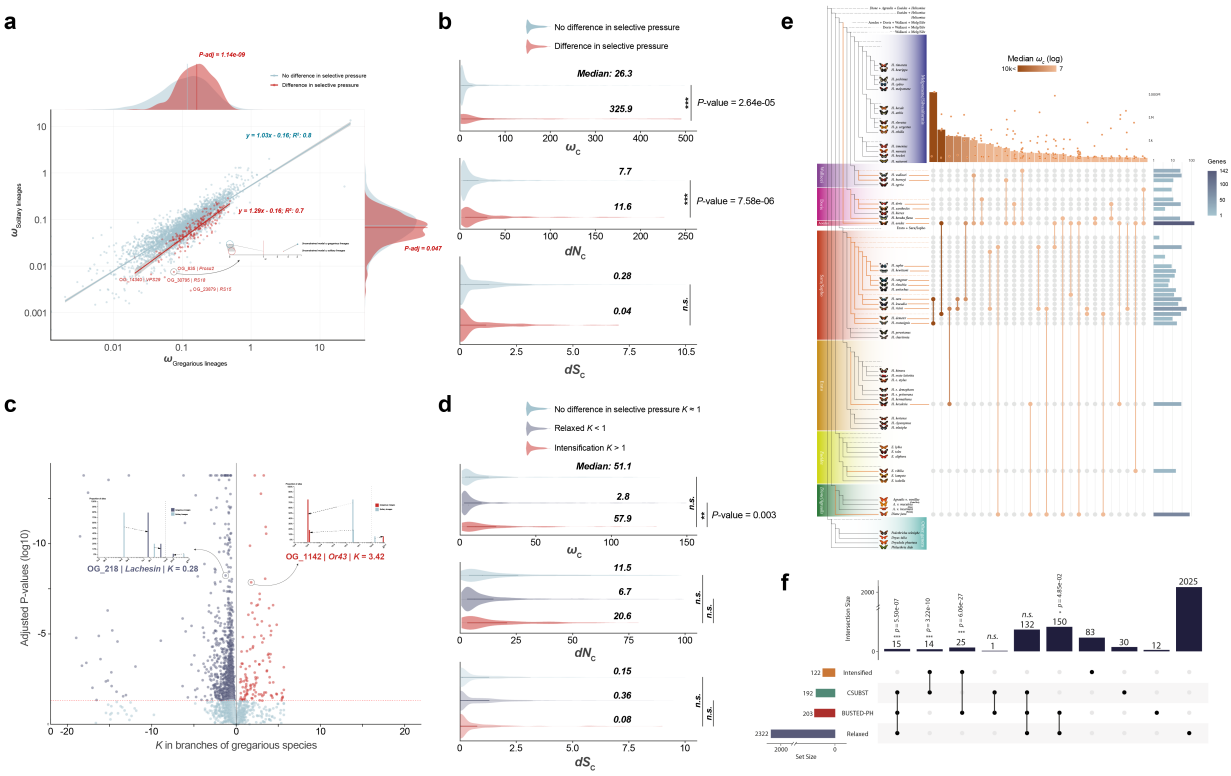

Figure 2. Convergence of molecular evolution in single-copy orthologous groups and their association with gregarious lineages. (A) Scatter plot of scOGs showing differential selecting regimes between gregarious and solitary lineages. Red dots correspond to genes that are differentially selected in gregarious lineages compared with solitary lineages, while light blue dots are not. Regression lines show significant slops. Along *x* and *y* axes the distributions of ω in gregarious and solitary lineages for the genes associated with gregarious and solitary lineages are plotted. The insert shows the three-rate classes of ω between gregarious vs solitary lineages in *Pros*α*2*, as an example. (B) Violin plots of convergent positive selection (ω_C), rate of non-synonymous convergent substitutions (dN_C), and rate of synonymous convergent substitutions (dS_C) between differentially (red) and not-differentially selected genes (light-blue). (C) Scatter plot of RELAX analysis. Light-blue dots represent genes under relaxation (k < 1; p adj. < 0.05); while red dots correspond to genes under intensification (k > 1; *p*-adj. < 0.05). *x*-axis is $\log_2$ transformed to centre k = 1 to 0. The red-dashed line corresponds to the *p* adj threshold of 0.05. The two inserts show the change in the three-rate classes of ω between gregarious vs solitary lineages in *Lachesin* and *Or43* loci, which are relaxed or intensified, respectively, as an example. *Lachesin* has been associated with change in crawling behaviour in *Drosophila*, while *Or43* is an odorant receptor known to be able to bind to the antimicrobial and larval food attractant compound citral. (D) Violin plots of convergent positive selection (ω_C), rate of non-synonymous convergent substitutions (dN_C), and rate of synonymous convergent substitutions (dS_C) between relaxed (light-purple), intensified (red), and no-difference genes. (E) CSUBST analysis across Heliconiini phylogeny. Next, the expanded tree diagram shows the lineages involved in convergent positive selection (ω^+_C). The histogram on top shows the median value of ω^+_C for specific pairs. On the right-most side the histogram shows the number of genes evolving convergently in the specific branch. (F) UpSet plot showing the intersection of the four gene lists. On top of the vertical histogram the corresponding p values of the probability of the overlap being due to a random distribution. In all plots asterisks the size of the p value: *n.s.* = not significant; * < 0.05; ** < 0.01; *** < 0.001.

454 ***ii) A subset of conserved non-exonic elements show accelerated rates of evolution in***
455 ***gregarious lineages.***
456 To determine the extent of convergent molecular evolution at non-coding loci, we next
457 compiled a total of 148k conserved non-exonic elements (CNEEs) across the 63-way
458 genome alignment[26], using *H. melpomene* as a reference. CNEEs are small genomic
459 regions previously shown to be enriched with *cis*-regulatory elements[26,70]. Within these 148k
460 CNEEs, we explored convergent shifts in molecular rate associated with transitions to
461 gregarious behaviour[71] (Figure 3). We identified 15.3% of CNEES that show variable rates of
462 evolution across the phylogeny (acceleration under a "full model" M2 has a higher
463 probability, expressed as Bayes Factor cutoff 1.0, than "null model" M0) and 9.8% that show
464 convergent acceleration (aCNEEs) in gregarious lineages (acceleration under a "target
465 model" M1 has a higher probability than M0 and M2), corresponding to ~10% (1.32Mbp,
466 14.6k CNEEs) of the reference genome. Of these regions roughly a third (29.2%) of all
467 aCNEEs are convergently accelerated in at least three different clades of the phylogeny
468 (Figure 3A, B).
469 We found 37 candidate genes associated with gregarious behaviour that harbour
470 more aCNEEs than expected by chance (Figure 3A, D; Table S3; p adj. < 0.05). These
471 include the putative *phenylalanyl-tRNA synthetase β-subunit homolog* (*β*-PheRS). *β*-PheRS
472 is an important enzyme that acts by charging tRNAs with their cognate amino acid and, in
473 *Drosophila*, manipulating *β*-PheRS expression affects growth speed, the timing of pupation,
474 and behaviours including feeding and roaming[72]. More broadly, aCNEEs are enriched in
475 putative regulatory domains of genes associated with more than 40 GO terms (Figure 3C;
476 Table S4). The most significantly enriched GO terms include those associated with viral
477 detection, defence response to oomycetes and fungi, regulation of antimicrobial peptide
478 production, and Toll signalling pathway - the innate immune response pathway in insects.
479 Finally, we explored the intersection between CNEEs and scOGs and found that in
480 putative regulatory domains of the scOGs, 15% of the CNEEs are accelerated (M2), and
481 9.4% are convergently accelerated (M1), of which 27% are accelerated in at least three
482 clades (Figure 3E). These are similar numbers to our genome-wide analyses, suggesting
483 that our scOG dataset is a robust representation of the entire coding gene repertoire.
484 Focusing on scOGs with putative associations with gregarious behaviour (*i.e.:* differentially
485 selected, relaxed, intensified and positively converging genes), we found a significant
486 enrichment of strict aCNEEs (*p* value < 0.05; 34.6% of CNEEs) in the putative regulatory
487 domains of scOGs with differential selective regimes in gregarious lineages. This translates
488 to a two-fold density of strict aCNEEs in the differentially selected scOGs (0.4 aCNEE/scOG)
489 compared with the background scOGs (0.2 aCNEE/scOG). We also observed that aCNEEs
490 in putative regulatory domains of differentially selected scOGs had higher acceleration rates

(median: 1.31 vs 1.28) and a lower conservation rate (median: 0.091 vs 0.093) (Figure 3F). Overall, our analyses of non-coding genomic regions show a strong signal of convergence associated with gregarious larvae. Importantly, the combination of analytical approaches applied to both coding and non-coding regions allow us to detect that selection for gregarious behaviour is simultaneously acting at the same loci: in protein coding regions, by changing amino-acid composition (purifying and/or positive selection), and through accelerated rates of sequence evolution in *cis*-regulatory elements (non-coding elements).

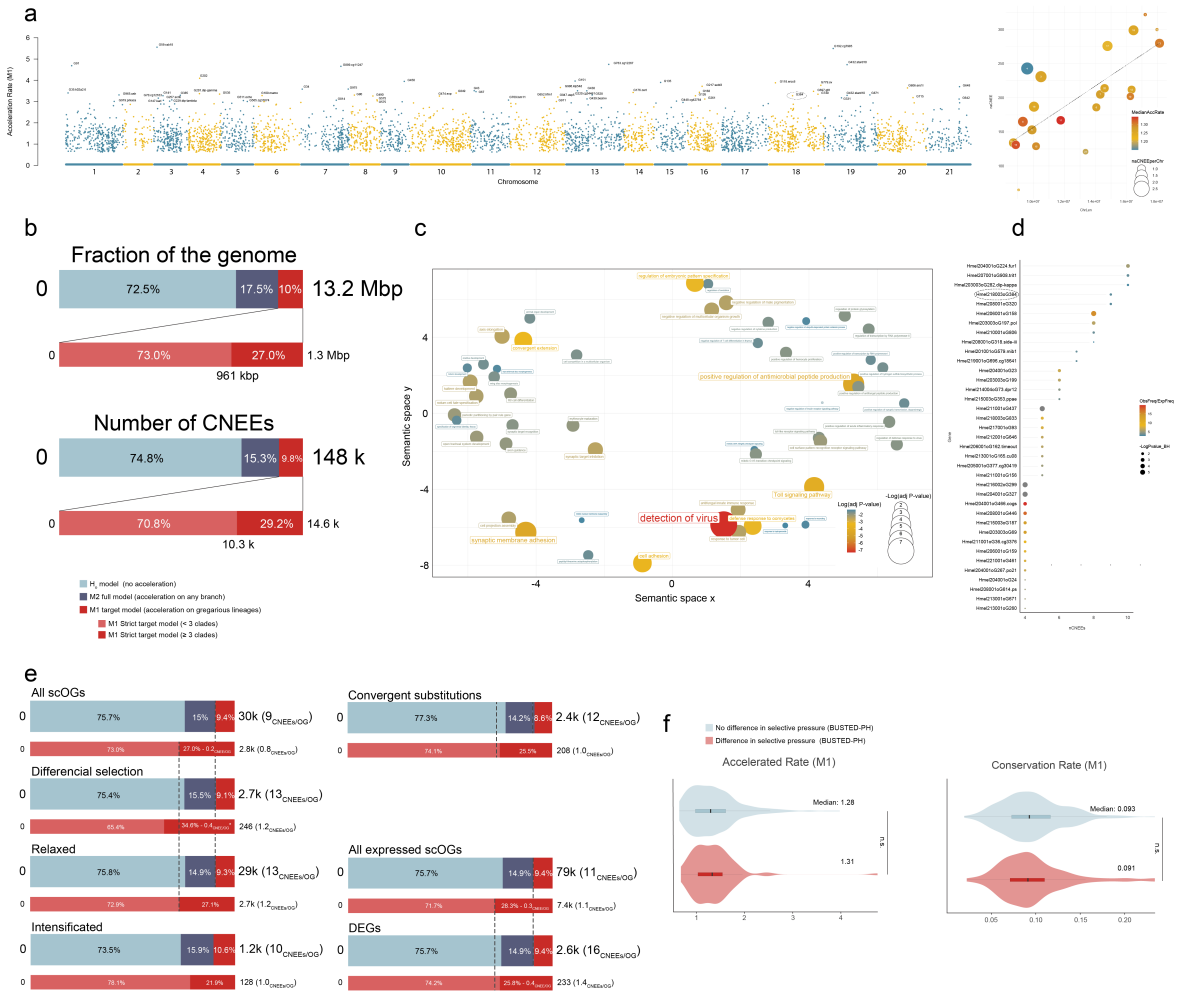

Figure 3. Convergent evolution of conserved non-coding elements in gregarious lineages throughout Heliconiini phylogeny. (A) Distribution of "strict" aCNEEs across the genome of *H. melpomene*, used as reference. *y*-axis is the acceleration rate under the M1 model. On the left side the positive correlation between the number of aCNEEs and chromosome lengths. Red-shift shows an enrichment of high rate. In particular chromosomes 11 and 14 show on average higher rates, while chromosome 4 show an excess of aCNEEs with respect its size, but with low rate on average; while chromosome 21, the sexual chromosome (Z), and 2 show a very low number of aCNEEs. (B) Staked histograms showing the proportion of the CNEEs as fraction of the genome (up) and in terms of model distribution (down). (C) Semantic plot of the enriched GO terms using genes interested by *strict* aCNEEs present in putative regulatory elements (see Methods). Large and red-shift circles correspond to GO terms with have smaller *p* values. (D) Bubble plot of some of the 73 genes enriched with aCNEEs in their regulatory elements (Table S3). Circle size shows the $\log_{10}$ of the *p* adj. values of the enrichment, while the red shift shows higher values of the ratio between the observed vs the expected frequency. Genes that are also differentially expressed in at least one comparison are listed in bold. (E) Staked histograms showing the proportion of the different classes CNEEs in different gene sets, including expressed and differentially expressed scOGs (DEGs). (F) Violin plots of accelerated rate and conservation rate for differentially selected genes (BUSTED-PH gene set). In all plots asterisks the size of the *p* value: *n.s.* = not significant; * < 0.05.

552 ***iii) Shifts in gene expression during the independent evolution of gregarious***
553 ***behaviour***
554 To further explore the impact of convergent shifts towards gregarious behaviour on neural
555 gene expression, we collected larval brain transcriptome data for three phylogenetically
556 independent pairs of solitary/gregarious species (Figure 4A), with five replicates per species.
557 Data normalisation, aligning gene lists between species pairs, and removing genes which
558 had zero reads across all species resulted in a final dataset of 4,874 scOGs shared across
559 all six species (obtained from Cicconardi *et al.*[26]). We next compared patterns of gene
560 expression between each pair of solitary/gregarious species. This revealed a relatively high
561 number of scOGs with significant differential expression within each pair (Figure 4B-D; Table
562 S5), with qualitatively greater divergence with increasing phylogenetic distance (Figure 4A),
563 suggesting a role of genetic drift in shaping gene expression divergence. Separate tests for
564 functional enrichment on differentially expressed genes (DEGs) between species pairs
565 revealed no significantly enriched GO terms after Benjamini-Hochberg p value correction for
566 false detection rate (all $p > 0.05$, Table S6).
567 We next examined the amount of DEG overlap between pairs. Permutation tests
568 revealed that the level of DEG (up- and downregulated in gregarious lineages) overlap
569 between *D. juno* and *H. doris* is significantly greater than expected by chance (47% and 56%
570 to total genes, respectively; $p < 0.001$, Figure 4E, Table S7), as is the degree of overlap at
571 the intersect between all three pairwise comparisons (23%, 28%, and 26% in *D. juno, H.*
572 *doris* and *H. sara*, respectively; $p < 0.001$, Figure 4E, Table S7). The higher strength of
573 overlap between *D. juno* and *H. doris* is notable, as the larvae of these species are more
574 strictly gregarious throughout larval development, relative to *H. sara*[37,38]. Considering
575 overlapping up- and downregulated genes separately, the level of upregulated gene overlap
576 between *D. juno* and *H. doris* is again significantly greater than expected by chance (29%
577 and 30%, respectively; $p = 0.002$, Table S7), and the same is true for the upregulated genes
578 at the intersect between all three pairs (10%, 10%, and 9% in *D. juno, H. doris* and *H. sara*,
579 respectively; $p = 0.008$). In contrast, overlap among downregulated genes does not deviate
580 from random expectation (Table S7). Considering scOGs with at least three aCNEEs in their
581 putative regulatory regions, we identified 14 of them that are differentially expressed
582 between one or more pairs of solitary and gregarious Heliconiini (Table S3). Six of these
583 harbour more aCNEEs than expected by chance (Table S3; p-adj. < 0.05). Of particular note
584 is *dpr12* (Figure 3D), which is differentially expressed in all three species pairs, and is
585 involved in dopamine signalling, a neuromodulator with known regulatory roles in several
586 insect behaviours linked to sociality, including feeding behaviour and aggression (reviewed
587 in Verlinden[73]); and two *semaphorins* (*sema-1a and sema-1b*, the former is shown in Figure
588 4F), which is involved in growth cone guidance through its role in axonal repulsion[74].

589 Several of the convergent DEGs with the greatest mean $\log_2$ fold-change identified in
590 our study are predicted to be involved in functions putatively linked to social behaviour, such
591 as hormonal pathways (*5htr*, *dop2r*), sensory system development and function (*CG31559,*
592 *ap-1σ, hdc, dan*), immune response (*rnf19b, CG3829, samdc, senju, ubc7*), feeding
593 behaviour and metabolism (*CG3552, mthl3, rmi1, hcrtr2, CG30344, indy*), and neural
594 development (*tao*) (Table S8). While genes associated with sensory functions may provide
595 clues as to the stimuli detected by social larvae, the convergent upregulation of genes
596 controlling hormonal pathways, such as serotonin and dopamine receptors, may also
597 suggest that modulation of how social cues are used plays an important role in the
598 evolutionary transition to social behaviour. As observed in other species, the activity and
599 regulation of these hormones can be crucial in the production of the social behavioural
600 phenotype[75-77], potentially because of their involvement in feeding[78-80] and aggression or
601 attraction/repulsion responses between conspecifics[75,77,81]. Elsewhere, these candidate
602 genes may be associated with secondary selection pressures which result from social living.
603 For example, feeding regulation is likely to be an important social phenotype given higher
604 competition for resources between group members[7,8] and the greater risk of starvation
605 during collective defoliation of a host plant[82,83]. Similarly, social organisms are more
606 susceptible to disease, given their close proximity to one another and sharing of food
607 sources[84], and group living is hypothesised to be associated with increased selection on
608 individual immunity[84,85].
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625

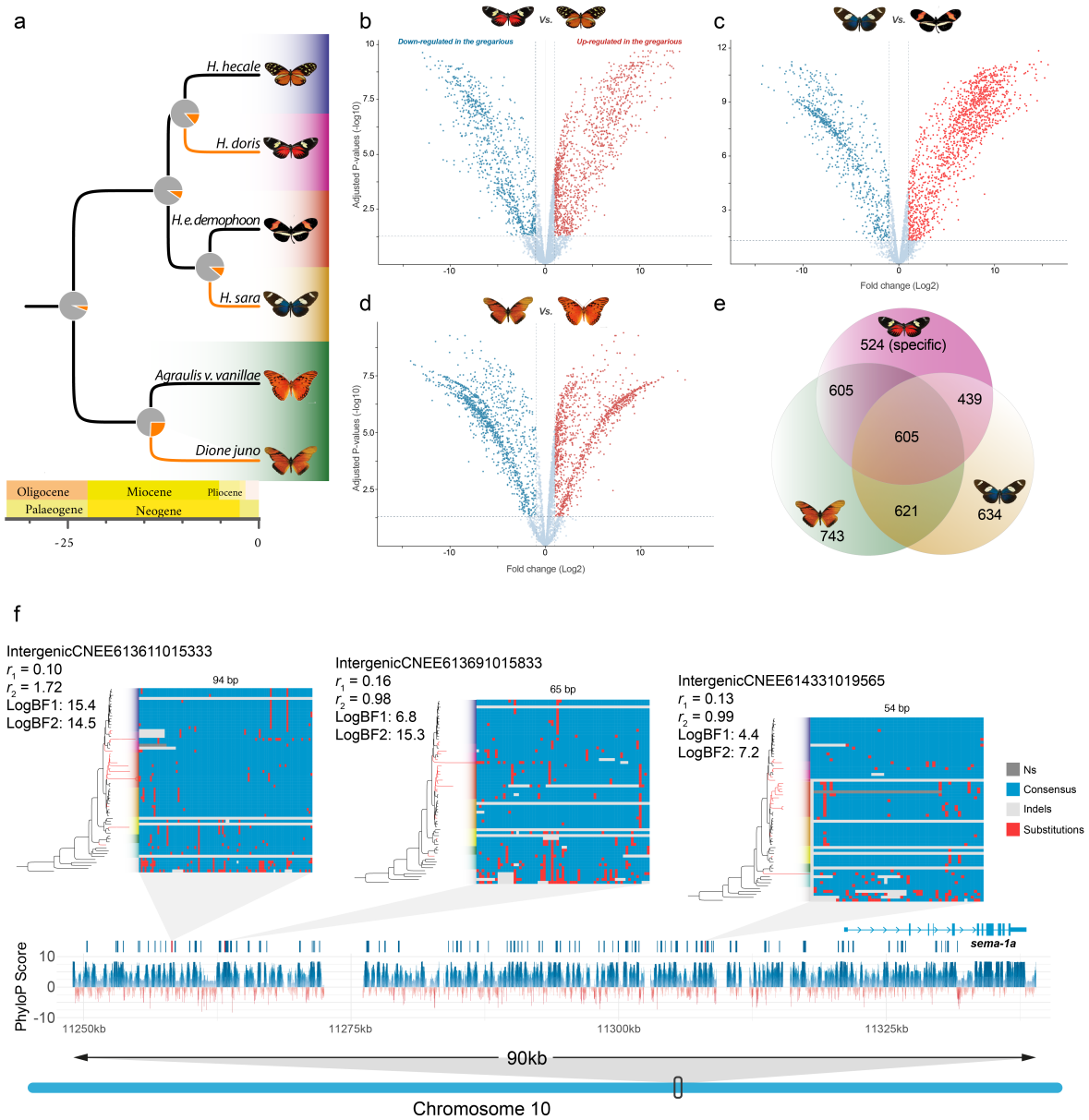

Figure 4. Convergent gene expression analysis across three pairs of species across the Heliconiini phylogeny. (A) Dated phylogeny of the three species pairs used as proxy to study the gene expression convergence in gregarious lineages. As in Figure 1, orange branches indicate lineages with gregarious larvae, while pie charts indicate the ancestral state reconstruction of the most recent common ancestor having gregarious larvae. (B-D) Volcano plot of the differentially expressed scOGs in the three pairs of species. In each pair the solitary larva species was used as "control", therefore red points correspond to upregulated genes in the gregarious species, while the light-blue points represent downregulated genes. Horizontal dashed lines indicate $p = 0.05$, vertical dashed lines indicate a ±1 $\log_2$FC threshold. (E) Venn diagram of the DEGs in common in the three studied pairs. Asterisks corresponds to the size of the p value: *n.s.* = not significant; * < 0.05; ** < 0.01; *** < 0.001. (F) Genomic locus around the gene *sema-1a*. The locus has a number of conserved loci as indicated by the phastCons and PhyloP tracks. The locus has three aCNEEs for each of them a phylogenetic tree is shown with branch lengths proportional to the corresponding acceleration rates for the gregarious lineages (red branches). Next to the phylogenetic tree the alignment of the CNEE is also showed. For each aCNEE the conservation (r_1), acceleration rates (r_2), and their respective log of the bayes factors (logBF) are also listed.

iv) Conclusions

Using multiple approaches across transcriptomic, protein coding and non-coding genomic data, our analyses provide a robust case that convergent transitions towards gregarious larval behaviour are associated with genome-wide convergent molecular evolution. We have identified protein-coding genes under differential selection in gregarious lineages, suggesting that changes at these loci may be important in the evolution of the gregarious phenotype. We also reveal a parallel signal of convergence at putative *cis*-regulatory elements, providing evidence of common shifts in selection affecting coding and non-coding elements of a wider locus. Finally, we have taken the first steps in exploring these gene expression profiles, revealing a convergent pattern of differential expression in association with convergence in the gregarious phenotype, with a number of genes associated with convergent aCNEEs being differentially expressed in independent gregarious lineages. Our comprehensive analyses identify the first candidate loci associated with gregarious larval behaviour in Lepidoptera, in particular highlighting the serotonergic and dopaminergic pathways as a potential mechanism for transitioning to larval gregariousness. Given the importance of these molecular mechanisms on other insect social behaviours, our data provide encouraging loci for future functional investigation. Crucially, our data also reveal the likely importance of genes involved in a wide range of physiological and behavioural traits, which likely evolved in concert with gregariousness *per se*. These include feeding behaviour, immune system function, and possibly aggression, as components of the repeated evolution of the gregarious phenotype across the Heliconiini. Little is known of the specific 'tolerance' genes that are likely to facilitate social behaviours[86], and our data provide a first step towards the identification of genes involved in anti-aggression, anti-cannibalism, and the prevention of egg consumption. More broadly, our analyses highlight the rich potential of phenotypically informed comparative analyses of densely sampled genomic datasets of closely related, but phenotypically divergent, phylogenetic groups.

ACKNOWLEDGMENTS

This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol. We are grateful to Christy Waterfall at the Bristol Genomics Facility for her assistance and detailed guidance. We thank Adriana Tapia, Moises Abanto, Oscar Paneso, Cruz Batista Saez, and STRI for support at the Gamboa insectaries, Panama. This work was funded by a Royal Commission for the Great Exhibition Research Fellowship, a Short-term STRI Fellowship, British Ecological Society Research Grant (3066), and a NERC IRF (NE/N014936/1) to SHM, and an SWBio DTP Studentship to CF (BB/M009122/1).

682 **AUTHOR CONTRIBUTIONS**

683 SHM conceived and designed the research, with input from FC and CFM. CFM collected
684 and analysed data the gene expression data, with assistance from AS, following larval
685 rearing and brain dissection performed by SHM with support from WOM. FC performed all
686 genomic analyses. CFM and FC wrote the first draft with assistance from SHM, with
687 subsequent edits performed equally between CFM, FC and SHM.

688

689 **DECLARATION OF INTEREST**

690 The authors declare no competing interests.

691

692 **Data and code availability**

693 The code used for this study and statistical analysis are available in the Git repository:
694 https://github.com/francicco/-ConvergentMolEvolutionGregLarvalBehaviour

695

696

697 **REFERENCES**

698 1. Rosenblum, E. B., Parent, C. E., and Brandt, E. E. (2014). The molecular basis of
699 phenotypic convergence. *Annu. Rev. Ecol. Evol. Syst.*, 45, 203-226.
700 2. Sackton, T. B., and Clark, N. (2019). Convergent evolution in the genomics era: new
701 insights and directions. *Philos. Trans. R. Soc. B*, 374, 20190102.
702 3. Christin, P. A., Weinreich, D. M., and Besnard, G. (2010). Causes and evolutionary
703 significance of genetic convergence. *TiG*, 26, 400-405.
704 4. Conte, G. L., Arnegard, M. E., Peichel, C. L., and Schluter, D. (2012). The probability of
705 genetic parallelism and convergence in natural populations. *Proc. R. Soc. B*, 279, 5039-
706 5047.
707 5. Fischer, E. K., Nowicki, J. P., and O'Connell, L. A. (2019). Evolution of affiliation:
708 patterns of convergence from genomes to behaviour. *Philos. Trans. R. Soc. B*, 374,
709 20180242.
710 6. Hochberg, M. E. (1991). Viruses as costs to gregarious feeding behaviour in the
711 Lepidoptera. *Oikos*, 61, 291-296.
712 7. Nilsson, M. and Forsman, A. (2003) Evolution of conspicuous colouration, body size and
713 gregariousness: a comparative analysis of lepidopteran larvae. *Evol. Ecol*. 17, 51–66.
714 (doi:10. 1023/A:1022417601010)
715 8. Reader, T. and Hochuli, D. F. (2003) Understanding gregariousness in a larval
716 lepidopteran: the roles of host plant, predation, and microclimate. *Ecol. Entomol*. 28,
717 729–737. (doi:10.1111/j.1365-2311.2003.00560.x)

9. Ezenwa, V. O., Ghai, R. R., McKay, A. F., and Williams, A. E. (2016). Group living and pathogen infection revisited. *Curr. Opin. Behav. Sci.*, 12, 66-72.
10. Li, H., Dai, C., Zhu, Y., & Hu, Y. (2021). Larvae Crowding Increases Development Rate, Improves Disease Resistance, and Induces Expression of Antioxidant Enzymes and Heat Shock Proteins in Mythimna separata (Lepidoptera: Noctuidae). J*. Econ. Entomol.*, 114, 1808-1816.
11. Wang, Y., Yang, P., Cui, F., and Kang, L. (2013). Altered immunity in crowded locust reduced fungal (*Metarhizium anisopliae*) pathogenesis. *PLoS pathogens*, 9(1), e1003102.
12. Ott, S. R., and Rogers, S. M. (2010). Gregarious desert locusts have substantially larger brains with altered proportions compared with the solitarious phase. *Proc. R. Soc. B*, 277, 3087-3096.
13. Costa, J. T. and Pierce, N. E. (1997). Social evolution in the Lepidoptera: ecological context and communication in larval societies. In: Choe JC & Crespi BJ (eds), *The evolution of social behavior in insects and arachnids*, 407-442. Cambridge University Press, Cambridge.
14. Hunter, A. F. (2000). Gregariousness and repellent defences in the survival of phytophagous insects. *Oikos*, 91(2), 213-224.
15. Costa, J. T. (2006). *The other insect societies*. Harvard University Press, Cambridge, USA.
16. Greeney, H. F., Dyer, L. A. and Smilanich, A. M. (2012). Feeding by lepidopteran larvae is dangerous: A review of caterpillars' chemical, physiological, morphological, and behavioral defenses against natural enemies. *Invertebrate Survival Journal*, 9(1), 7-34.
17. Sillén-Tullberg, B. (1988). Evolution of gregariousness in aposematic butterfly larvae: a phylogenetic analysis. *Evolution*, 42, 293-305.
18. Tullberg, B. S., and Hunter, A. F. (1996). Evolution of larval gregariousness in relation to repellent defences and warning coloration in tree-feeding Macrolepidoptera: a phylogenetic analysis based on independent contrasts. *Biol. J. Linn. Soc. Lond.*, 57, 253-276.
19. McLellan, C. F., Cuthill, I. C., and Montgomery, S. H. (2023). Warning coloration, body size, and the evolution of gregarious behavior in butterfly larvae. *Am. Nat.*, 202, 64-77.
20. Rittschof, C. C., and Robinson, G. E. (2014). Genomics: moving behavioural ecology beyond the phenotypic gambit. *Anim. behav.*, 92, 263-270.
21. Despland, E. (2021). Selection forces driving herding of herbivorous insect larvae. *Front. Ecol. Evol.*, 9, 760806.

22. Rentería, J., Despland, E. and Checa, M. F. (2022). Grouping as a strategy to mitigate top-down and bottom-up pressures for survival and growth in *Methona confusa* (Butler, 1873) (Nymphalidae, Ithomiini). *Trop. Lepid. Res.*, 32, 32-37.
23. McLellan, C. F., and Montgomery, S. H. (2023). Towards an integrative approach to understanding collective behaviour in caterpillars. *Philos. Trans. R. Soc. B*, 378, 20220072.
24. McLellan, C. F., and Montgomery, S. H. (2024). Evolution of larval gregariousness is associated with host plant specialisation, but not host morphology, in Heliconiini butterflies. *Ecol. Evol.*, 14, e11002.
25. Beltran, M., Jiggins C. D., Brower, A. V., Bermingham, E. and Mallet, J. (2007). Do pollen feeding, pupal-mating and larval gregariousness have a single origin in *Heliconius* butterflies? Inferences from multilocus DNA sequence data. *Biol. J. Linn. Soc. Lond.*, 92, 221-39.
26. Cicconardi, F., Milanetti, E., Castro, E.C.P. De, Mazo-vargas, A., Belleghem, S.M. Van, Ruggieri, A.A., Rastas, P., Hanly, J., Evans, E., Jiggins, C.D., et al. (2023). Evolutionary dynamics of genome size and content during the adaptive radiation of Heliconiini butterflies. *Nat. Commun*. 14, 5620. 10.1038/s41467-023-41412-5.
27. Pond, S. (2023). iGEM/UCSD evolutionary biology and bioinformatics group. hyphy-analyses. https://github.com/veg/hyphy-analyses.
28. Kosakovsky Pond, S.L., Poon, A.F.Y., Velazquez, R., Weaver, S., Hepler, N.L., Murrell, B., Shank, S.D., Magalis, B.R., Bouvier, D., Nekrutenko, A., et al. (2020). HyPhy 2.5 - A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol. Biol. Evol*. 37, 295–299. 10.1093/molbev/msz197.
29. Warton, D.I., Duursma, R.A., Falster, D.S., and Taskinen, S. (2012). smatr 3- an R package for estimation and inference about allometric lines. *Methods in Ecol. Evol.*. https://doi.org/10.1111/j.2041-210X.2011.00153.x.
30. Wertheim, J.O., Murrell, B., Smith, M.D., Kosakovsky Pond, S.L., and Scheffler, K. (2014). RELAX: Detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol*. 32, 1–13. 10.1093/molbev/msu400.
31. Fukushima, K., and Pollock, D.D. (2023). Detecting macroevolutionary genotype–phenotype associations using error-corrected rates of protein convergence. *Nat. Ecol. Evol*. 7, 155–170. 10.1038/s41559-022-01932-7.
32. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Spieth, J., Hillier, L.W., Richards, S., Weinstock, G.M., et al. (2005). Evolutionarily conserved elements in vertebrate , insect , worm , and yeast genomes. *Genome research* 15, 1034–1050. https://doi.org/10.1101/gr.3715005.

33. Hubisz, M.J., Pollard, K.S., and Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. *BBIMFX,* 12, 41–51. https://doi.org/10.1093/bib/bbq072.
34. Yan, H., Hu, Z., Thomas, G.W.C., Edwards, S. V., Sackton, T.B., and Liu, J.S. (2023). PhyloAcc-GT: A Bayesian Method for Inferring Patterns of Substitution Rate Shifts on Targeted Lineages Accounting for Gene Tree Discordance. *Mol. Biol. Evol.,* 40, 1–21. https://doi.org/10.1093/molbev/msad195.
35. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotech.,* 28, 495–501. https://doi.org/10.1038/nbt.1630.
36. Tomislav, S. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. 6. https://doi.org/10.1371/journal.pone.0021800.
37. Muyshondt, A., and Young, A. M. (1973). The biology of the butterfly *Dione juno huascama* (Nymphalidae: Heliconiinae) in El Salvador. *J. N. Y. Entomol.*, 137-151.
38. Jiggins, C. D. (2017). The Ecol. Evol. of *Heliconius* butterflies. Oxford University Press.
39. Sobral-Souza, T., Bastos Francini, R., Guimarães, M., and Benson, W. W. (2015). Short-term dynamics reveals seasonality in a subtropical *Heliconius* butterfly. *J. Insects*, 2015.
40. Smallegange, R. C., Van Loon, J. J. A., Blatt, S. E., Harvey, J. A., Agerbirk, N. and Dicke, M. (2007). Flower vs. leaf feeding by *Pieris brassicae*: glucosinolate-rich flower tissues are preferred and sustain higher growth rate. *J. Chem. Ecol.*, 33, 1831-1844.
41. Winnebeck, E. C., Millar, C. D. and Warman, G. R. (2010). Why does insect RNA look degraded?. *J. Insect Sci.*, 10, 159.
42. Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.*, 7, 1-14.
43. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data (version 0.11.4). https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
44. Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114-2120.
45. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P. Chaisson, M. and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15-21.
46. Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31, 166-169.
47. R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

825 48. Phipson, B, Lee, S, Majewski, IJ, Alexander, WS, and Smyth, GK (2016). Robust
826 hyperparameter estimation protects against hypervariable genes and improves power to
827 detect differential expression. *Ann. Appl. Stat.,* 10:946–963.
828 49. Hansen, KD, Irizarry, RA and Wu, Z (2012). Removing technical variability in RNA-seq
829 data using conditional quantile normalization. *Biostatistics*, 13, 204-216.
830 50. Zhou, Y., Zhu, J., Tong, T., Wang, J., Lin, B. and Zhang, J. (2019). A statistical
831 normalization method and differential expression analysis for RNA-seq data between
832 different species. *BMC Bioinformatics*, 20, 1-10.
833 51. Montgomery, S. H. and Mank, J. E. (2016). Inferring regulatory change from gene
834 expression: the confounding effects of tissue scaling. *Mol. Ecol.*, 25, 5114-5128.
835 52. Ashburner, M. et al. and Sherlock, G. (2000). Gene ontology: tool for the unification of
836 biology. *Nature Genetics*, 25, 25-29.
837 53. Alexa, A. and Rahnenführer, J. (2009). Gene set enrichment analysis with topGO.
838 *Bioconductor Improv*, 27, 1-26.Rana, V.S., Popli, S., Saurav, G.K., Yadav, K., Kumar, A.,
839 Sunil, S., Kumar, N., Singh, O.P., Natarajan, K., and Rajagopal, R. (2020). *Aedes*
840 *aegypti* lachesin protein binds to the domain III of envelop protein of Dengue virus-2 and
841 inhibits viral replication. *Cell. Microbiol*. 22. 10.1111/cmi.13200.
842 54. Rana, V.S., Popli, S., Saurav, G.K., Yadav, K., Kumar, A., Sunil, S., Kumar, N., Singh,
843 O.P., Natarajan, K., and Rajagopal, R. (2020). *Aedes aegypti* lachesin protein binds to
844 the domain III of envelop protein of Dengue virus-2 and inhibits viral replication. *Cell.*
845 *Microbiol*. 22. 10.1111/cmi.13200.
846 55. Chaudhary, P., Proulx, J., and Park, I.W. (2023). Ubiquitin-protein ligase E3A (UBE3A)
847 mediation of viral infection and human diseases. *Virus Res*. 335, 199191.
848 10.1016/j.virusres.2023.199191.
849 56. Wang, X. yang, Zhao, C. xiao, Wang, X., Zhao, Z. qin, Su, Z. hao, Xu, P. zhen, Li, M.
850 wang, and Wu, Y. chun (2021). The validation of the role of several genes related to
851 *Bombyx mori* nucleopolyhedrovirus infection in vivo. *Arch. Insect Biochem. Physiol*. 106,
852 1–14. 10.1002/arch.21762.
853 57. Cicconardi, F., Morris, B., Martelossi, J., Ray, D.A., and Montgomery, S. (2023). Novel
854 sex-specific genes and diverse interspecific expression in the antennal transcriptomes
855 of ithomiine butterflies. *Authorea Prepr*.
856 58. Carraher, C., Authier, A., Steinwender, B., and Newcomb, R.D. (2012). Sequence
857 comparisons of odorant receptors among tortricid moths reveal different rates of
858 molecular evolution among family members. *PLoS One* 7, 1–9.
859 10.1371/journal.pone.0038391.

860 59. Hamamura, Y., and Naito, K.I. (1961). Food selection by silkworm larvæ, *Bombyx mori*:
861 Citral, linalyl acetate, linalol, and terpinyl acetate as attractants of larvæ. *Nature*.
862 10.1038/190879a0.
863 60. Zhang, Q., Wang, C., Liu, Z., Liu, X., Han, C., Cao, X., and Li, N. (2012). Notch signal
864 suppresses toll-like receptor-triggered inflammatory responses in macrophages by
865 inhibiting extracellular signal-regulated kinase 1/2-mediated nuclear factor κB activation.
866 *J. Biol. Chem*. 287, 6208–6217. 10.1074/jbc.M111.310375.
867 61. Abhyankar, V., Kaduskar, B., Kamat, S.S., Deobagkar, D., and Ratnaparkhi, G.S.
868 (2018). *Drosophila* DNA/RNA methyltransferase contributes to robust host defense in
869 aging animals by regulating sphingolipid metabolism. *J. Exp. Biol*. 221, 1–10.
870 10.1242/jeb.187989.
871 62. Dasso, M., Nishitani, H., Kornbluth, S., Nishimoto, T., and Newport, J. W. (1992). RCC1,
872 a regulator of mitosis, is essential for DNA replication. *MCB*, 12, 3337-3345.
873 63. Hadjebi, O., Casas-Terradellas, E., Garcia-Gonzalo, F. R., and Rosa, J. L. (2008). The
874 RCC1 superfamily: from genes, to function, to disease. *Biochim. Biophys. Acta, Mol.*
875 *Cell Res.*, 1783, 1467-1479.
876 64. Chen, R., and Swale, D.R. (2023). Functional interactions between potassium-chloride
877 cotransporter (KCC) and inward rectifier potassium (Kir) channels in the insect central
878 nervous system. *Pestic. Biochem. Physiol*. 192, 105389. 10.1016/j.pestbp.2023.105389.
879 65. Reggiori, F., Wang, C. W., Stromhaug, P. E., Shintani, T., and Klionsky, D. J. (2003).
880 Vps51 is part of the yeast Vps fifty-three tethering complex essential for retrograde
881 traffic from the early endosome and Cvt vesicle completion. *JBC*, 278, 5009-5020.
882 66. Hodson, C., Cole, A. R., Lewis, L. P., Miles, J. A., Purkiss, A., and Walden, H. (2011).
883 Structural analysis of human FANCL, the E3 ligase in the Fanconi anemia pathway.
884 *JBC*, 286, 32628-32637.
885 67. Planche, A., Bacac, M., and Stamenkovic, I. (2012). The Golgi protein GLG1
886 participates in tumor progression. *Cancer Research*, 72, 3245-3245.
887 68. Garrido, D., Bourouh, M., Bonneil, É., Thibault, P., Swan, A., and Archambault, V.
888 (2020). Cyclin B3 activates the anaphase-promoting complex/cyclosome in meiosis and
889 mitosis. *PLoS genetics*, 16, e1009184.
890 69. Bloomquist, B. T., Shortridge, R. D., Schneuwly, S., Perdew, M., Montell, C., Steller, H.,
891 Rubin, G. and Pak, W. L. (1988). Isolation of a putative phospholipase C gene of
892 *Drosophila*, norpA, and its role in phototransduction. *Cell*, 54, 723-733.
893 70. Xue, J.R., Mouri, K., Garcia, M.F., Akers, J.F., Noble, M., Noonan, J.P., Capellini, T.D.,
894 Brennand, K.J., Tewhey, R., Reilly, S.K., et al. (2023). The functional and evolutionary
895 impacts of human-specific deletions in conserved elements. *Science,* 380.
896 10.1126/SCIENCE.ABN2253.

897 71. Sackton, T.B., Grayson, P., Cloutier, A., Hu, Z., Liu, J.S., Wheeler, N.E., Gardner, P.P.,
898 Clarke, J.A., Baker, A.J., Clamp, M., et al. (2019). Convergent regulatory evolution and
899 loss of flight in paleognathous birds. *Science,* 364, 74–78. 10.1126/science.aat7244.
900 72. Brunßen, D., and Suter, B. (2024). Effects of unstable β-PheRS on food avoidance,
901 growth, and development are suppressed by the appetite hormone CCHa2. *Fly (Austin)*.
902 18. 10.1080/19336934.2024.2308737.
903 73. Verlinden, H. (2018). Dopamine signalling in locusts and other insects. *Insect Biochem.*
904 *Mol. Biol.*, 97, 40-52.
905 74. Nguyen, C. T. and Jeong, S. (2022). Regulation of Off-track bidirectional signaling by
906 Semaphorin-1a and Wnt signaling in the *Drosophila* motor axon guidance. *Insect*
907 *Biochem. Mol. Biol.*, *150*, 103857.
908 75. Anstey, M. L., Rogers, S. M., Ott, S. R., Burrows, M., and Simpson, S. J. (2009).
909 Serotonin mediates behavioral gregarization underlying swarm formation in desert
910 locusts. *Science*, 323, 627-630.
911 76. Vleugels, R., Verlinden, H., and Vanden Broeck, J. (2015). Serotonin, serotonin
912 receptors and their actions in insects. *Neurotransmitter*, 2.
913 77. Sasaki, K., Okada, Y., Shimoji, H., Aonuma, H., Miura, T., and Tsuji, K. (2021). Social
914 evolution with decoupling of multiple roles of biogenic amines into different phenotypes
915 in Hymenoptera. *Front. Ecol. Evol.*, *9*, 659160.
916 78. Dacks, A. M., Nickel, T., and Mitchell, B. K. (2003). An examination of serotonin and
917 feeding in the flesh fly *Neobellieria bullata* (Sarcophagidae: Diptera). *J. Insect. Behav.*,
918 16, 1-21.
919 79. Neckameyer, W. S. (2010). A trophic role for serotonin in the development of a simple
920 feeding circuit. *Dev. Neurosci.*, 32, 217-237.
921 80. French, A. S., Simcock, K. L., Rolke, D., Gartside, S. E., Blenau, W., and Wright, G. A.
922 (2014). The role of serotonin in feeding and gut contractions in the honeybee. *J. Insect.*
923 *Physiol.*, 61, 8-15.
924 81. Dierick, H. A., and Greenspan, R. J. (2007). Serotonin and neuropeptide F have
925 opposite modulatory effects on fly aggression. *Nature genetics*, 39, 678-682.
926 82. Denno, R., and Benrey, B. (1997). Aggregation facilitates larval growth in the neotropical
927 nymphalid butterfly *Chlosyne janais*. *Ecol. Entomol.*, 22, 133-141.
928 83. Daly, D., Higginson, A. D., Chen, D., Ruxton, G. D. and Speed, M. P. (2012). Density-
929 dependent investment in costly anti-predator defences: an explanation for the weak
930 survival benefit of group living. *Ecol. Lett.*, 15, 576-583.
931 84. López-Uribe, M. M., Sconiers, W. B., Frank, S. D., Dunn, R. R., and Tarpy, D. R. (2016).
932 Reduced cellular immune response in social insect lineages. *Biol. Lett.*, 12, 20150984.

933 85. Evans, J. D., Aronstein, K., Chen, Y. P., Hetru, C., Imler, J. L., Jiang, H., Kanost, M.,
934 Thompson, G. J., Zou, Z. and Hultmark, D. (2006). Immune pathways and defence
935 mechanisms in honey bees *Apis mellifera*. *Insect Mol. Biol.*, 15, 645-656.
936 86. Mayhew, P. J., and van Alphen, J. J. (1999). Gregarious development in alysiine
937 parasitoids evolved through a reduction in larval aggression. *Anim. behav.*, 58, 131-141.