Manuscript type: Research Article

Title: The developmental gene *disco* regulates diel-niche evolution in adult moths

Authors: Yash Sondhi[1,2] (ORCID:0000-0002-7704-3944), Rebeccah L. Messcher[2], Anthony J. Bellantuano[1], Caroline G. Storer[2], Scott D. Cinel[2], R. Keating Godfrey[2], Deborah Glass[4], Ryan A. St Laurent[2,6], Chris A. Hamilton[5], Chandra Earl[7], Colin J. Brislawn[2], Ian J. Kitching[4], Seth M. Bybee[3], Jamie C. Theobald[1], Akito Y. Kawahara[2]

Author Affiliation:

[1]Department of Biology, Florida International University, Miami FL 33174 USA

[2]McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, University of Florida, Gainesville, FL 32611 USA

[3]Department of Biology, Monte L. Bean Museum, Brigham Young University, 4102 Life Science Building, Provo, UT 84602, USA

[4]Natural History Museum, Cromwell Road, London SW7 5BD, UK.

[5]Department of Entomology, Plant Pathology & Nematology, University of Idaho, Moscow, ID 83844, USA

[6]Department of Entomology, Smithsonian Institution, National Museum of Natural History, Washington, D.C., U.S.A

[7]Bernice P. Bishop Museum, Honolulu, HI 96817, USA

Corresponding author: Yash Sondhi, yashsondhi@gmail.com

Keywords: circadian; development, diel-niche; gene expression; insects; Lepidoptera; light cycle; RNA-seq transcriptome, transcription factor

Significance: Insect diel-activity patterns are diverse, yet the underlying evolutionary processes are poorly understood. Light environment powerfully entrains circadian rhythms and drives diel-niche and sensory evolution. To investigate its impact, we compared gene expression in closely related day- and night-active wild silk moths, with otherwise similar ecologies. Expression patterns that varied with diel activity included genes linked to eye development, neural plasticity and cellular maintenance. Notably, *disco*, which encodes a zinc-finger transcription factor involved in pupal

Drosophila optic lobe and antennal development, shows robust adult circadian mRNA cycling in moth heads, is highly conserved in moths, and has additional zinc-finger domains with specific nocturnal and diurnal mutations. We hypothesize that *disco* may contribute to diversification of adult diel-activity patterns in moths.

Abstract: Animals shift activity periods to reduce predation, minimize competition, or exploit new resources, and this can drive sensory system evolution. But adaptive mechanisms underlying niche-shifts are poorly understood, and model organisms are often too distantly related to reveal the genetic drivers. To address this, we examined expression patterns between two closely related silk moths that have undergone temporal niche divergence. We found 200-700 differentially expressed genes, including day upregulation in eye development and visual processing genes, and night upregulation of antennal and olfactory brain development genes. Further, clusters of circadian, sensory, and brain development genes co-expressed with diel-activity. In both species, eight genes showed expression significantly correlated to diel activity, and are involved in vision, olfaction, brain development, neural plasticity, energy utilization, and cellular maintenance. We repeatedly recovered *disco*, a zinc-finger transcription factor involved in antennal development, circadian activity, and optic lobe brain development in flies. While *disco* mutants have circadian arrhythmia, most studies attribute this to improper clock neuron development, not adult circadian maintenance. Comparing predicted 3D protein structure across moth and fly genetic models revealed *disco* likely retained developmental function with a conserved zinc finger domain, but gained functional zinc finger domains absent in *D. melanogaster.* These regions have several mutations between nocturnal and diurnal species that co-occur with higher levels of predicted phosphorylation sites. With robust circadian expression, functional nocturnal and diurnal mutations, and structural and sequence conservation, we hypothesize that *disco* may be a master regulator contributing to diel-activity diversification in adult moths.

Introduction

Circadian rhythms regulate many biological processes, including day-night activity patterns. Research to date has explored genes, circuits, and environmental cues mostly in the context of control within a single organism (1, 2). Diel-niche stems from the concept of temporal partitioning of activity periods, driven by the variation in resources, temperature, light level, or predation, across the

day (3–6). Periods are often binned into three distinct groups: diurnal (day active), nocturnal (night active), or crepuscular (active at dusk, dawn, or both), and although this is useful for comparative analysis, it is often an overgeneralization (7–11), and the evolution of diverse diel-niches across organisms (7, 11–15) is poorly understood (16–19).

Several studies have documented sensory adaptation accompanying evolutionary diel-niche transitions in mammals, birds, and insects (8, 18, 20, 21). Specific examples include colour vision gene expansions in diurnal moths (22, 23), improved olfactory senses in dark-bred flies (24), and the evolution of hearing organs in nocturnal butterflies (25). Comparisons of circadian genes across model organisms catalogue variation at long evolutionary timescales and highlight conserved elements (26, 27), but diel switches can occur over short timescales (28), and distantly related species provide less insight into mechanisms that enable faster switches. Moths and butterflies (Lepidoptera) are an ideal group to address this, as their evolutionary history is well known and there are many diel-niche switches, often between closely related species (7, 29). They are also one of the few insect groups, outside of *Drosophila*, in which both sensory and circadian genes have been characterized (23, 30–35).

Wild silk moths (Saturniidae) are important models for understanding chronobiology, with seminal experiments confirming the brain functions as the primary circadian control (36, 37). While most are nocturnal, many fly during the day, and temporal niche switches may function as a mechanism for reproductive isolation in sympatric species (38, 39). Saturniid phylogeny is relatively well known, and there are several annotated genomes, chromosomal maps. Furthermore, they belong to the superfamily Bombycoidea, which includes the domesticated silk moth, *Bombyx mori*, a key biological model organism serving as a reference taxon for comparative genomic studies (32, 40–43). Saturniids such as *Antheraea* are of major importance to the silk industry, and they have a well-resolved taxonomy and well-documented life histories (38, 44–46). Despite this, the genetic elements that control day-night activity and diel-niche evolution, in this group and insects in general, remain largely unknown (28, 37).

We used RNA-Seq to characterize expression patterns during peaks and troughs (midday and midnight) across two closely related wild silk moths: the diurnal *Anisota pellucida*, and nocturnal *Dryocampa rubicunda*. They are sister genera, feeding on large deciduous trees in the temperate forests in Eastern North America, with a recent divergence in temporal niche ~3.8 Mya (47). We

92 sequenced head tissue at different time periods to identify expression in sensory, circadian, and
93 neural genes that correlate with diel activity. Expression in eight genes clearly and significantly
94 correlates with diel activity, with functions in vision, olfaction, brain development, neural plasticity,
95 energy utilization, and cellular maintenance. Of these, a single gene emerges consistently across
96 analyses: *disco*, which encodes a zinc-finger transcription factor. In moths, it has likely retained
97 functions also found in flies (eye and brain development during the pupal stage) through a conserved
98 zinc-finger domain. But it has also gained an extra zinc-finger domain, surrounded by
99 phosphorylated sites and with mutations in both nocturnal and diurnal species, which may contribute
100 to adult moth circadian regulation.
101

102 ## Results

103 We compared gene expression across two wild silk moth species, *Anisota pellucida* and *Dryocampa*
104 *rubicunda*, whose males are diurnal and nocturnal, respectively (Fig. 1, Table S1). We generated
105 head (eyes, brain) transcriptomes from moths collected and flash frozen at midday and midnight,
106 referred to as 'day' and 'night' hereafter. Using multiple programs to assemble high-quality *de novo*
107 assemblies (Table S2), we characterized the level of gene (mRNA) expression with quasi-read
108 mapping, which can be used to correlate protein expression (48).
109

110 ***Day-night gene expression patterns switch between nocturnal and diurnal species***

111 We found 350 & 393 significantly differentially expressed genes (DEGs) when comparing day and
112 night treatments for each species (Table S3, Fig. 2A, Supplementary Data 1). *Anisota* had more day-
113 upregulated genes (56%) and *Dryocampa* was slightly more night-upregulated (53%). In order to
114 compare DEG sets between species, we mapped our DEGs to *Bombyx mori* orthologs using the
115 Orthofinder software (49) (Supplementary Data 1). Approximately 60% of DEGs from each species
116 had identifiable orthologs in *B. mori* (Table S3), and only a small number of DEGs (6-8 genes)
117 overlapped between both species (Fig. 2B). We also replicated this analysis using other software to
118 assure that our results were robust, despite different normalization methods (50). With DESeq2, we
119 found 498 & 697 DEGs (Fig. 2C, Table S3, Supplementary Data 2), with similar *Bombyx* annotation
120 rates (61%), although the proportion of day upregulated genes increased considerably in the *Anisota*
121 (79%) compared to being more even split in *Dryocampa* (50%) (Table S3). The total number of

122 overlapping genes increased (19-26, Fig. 2D) when using DESeq2. A comparison of the two methods
123 revealed that 174 and 216 genes were shared between *Anisota* and *Dryocampa*, respectively.
124
125 ***Divergently expressed genes are linked to brain optic lobe, antennal and neural development***
126 To identify important regulators involved in diel-niche evolution, we applied two filtering criteria to
127 our gene expression data. First, we selected genes that exhibited highly significant differential
128 expression in both species. Second, we focused on genes that displayed upregulation patterns
129 consistent with the natural diel-activity of each species. Our rationale was that this subset of genes
130 was more likely to contain key regulators. To compare DEG overlap between the two species, we
131 grouped the transcripts to their matching orthologs from *Bombyx mori*; if two transcripts from
132 different species mapped to the same ortholog, we treated them as being the same. This allowed us to
133 examine overlapping genes between the species to see if any genes switched fold-change sign from
134 positive to negative or vice versa. (Fig. 2, Supplementary Data 3). We found 51 overlapping DEG
135 transcripts that mapped to 28 unique *Bombyx* genes. Nine genes showed flipped patterns of
136 expression between the two species, and eight of them coincided with known diel activity patterns
137 (Table S7). Examining gene ontology (GO) annotations and comparing orthologs from flybase
138 (https://flybase.org/), we found genes linked to optic lobe and antennal development (*disco*),
139 locomotion and energy use (*SLC2A6, SLC17A5*), brain and neural development (*TUBG1*) and other
140 essential biological processes like transcription, ribosomal translation, protein processing,
141 mitochondrial maintenance (*RpS4, PARL, Mrps5*) and wound response (*PRP2*) (Table S7, Fig. 2B,
142 D). Of these, only *disco* was recovered with both methods.
143
144 ***Gene network analysis identifies diel activity and species-specific co-expressed clusters***
145 Identifying highly expressed genes helps understand which genes are activated during particular
146 biological processes. However, determining only those that are highly expressed can often overlook
147 genes with important biological functions (51). We examined co-expressed genes that may be
148 correlated with diel-niche or RNA collection time. We used WGCNA, a weighted correlation
149 network analysis tool to cluster genes together based on their normalized counts (52). After
150 examining co-expression patterns for each species separately, we found two modules in each species
151 that clustered with day-night treatment (cluster-*grey60* and cluster-*tan*) (Fig. S6, Supplementary data

152 4). Since we were interested in species-specific differences, we reran analyses and combined reads
153 from *Anisota* and *Dryocampa,* using only normalized counts for genes that had valid *Bombyx*
154 annotations for both species. This approach narrowed our focus to 2000 genes. Among these, we
155 discovered two clusters (cluster-blue and cluster-turquoise) consisting of 50 genes each that exhibited
156 different expression patterns across species (Fig. 4).
157
158 ***GO enrichment of photoperiodism, circadian control, muscle and neural growth genes***
159 We used a gene enrichment analysis to determine if gene ontology (GO) terms were significantly
160 overrepresented in the DEG and WGCNA sets compared to the appropriate background of GO terms.
161 Using TopGo, which allows custom gene sets, we found an overrepresentation of genes involved in
162 several biological functions (Supplementary Data 5). Those that seemed significant for diel-niche and
163 vision were 'response to stimulus' and 'smooth muscle control', as well as folic acid serine, glycine
164 and retinoic acid metabolism. We also used ShinyGo to examine WGCNA clusters (Supplementary
165 Data 5). We matched orthologous genes in the less duplicated, filtered transcriptomes to obtain
166 identifications of the most related *Bombyx* genes (Table S4). We examined the enrichment of both
167 tan and grey60 modules, listing the non-redundant terms using ReviGo (Fig. S7, Supplementary Data
168 5). Gene clusters that co-expressed in the same direction together in the day and night treatments of
169 both species included photoperiodism, circadian control, negative phototaxis and nervous system
170 development. We next checked for the enrichment of the modules that showed species-specific
171 patterns (blue and turquoise, Fig. S8, Supplementary Data 5). These included genes involved in
172 muscle proliferation and nerve growth, neural signaling, glycolysis, oxidative stress response, and
173 basic cellular functioning such as protein processing and transcriptional regulation.
174
175 ***Day upregulation of vision genes in the diurnal moth***
176 Since both EdgeR and DESeq2 analyses use different normalization methods and statistical model
177 assumptions (50, 53), we repeated enrichment analyses by combining datasets and examining genes
178 that appeared in both analyses. For the *Anisota,* we tested over-enrichment of a smaller subset (FC ≤
179 -5) of diurnally highly upregulated genes (Fig. 3, Table S5). We found gene enrichment for visual
180 perception, excretion regulation, negative gravitaxis, synaptic plasticity, along with genes associated
181 with other biological processes, such as RNA interference, endopeptidase activity and endocytosis. A

182 reduction in stringency (FC ≤- 2) did not alter results considerably (Fig. S4). Night-upregulated
183 genes (FC ≥ 2) included ocellar pigment genes, eye-photoreceptor cell development, snRNA
184 processing, post embryonic development and neurotransmitter secretion, among a host of other
185 processes that may be required for growth, development and metabolism (wnt signaling,
186 tricarboxylic acid cycle, and cellular response to insulin, glucose transport) (Fig. S4).
187
188 ***Night upregulation of antennal and olfactory brain regions mushroom development genes***
189 We repeated the same analyses for *Dryocampa* and tested highly nocturnally upregulated genes (FC ≥ 5).
190 Our results show upregulation in genes known to be associated with mushroom body development,
191 locomotor rhythm, synaptic growth, energy utilization (Sialin transport) and mitochondrial translation
192 (Fig. 3B, Table S6). A reduction in stringency (|FC| ≥ 2) showed entrainment of the clock cycle, and
193 antennal development genes. Genes associated with innate immune response, DNA repair, cell division,
194 histone acetylation, circadian rhythm, retinoid cycle were upregulated during the day, possibly indicating
195 a period of cellular repair during a time when these moths are inactive (Fig. S5).
196
197 ***Key sensory, circadian, eye development and behavioral genes can drive diel-niche switches***
198 We combined results from the DEG (EdgeR and DESeq2) and gene network analyses (WGCNA) to
199 create a cumulative list of 1700 transcripts (Supplementary Data 6). Focusing on genes that were
200 recovered across *Anisota* and *Dryocampa* reduced the set to 274 transcripts (Supplementary Data 7).
201 Because many transcripts had poorly annotated *Bombyx* hits, we improved annotations using the
202 program eggNog mapper (54). We tested if these genes had GO terms associated with sensory,
203 circadian, brain and neural development, or behavioral regulatory genes (Fig. 4, Table S7,
204 Supplementary Data 8). We found that several genes in each category had associated GO terms, with
205 a predominance of vision and brain development genes (Supplementary Data 9).
206
207 ***Predicted functional regions and homology patterns identified for genes of interest***
208 We examined protein and gene evolution for a set of genes which we found were of interest based on
209 results from DGE, WGCNA, GO annotations (Supplementary Data 10). In order to infer functional
210 homology from structural conservation, we downloaded high quality genomes of Bombycoidea
211 moths and relatives from Darwin Tree of Life [https://www.darwintreeoflife.org/] (Supplementary

212 Data 11). We assigned a reference protein sequence for each gene of interest from the *Bombyx mori*
213 predicted proteome. We used Orthofinder to identify orthologs and filtered orthogroups containing
214 the reference sequence (49, 55). Since we had more than one gene per species, we reduced the
215 number down to a single gene by choosing the highest identity sequence relative to the *B. mori*
216 reference sequence. We also modelled the 3D structure of *B. mori* proteins and mapped the
217 evolutionary conservation onto the 3D predicted structure for proteins above a certain conservation
218 threshold (Supplementary Data 12). These analyses predict structurally and functional conserved
219 regions of proteins (Fig S9, Supplementary Data 13). We repeated this analysis with 38 insect
220 genomes (Supplemental Data 11) and mapped evolutionary conservation onto 3D protein structure
221 (Supplementary Data 12-13). We include results of evolutionary conservation analyses for two
222 regulatory candidates (*disco* and *tk*) that showed varying levels of sequence and protein evolution
223 between insects and moths (Fig. S10).
224
225 ***Modeling predicts additional functional zinc-finger domains for disco in Lepidoptera***
226 *Disco* was recovered across multiple analyses. In *Drosophila melanogaster* it is known for its role in
227 eye development, important for circadian maintenance, and for leg and antennal appendage
228 formation (56–60). To determine if *disco* was conserved between moths and *Drosophila*, we
229 compared the primary sequence and 3D protein structure of *Drosophila* and *Bombyx mori*. In the
230 moth, the sequence length of *disco* was nearly double that of *Drosophila* (Fig 6A). However, a region
231 spanning over 100 amino acids was highly conserved, contained the zinc-finger domain important for
232 its function, and showed strong 3D structural conservation (*Whole protein alignment*: RMSD: align,
233 super) = 37.833, 2.566, MatchAlign score (align, super) = 540 (2234 atoms), 333.1 (554 atoms) vs.
234 *alignment of conserved region*: RMSD=2.699,2.566, MatchAlignScore =443(671 atoms), 351 (615)
235 atoms). This indicates the DNA binding function of *disco* has likely been conserved. However, an
236 additional ~500 amino acid region absent in *Drosophila* is highly conserved across moths, and
237 includes several regions predicted to be functional (Fig. 6A, Fig. S10). We hypothesize that *disco* has
238 a novel role in moths for diel-niche regulation. To further test this, we compared *disco* sequences
239 across *Anisota* and *Dryocampa* and found 23 mutations between them, three of which mapped to the
240 predicted functional region (Fig. 6B, Supplementary Data 14). InterProScan predicted four zinc-
241 finger domains, three in this region, although the CATH-Gene3D databases prediction combined the

242 two separate domains into a single predicted domain (Fig. S11). We also found 53 sites with
243 predicted phosphorylation potential, especially around the second zinc-finger domain (18). Reducing
244 the stringency increased the total to 142, which were still enriched around the second domain.
245

246 Discussion

247 We identified the genetic mechanisms of diel-niche switches by examining gene expression during
248 the day and night in two closely related moth species. We used RNA-Seq to measure gene expression
249 profiles of head tissue and isolated sensory, circadian and neural processes. We found between 300
250 and 700 genes that significantly altered day night circadian expression patterns. The diurnal species
251 had enriched visual perception genes during the day and the nocturnal species had locomotor and
252 olfactory genes upregulated at night.
253 Thirty overlapping DEGs were present in both species, with some DGEs showing divergent
254 patterns of expression matching the species' diel-niche. Examining expression data with a sensitive
255 clustering analysis yielded over 170 genes that showed day-night or species specific co-expressed
256 clusters. We also used GO terms to search for genes associated with sensory, neural, or circadian
257 patterns. We found genes in each category and modeled the 3D structure and predicted functional
258 regions for a subset. The divergently expressed gene *disconnected* (*disco*) was implicated in vision,
259 hearing, locomotion and brain development. Given the multiple lines of evidence supporting the
260 importance of this gene in regulating diurnal and nocturnal activity in moths, we explored the
261 evolution of *disco* by modeling fly and moth *disco* structures. This analysis revealed novel zinc-
262 finger conserved domains in moths, which were lacking in *Drosophila,* surrounded by
263 phosphorylated sites. Several mutations between *Anisota* and *Dryocampa* also mapped to these
264 regions, further strengthening evidence for its role in diel-niche shifts.
265

266 *Visual and olfaction*

267 Visual systems often accompany diel and photic environment shifts (18). For example, nocturnal
268 carpenter bees have much larger facets in their eyes than their diurnal counterparts (61). Nocturnal
269 moths have higher photoreceptor light sensitivity and have neurons more suited to pooling than close
270 diurnal relatives (62). Some of these trends seem to hold across the phylogeny, with diurnal species
271 evolving more complex color visual systems, often reflected in visual opsin sequence evolution (22,

272 23). We did not find diel-expression patterns in color vision opsins, a result corroborated by a recent
273 study (22). However, we discovered a cerebral opsin (ceropsin) upregulated during the day. Ceropsin
274 has been implicated in photoperiodism and is expressed in the brain (63).
275 We also found several eye development genes (*TENM2, ANKRD17, EHD4, JAK2*),
276 phototransduction genes (*PPAP2*, *RDH11*), and retina homeostasis, eye-antennal disc development,
277 and photoreceptor cell maintenance genes (*disco, glass*) (64). Surprisingly a few visual genes, such
278 as *garnet* and *rugose* also appeared to have different isoforms present, showing both day and night
279 upregulation. *Garnet* is an eye color mutant gene in flies (65) and *rugose* is implicated in retinal
280 pattern formation (66).
281 One of the classical differences between day flying butterflies and night flying moths is that
282 butterflies have a clubbed antennae, while moths often have a sensilla-covered antenna; likely a
283 function of increased investment in olfactory sensing, similarly, there is also evidence for increased
284 investment in olfactory sensing in the brain, with relatively larger mushroom bodies in nocturnal
285 species (67, 68). In our diel gene-expression dataset, we found several olfactory genes, including
286 those involved in odorant binding (*Obp84a, Obp58b*), pheromone response (*tk*), mushroom body
287 development (*DAAM2 and DST*) and antennal development (*disco*).
288
289 ***Hearing and mechanoreception***
290 Many moths have active hearing organs, and these have evolved repeatedly across Lepidoptera (69),
291 and their use varies depending on the ecological context of the moth. Several drivers for moth
292 hearing include sexual selection and as a defense against insectivorous bats (70). The primary
293 nocturnal group of butterflies, Hedylidae, has reverted to a nocturnal niche and regained hearing
294 organs (25). Thus, although rudimentary vibrational hearing exists in some diurnal moths (71) it is
295 significantly more developed and used at broader frequency ranges at night, when light is less
296 available. While Saturniidae lack hearing organs (72) they may instead still have vibration
297 mechanosensors that can be useful in evading predation similar to those in locusts and crickets (73–
298 76). Indeed, we did find diel co-expression of several mechanosensory and ear development genes,
299 including *PI4KB, KCTD15, unc-22,* and *RhoGAP92B* (77)
300
301 ***Brain and neural rewiring***

302 Finding an upregulation of neural and brain development was expected since we sequenced head
303 transcriptomes. However, we tested gene expression two days after pupal eclosion, where
304 presumably gross structures are already developed, so it is possible these expressed genes we found,
305 regulate adult plasticity. Adult neural plasticity has been showed in Lepidoptera and other insects
306 (78–80), and we recovered several genes linked to axon regeneration (*APOD*), central complex
307 development (*Ten-a, TENM2, DST, ALDH3A2, OGT*), central nervous system development (*disco*,
308 *RpL4*), and neuropeptide hormonal activity (*tk*). Several genes were specific to retinal ganglion cell
309 axon guidance (*TENM2*) and mushroom body development (*DAAM2*), leading us to speculate that
310 plasticity could occur through neural wiring and plasticity are shaping sensory adaptation. Many
311 Lepidoptera have distinct phenotypic plasticity in wing patterns and coloration in different seasons
312 (81), and there is some evidence that they also have seasonal plasticity in their behavior and foraging
313 preferences (82). Research has found Lepidoptera can override their innate preferences and learning
314 preferences for new visual and olfactory cues after eclosion (83). It is possible the diel-niche and
315 circadian rhythms too have plasticity, there are reports of some moth species like *Hyles lineata*
316 showing relatively labile diel-niches possibly driven by temperature and resources (84–86). These
317 same mechanisms allowing flexibility within a species, might be involved in diel-niche evolution
318 between species.
319
320 ***Circadian and behavioral regulators***
321 The behavioral state of an animal to engage in any activity, e.g., feeding, flight or mating, is likely a
322 function of circadian and behavioral regulators that respond to certain stimuli. We found differential
323 expression of genes involved in locomotion (*unc-22, KCTD15,Tk*) and circadian or rhythmic
324 behavior (*SREBF1/SREBF2, OGT, disco, JAK2*) in both species. We also found several key circadian
325 regulators like *per*, *tim* although they were downregulated only in *Dryocampa*. C*lock-like* or *takeout*
326 another gene under circadian control, (87) was expressed in both species, although without any
327 significant up or down regulation. *Takeout* was moderately conserved in the moths we examined, but
328 had diverged considerably among other insects, to the point where orthology searches with
329 Orthofinder failed to recover orthologs among insects. Even among *D. melanogaster*, its closest
330 homologs were *Jhbp5* and *takeout* which had 21-25% sequence identity (Supplementary Data 10,14).
331 Surprising, it's 3D structure was highly conserved (RMSD (align, super) = 2.556, 2.755,

332 MatchAlign score (align, super) =199 , 642.937, Supplementary Data 14, Fig. S12), highlighting an
333 interesting example of likely functional convergence despite primary sequence divergence.
334
335 ***Energy use and general cellular maintenance***
336 In addition to the genes mentioned above, several key divergently expressed genes were involved in
337 energy utilization. This finding makes biological sense, as once activity is initiated, energy
338 mobilization and upregulation of basic cellular processes need to be maintained. Two of these genes
339 were *SLC175A/MFS10* and *SLC2A6/Tret1-1*, which encode different trehalose sugar transporters.
340 Trehalose is an important sugar present in insect hemolymph (88). Other genes found were ribosomal
341 (*RpL4, mRpS5*), golgi (*GCC*), and mitochondrial maintenance (*PARL/rho7*) along with
342 transcriptional regulation (*Spt5).*
343
344 ***Disco as putative master diel-niche regulator in adult moths***
345 Many genes that we discovered likely play a critical role in maintaining sensory shifts, but in order to
346 identify genes that might regulate upstream, we looked for genes that were: 1) expressed in both
347 species, 2) showed coincident expression patterns with respect to diel-niche, and 3) played a role in
348 sensory and neural functioning and circadian control. Only *disconnected* (*disco*) fit these criteria.
349 *Disco* is a key developmental and patterning gene, first discovered for its role in neural migration
350 from the disconnected optic lobe mutant phenotype (89). This gene is now considered an appendage
351 development gene, with a well characterized pupal role in antennal and leg distal patterning in
352 *Drosophila melanogaster (60)*. Mutants also have a disrupted circadian locomotor rhythm due to the
353 improper formation of neurons that express clock genes (90, 91). As a gene that is involved both in
354 optic lobe, antennal formation, and in neural wiring and circadian control, it is a very strong
355 candidate for driving behavioral diel-niche shifts and sensory adaptations (56, 57, 92). *Disco*
356 expression data from *B. mori* suggests strong adult head, antennal, and nucleus expression,
357 suggesting that it still acts as a transcription factor, and can regulate other genes (40).
358 *Disco* encodes the acid zinc-finger transcription factor and is approximately 500 and 1000
359 amino acids in length in *Drosophila* and *B. mori,* respectively. Our modeling showed that a ~150
360 amino acid region is conserved structurally at the sequence level, and this region also contains zinc-
361 finger motifs associated with DNA binding (93). This indicates that *disco* likely has retained its DNA

binding and likely pupal and appendage patterning function in moths. We predicted the functional regions of *disco* in *Bombyx* based on evolutionary conservation modelling, and found an additional 500 amino acids, absent in *Drosophila*, were predicted to be functional (conserved and exposed). Domain and family level modeling predicted at least two additional zinc-finger domains in this region (Fig. S10,Fig. S11). We also found many phosphorylated sites surrounding the zinc-finger domains. Examining mutations between *Anisota* and *Dryocampa* revealed three mutations mapped to these predicted functional regions (Fig. 6B).

Given *disco's* adult diel-specific expression in moths, additional zinc-finger DNA binding domains, and the high number of phosphorylated sites, we propose it as a candidate transcriptional regulator for diel-regulation in adult moths.

Study species and their evolution and variation in life histories

Anisota and *Dryocampa* are closely related and ecologically similar saturniids of the subfamily Ceratocampinae, thought to have diverged ~3.81 Mya (47). They largely occupy the same range and forest habitat in the Americas, feeding on large trees (*Anisota* on oaks and *Dryocampa* on maples) where population flares cause host plant defoliation (38). Because they reportedly hybridize (94), and pre- and postzygotic barriers are potentially ineffective, temporal partitioning may be important in their evolutionary history.

Conclusion

This study provides a framework for how diel-niche evolution in Lepidoptera can occur. Sensory and neural developmental genes appear to be key. We identify the pivotal role of the transcription factor *disco*, uncovering a second functional region with species-specific mutations in moths. Our findings provide useful targets for further genetic manipulation and highlight the insight non-model systems provide to genetics and development.

Supplementary information

Supplementary tables and data are available for the reviewers and will be uploaded to repository [xxxxx] on submission.

.

Author contributions

Y.S.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, visualization, writing–original draft. R.M: investigation, methodology, project administration, data curation, writing - review & editing. A.J.: conceptualization, formal analysis, writing - review & editing C.S.: conceptualization, data curation, methodology, project administration, supervision. S.C.: formal analysis, software. K.G.: formal analysis, software, visualization, writing-review and editing, D.G.: investigation, project administration, data curation. R.L.: conceptualization, data curation, investigation, methodology, writing-original draft, writing -review & editing. C.H.: conceptualization, data curation, investigation, methodology, project administration, writing - review and editing I.K.: conceptualization, funding acquisition, methodology, resources, writing–review and editing. C.E.: formal analysis, software, visualization, data curation, writing- original draft, writing- review & editing. C.B.: software, validation, writing -review and editing. S.B.: funding acquisition, methodology, validation, writing–review and editing. J.T: conceptualization, funding acquisition, supervision, validation, writing–review & editing. A.K: conceptualization, methodology, resources, writing–review and editing, visualization, supervision, funding acquisition

Acknowledgements

We thank Jesse Breinholt, Martijn Timmermans and Andreas Zwick for help with project conceptualization. Kelly Dexter, Amanda Markee, and members of the McGuire Center for Lepidoptera and Biodiversity at the University of Florida assisted with animal care and wet-lab troubleshooting. We thank Belinda Pinto, Danielle DeLeo, Heather Bracken-Grissom, Jorge Perez-Moreno, Zhou Lei for discussions, advice and assistance with bioinformatics pipelines and wet-lab protocols. We thank Nicolas Alexandre, Sachit Daniel, Riddhi Deshmukh, David Plotkin, Chelsea Skojec and Nitin Ravikanthachari for comments and feedback on the manuscript. Janelle Nunez-Castilla, Jessica Liberles, Raghavan Venket, Yi-Ming Weng, Yu Fahong and Xiaokang Zhang assisted with troubleshooting bioinformatic pipelines and provided analytical recommendations. The authors acknowledge the University of Florida Research Computing for providing computational resources and support that have contributed to the research results reported in this publication (http://researchcomputing.ufl.edu). We acknowledge NBAF and UF-ICBR for assistance with sequencing and generating libraries.

Funding for the project was through NSF DEB-1557007, NSF PRFB-1612862, and NSF IOS-1750833 and NERC NE/P003915/1. The Florida International University Presidential fellowship supported YS.

Figure Legends and Figures:

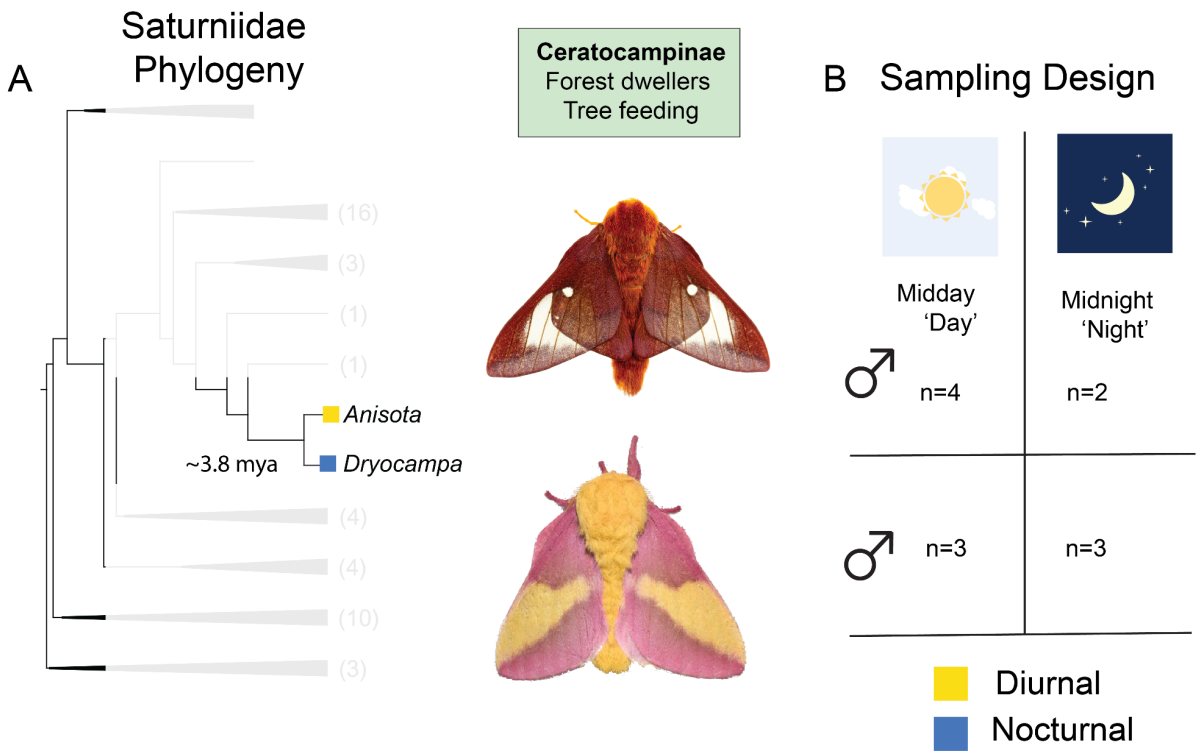

Figure 1: Nocturnal and diurnal moths on a phylogeny with RNA-seq sampling design.

(A): Collapsed phylogeny of Saturniidae, adapted to show where the two study species, diurnal *Anisota pellucida* (Pink-striped oakworm moth) and nocturnal *Dryocampa rubicunda* (Rosy Maple moth) sit. Faded numbers on the tip represent the number of genera in the tree before collapsing. Phylogeny adapted from Rougerie et al. 2022 (**B)**: Sampling design showing the number of replicates for each species and collection period (day/night) sampled. Collection of heads was done 2-days post eclosion at midday (sun) and midnight (moon) and tissue was flash frozen for RNA preservation. Care was taken that the eclosed moths were exposed to a natural light cycle and red lights were used when collecting the moths at night. Photo credits *Anisota pellucida* © Mike Chapman; *Dryocampa rubicunda* (CC) Andy Reacgo and Chrissy Mclearan;

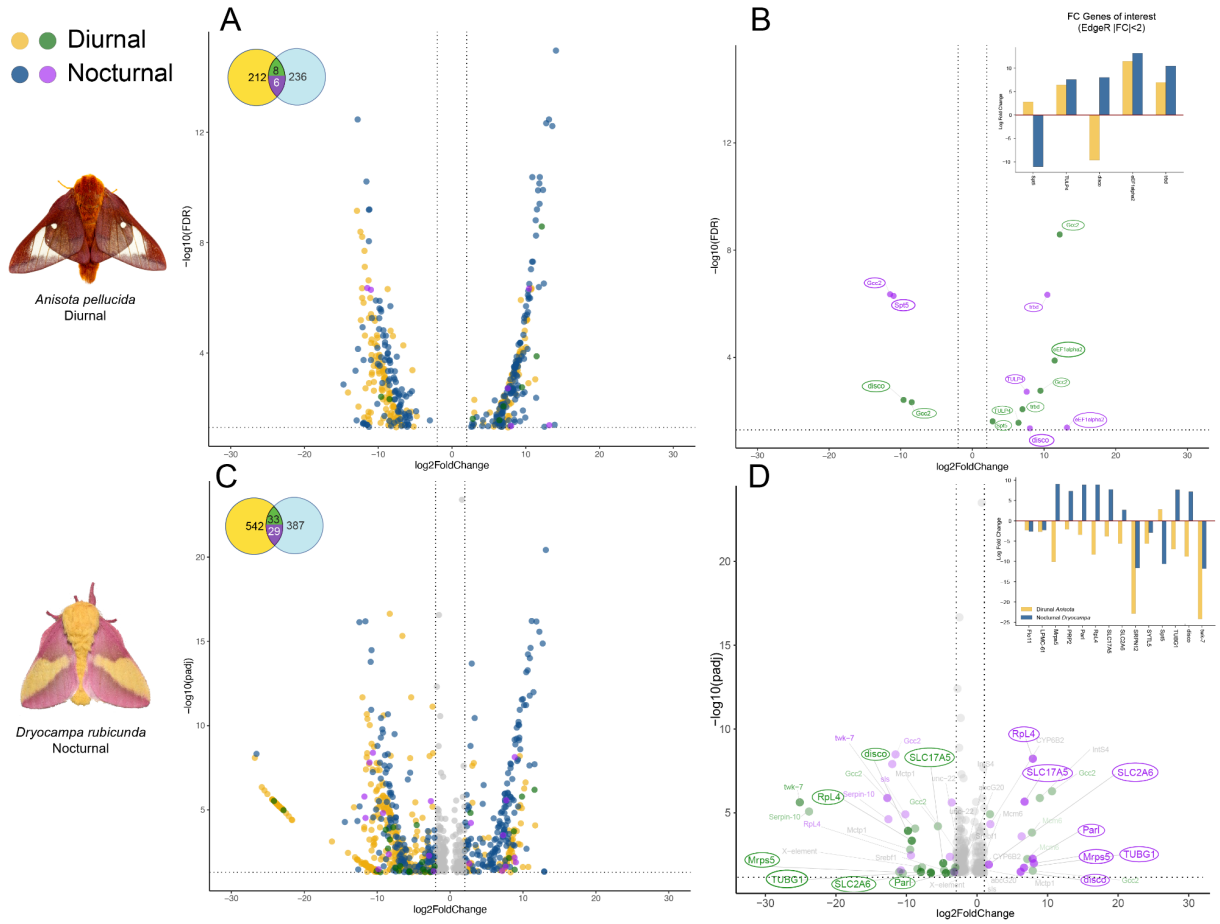

Figure 2: Nocturnal and diurnal species show divergent patterns of gene expression across two different analyses software (EdgeR and DESeq2)

Volcano plots and Venn diagrams showing EdgeR (A-B) and DESeq2 (C-D) results across **day vs. night** sampling times for both nocturnal and diurnal species. Venn diagrams represent the number of differentially expressed genes (DEGs) across both species with the number of common DEGs across each pair. **(A/C)**: Volcano diagrams illustrating fold change and adjusted p-values for the significant differentially expressed genes between midday and midnight samples. Circles in the top left represent the number of genes expressed in both species and the colors correspond to FC values for those genes in the volcano plots. Yellow and blue represent genes expressed only in the nocturnal or diurnal species, green/purple indicates DEGs present in both species. **(B/D)**: Only genes expressed in both species are shown and genes that display opposite trends in expression are highlighted. See Supplementary Material for a detailed list of DEGs. Positive fold change indicates night overexpression and negative fold change indicates day overexpression. Genes had FDR or adjusted p-value < 0.05 and FC > |2|. Identification of common genes was done using orthogroup clustering with Orthofinder with *Bombyx mori*. Gene names and annotations were transferred from *B. mori.*

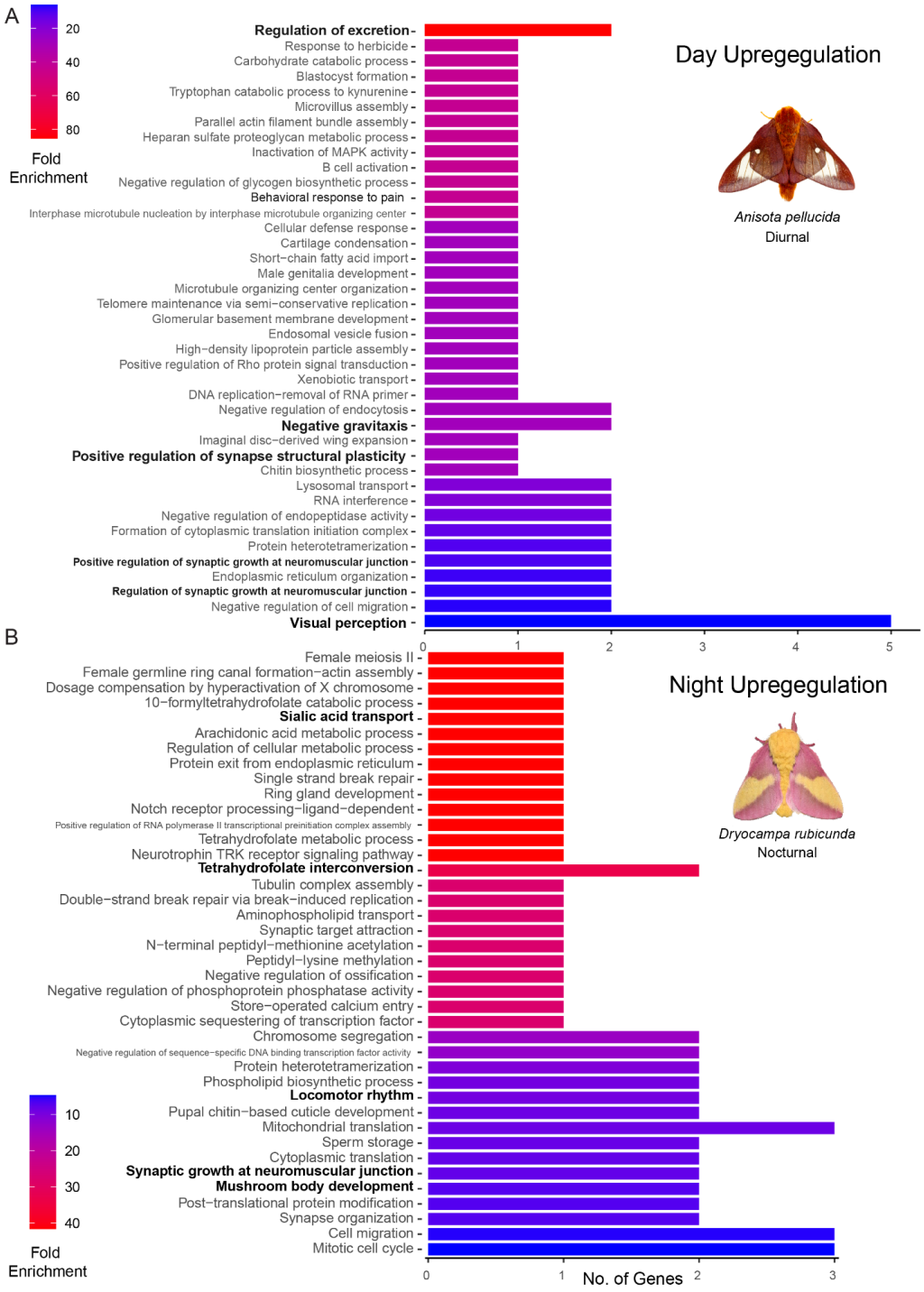
A
20
40
60
80
Fold Enrichment
Regulation of excretion
Response to herbicide
Carbohydrate catabolic process
Blastocyst formation
Tryptophan catabolic process to kynurenine
Microvillus assembly
Parallel actin filament bundle assembly
Heparan sulfate proteoglycan metabolic process
Inactivation of MAPK activity
B cell activation
Negative regulation of glycogen biosynthetic process
Behavioral response to pain
Interphase microtubule nucleation by interphase microtubule organizing center
Cellular defense response
Cartilage condensation
Short−chain fatty acid import
Male genitalia development
Microtubule organizing center organization
Telomere maintenance via semi−conservative replication
Glomerular basement membrane development
Endosomal vesicle fusion
High−density lipoprotein particle assembly
Positive regulation of Rho protein signal transduction
Xenobiotic transport
DNA replication−removal of RNA primer
Negative regulation of endocytosis
Negative gravitaxis
Imaginal disc−derived wing expansion
Positive regulation of synapse structural plasticity
Chitin biosynthetic process
Lysosomal transport
RNA interference
Negative regulation of endopeptidase activity
Formation of cytoplasmic translation initiation complex
Protein heterotetramerization
Positive regulation of synaptic growth at neuromuscular junction
Endoplasmic reticulum organization
Regulation of synaptic growth at neuromuscular junction
Negative regulation of cell migration
Visual perception
0
1
2
3
4
5
Day Upregegulation
Anisota pellucida
Diurnal
B
Female meiosis II
Female germline ring canal formation−actin assembly
Dosage compensation by hyperactivation of X chromosome
10−formyltetrahydrofolate catabolic process
Sialic acid transport
Arachidonic acid metabolic process
Regulation of cellular metabolic process
Protein exit from endoplasmic reticulum
Single strand break repair
Ring gland development
Notch receptor processing−ligand−dependent
Positive regulation of RNA polymerase II transcriptional preinitiation complex assembly
Tetrahydrofolate metabolic process
Neurotrophin TRK receptor signaling pathway
Tetrahydrofolate interconversion
Tubulin complex assembly
Double−strand break repair via break−induced replication
Aminophospholipid transport
Synaptic target attraction
N−terminal peptidyl−methionine acetylation
Peptidyl−lysine methylation
Negative regulation of ossification
Negative regulation of phosphoprotein phosphatase activity
Store−operated calcium entry
Cytoplasmic sequestering of transcription factor
Chromosome segregation
Negative regulation of sequence−specific DNA binding transcription factor activity
Protein heterotetramerization
Phospholipid biosynthetic process
Locomotor rhythm
Pupal chitin−based cuticle development
Mitochondrial translation
Sperm storage
Cytoplasmic translation
Synaptic growth at neuromuscular junction
Mushroom body development
Post−translational protein modification
Synapse organization
Cell migration
Mitotic cell cycle
Night Upregegulation
Dryocampa rubicunda
Nocturnal
10
20
30
40
Fold Enrichment
0
1
2
3
No. of Genes

Figure 3: Visual genes are upregulated in the diurnal species during the day, and energy utilization, brain olfactory region and locomotion in the nocturnal species during the night.
Go enrichment of highly upregulated genes recovered from both DEG analyses coinciding with the species. highlighting the two modules that showed species specific clustering patterns.
A: Enrichment of **day** upregulated genes in diurnal *Dryocampa*. B Enrichment of **night** upregulated genes in nocturnal *Anisota.* Go enrichment was done using the custom ShinyGo database v0.75c using *Bombyx* gene IDs. FDR cutoff <=0.17, and only Biological Process GO terms were selected with Min. pathway size =1. Input genes had FC>5, padj. <0.05 and recovered both in EdgeR and DeSeq2 were used with the background being all orthologous *Bombyx mori* genes for either species.

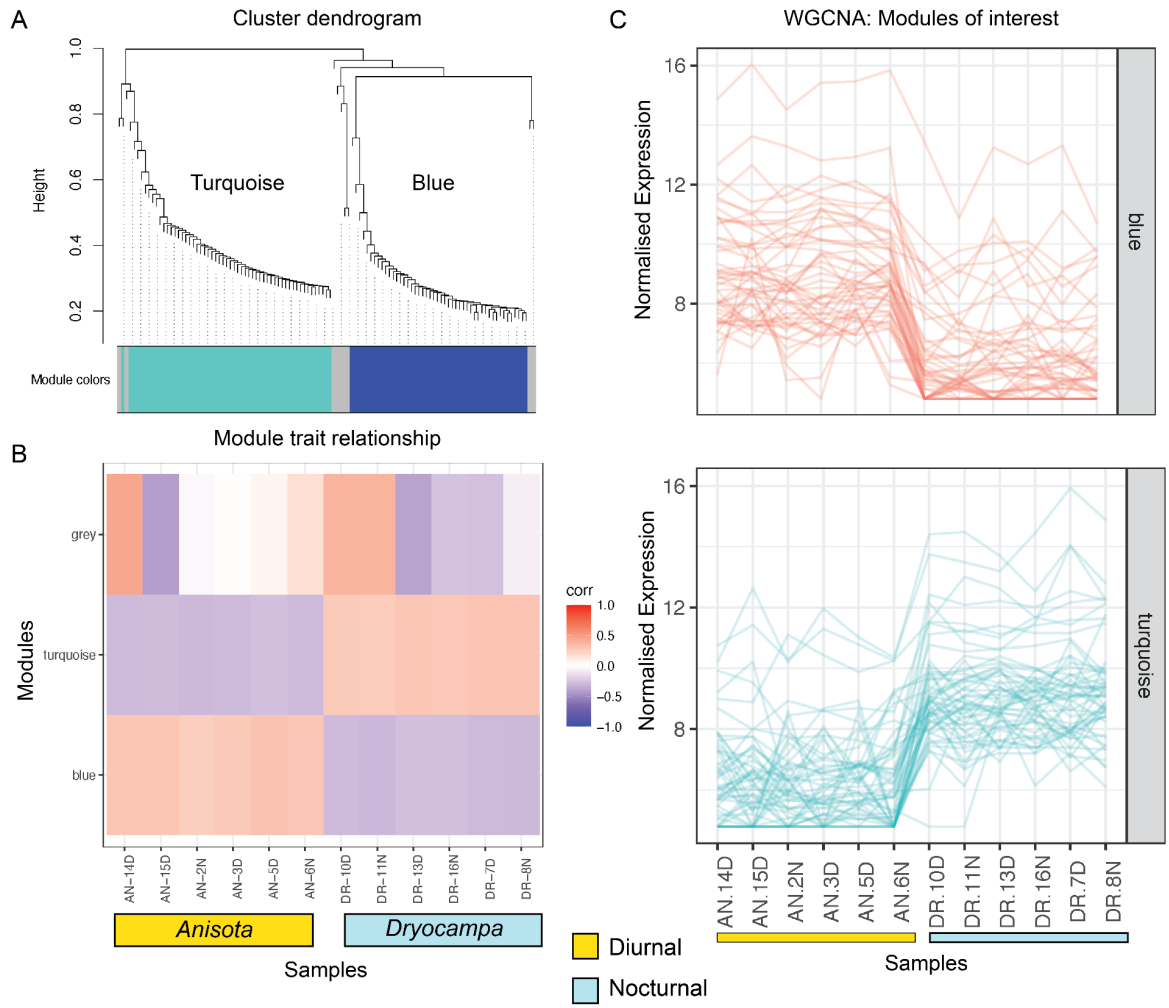

Figure 4: Modules of clustered co-expressed genes grouped using normalized expression highlighting the two modules that showed species specific clustering patterns.
A: Cluster dendrogram shows WGCNA clusters. B: Shows how patterns of gene expression correlate

474 across samples and modules. C: Shows the normalized expression for all
475 genes across samples. The species show two sets of 50 genes (Blue) and (Turquoise) that have clear
476 species specific expression patterns. Normalization was done with DESeq2 and reads were mapped
477 to the more stringently filtered transcriptome. A soft power analysis was done and the picked
478 power=9 for the WGCNA analysis. AN: *Anisota*, DR: *Dryocampa,* Numbered D and N represent the
479 different samples collected at day and night time points.

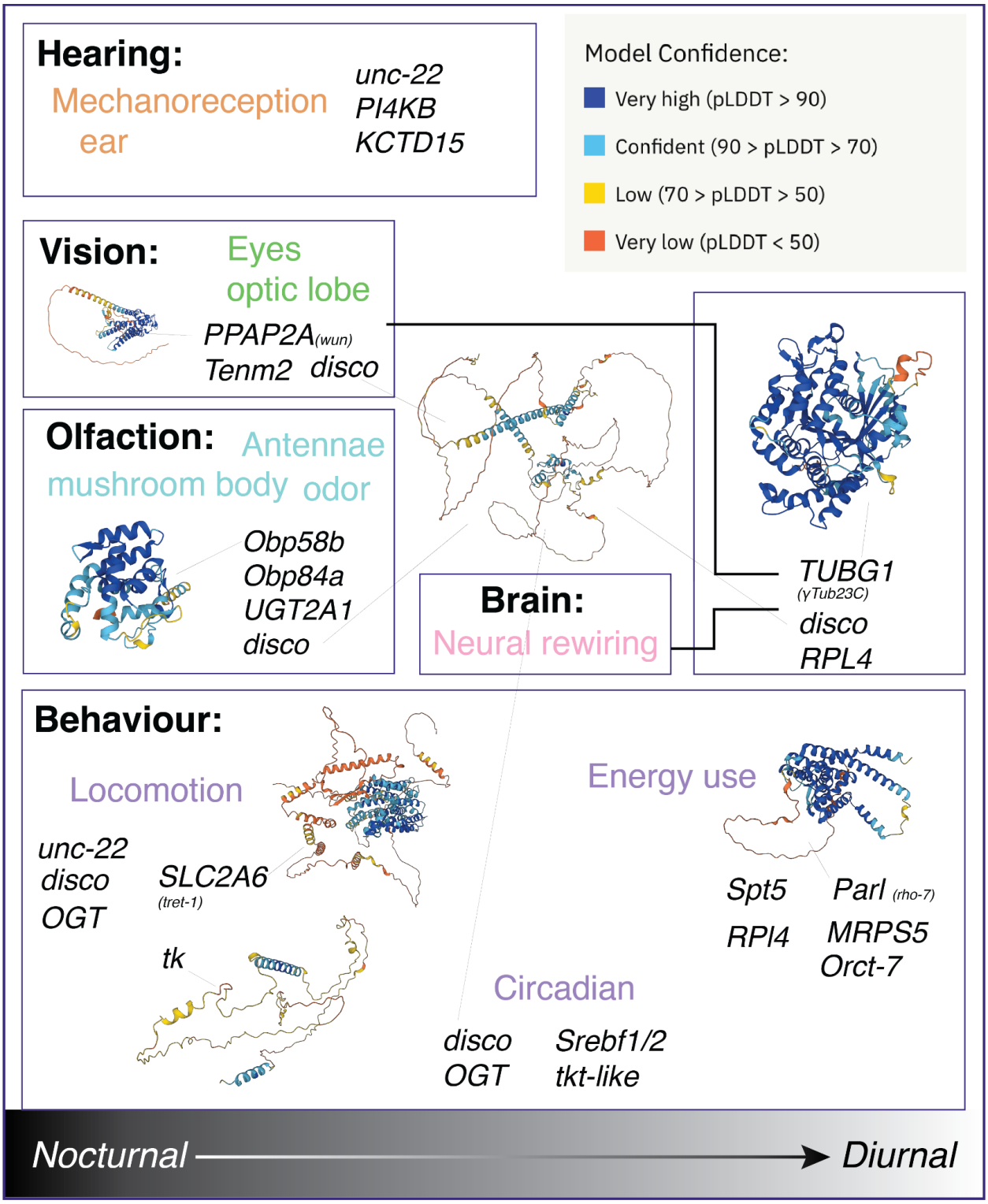

Figure 5: Model for diel-niche shift from and genes of interest obtained from GO searches and annotating divergently expressed genes. For representative purposes, homologs from *Drosophila*

melanogaster have been used, with colors in the 3d structures representing Alpha fold per-residue confidence metric (pLDDT), the range for each color is shown on the top left. Model moth *(Bombyx mori)* proteins were also modeled for a subset of proteins, see Supplementary data.

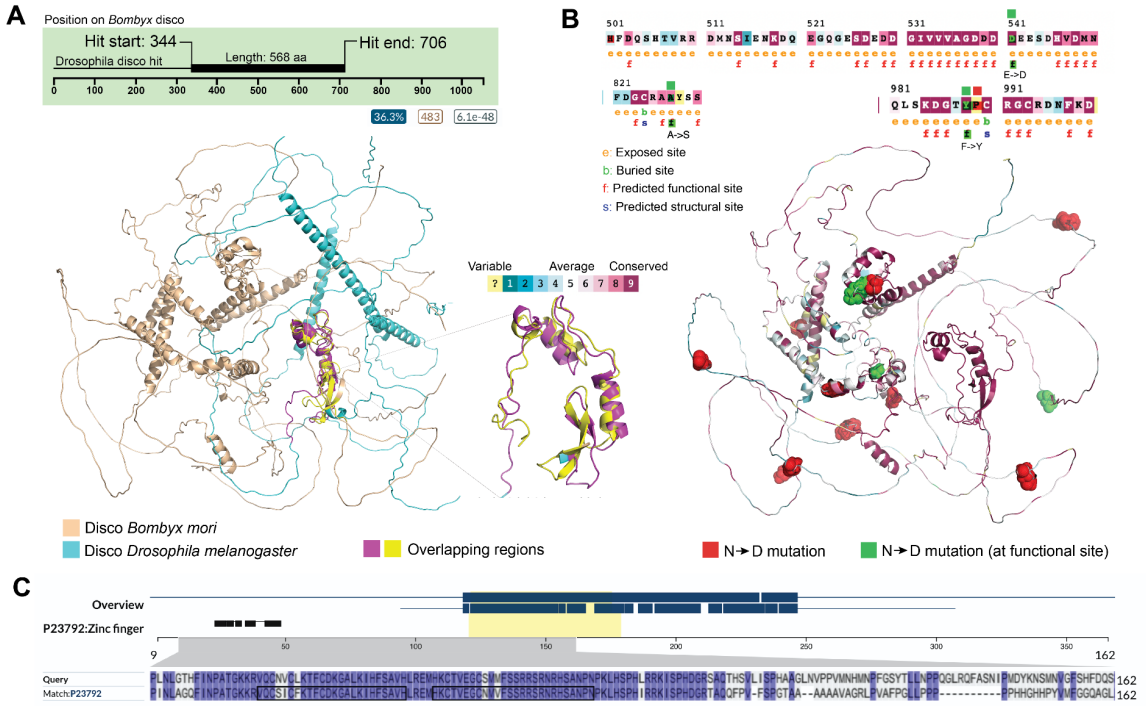

Figure 6: *Disco* has 150 amino acid long highly conserved region across *Bombyx* (1054 aa) & *Drosophila* (568 aa) involved in its role as a transcription factor, but it also has other roles in Lepidoptera due to the presence of other highly conserved and predicted functional regions, that are not present in flies, and only in Lepidoptera. 3/14 of the mutations between nocturnal to diurnal species map to these regions (>500 aa).
(A) Top: *Disco* best Uniprot hit in *Drosophila* using default settings (blastp, e-threshold 10, Auto-Blosum62). Bottom: *disco* aligned and superimposed predicted structures for the silk moth and vinegar fly. Peach structure: *Bombyx mori* alpha-fold predicted structure for *disco,* Cyan: *Drosophila melanogaster* alpha-fold predicted structure for *disco* **(B)** Top: Partial views of the Consurf predicted residues for *disco.* The views encompassing the region where 3 mutations of the 14 mutated sites between Nocturnal *Dryocampa* and Diurnal *Anisota* overlap with predicted functional residues. Bottom: Residues that were mutated between the nocturnal and diurnal species are highlighted on an overlay of the Consurf scores mapped onto the predicted alpha fold structure of disco from *Bombyx mori.* Green residues indicate mutated and predicted functional sites, red indicate mutated sites that did not have a predicted residue. **(C)** Overlap of the highly conserved region of *disco* this region includes the zinc-finger domain that is characteristic of disco transcription factor.

505

506 # Methods

507 **Insect rearing:**

508 Moths were reared under natural light-dark cycles at room temperature (25°C) and natural light
509 conditions at the McGuire Center for Lepidoptera and Biodiversity (MGCL). Adults were sampled
510 two days after eclosion at the two time points by C.H. and R.S.L.

511

512 **Sampling design:**

513 We collected 3-5 replicates (i.e., individual moths) per experimental treatment. We sampled at two
514 time points per species at two midpoints of their circadian cycle at noon ('day') and midnight
515 ('night') (Fig. 1, Table 1). Moths were allowed to eclose naturally and sexed, To ensure vision genes
516 were not artificially stimulated, for collection at night we used red light, which is thought to not
517 stimulate most Lepidoptera insect visual systems (95). We used a sterilized scalpel to remove the
518 head, which was placed in a 1.5 ml Eppendorf tube, which was immediately placed in liquid nitrogen
519 and stored at -80°C. It is worth noting that only males of most *Anisota* species are diurnal, including
520 *A. pellucida*, our focal taxon. Females of all *Anisota* and both sexes of at least *A. stigma*, are
521 nocturnal. In *Dryocampa*, both males and females are nocturnal. We therefore limited our
522 transcriptomic work to males of *A. pellucida* and *D. rubicunda*.

523

524 **RNA extraction and quality control:**

525 Tissue was removed from liquid nitrogen and placed into a flat-bottomed cryo-tube with Trizol and
526 2-4 sterile, stainless-steel beads. The tissue was homogenized using a bead beater for 2.5 minutes at
527 1900 RPM. Beads were removed with a magnet, and chloroform added to separate phases.
528 Isopropanol precipitated the RNA, followed by several ethanol wash steps to pellet the RNA. Quality
529 Control (QC) was undertaken with a high sensitivity Qubit assay. Samples with yields > 20 ng/µL
530 were cleaned using a Thermo Scientific RapidOut DNA Removal Kit (Waltham, Massachusetts,
531 USA). Sample quality was verified using Agilent 2200 Tapestation. Since heads comprised of a
532 relatively small amount of tissue, initial extractions using kit-based methods resulted in low yields.
533 Several rounds of troubleshooting were required before completion of extraction using a modified
534 Trizol based protocol with a DNA clean up step. We measured the quantity of RNA using a Qubit

535 fluorometer and used Agilent tapestation or Bioanalyzer to characterize extraction quality, weight,
536 and size distributions. Where possible, a concentration of more than 10 ng/µL was used for library
537 preparation. Yields below 10 ng/µL were prone to higher loss when purifying with a DNAse agent,
538 which often caused 20-30 % loss in RNA yield, approaching 5ng/µL is below the Qubit detection
539 threshold.
540
541 **Library preparation and NGS sequencing:**
542 Libraries were prepared by sequencing cores at NBAF-Liverpool using the NEB polyA selection and
543 NEBNext Ultra II directional stranded kit, suitable for low input yields. 12 samples were run on one
544 lane of an Illumina NovaSeq using SP chemistry (Paired-end, 2x150 bp sequencing). Samples were
545 shipped overnight from MCGL to the NERC-NBAF after dehydrating in a biosafety chamber using
546 GenTegra RNA tubes.
547
548 **Read trimming and cleanup:**
549 For *Anisota pellucida* and *Dryocampa rubicunda*, trimming was undertaken by the NERC-NBAF
550 core and the raw Fastq files were trimmed for the presence of Illumina adapter sequences using
551 Cutadapt version 1.2 (96). The option -O 3 was used, so the 3' end of any reads that match the
552 adapter sequence for 3 bp or more were trimmed. Trimming was also done with trimmomatic. The
553 reads are further trimmed using Sickle version 1.200 with a minimum window quality score of 20.
554 Reads shorter than 15 bp after trimming were removed.
555
556 **Transcriptome library sizes, QC, and PCA:**
557 Quality Control (QC) was conducted on the trimmed reads, library size varied for each species (Fig.
558 S1). We examined expression data and removed genes with TPM < 1. PCA results showed that
559 *Anisota* and *Dryocampa* did not separate with diel-niche (Fig. S3).
560
561 ***De novo* assemblies:**
562 We combined reads from multiple samples and generated several reference de novo transcriptomes
563 using different assemblers, and then combining them, compared their quality using BUSCO scores
564 and number of single-copy BUSCO genes across the different versions (see Table 2 for the BUSCO

565 scores, redundancy and duplication for the various assemblies). We found that unfiltered assemblies
566 were highly redundant, but had the highest BUSCO scores. These assemblies were too
567 overrepresented to be used in any downstream analysis, so we used Transrate, CD-HIT, MMeqs2, as
568 well as Transdecoder, to filter the assemblies. We chose v5 and v6 to use in downstream analysis,
569 listed as the last two assemblies for each species (Table 3) and repeated certain analyses with both
570 assemblies to see how stringent or a less stringent filter would affect the results.
571
572 **Multi assembler *de novo* transcriptome assembly:**
573 To our knowledge, there is no publicly available annotated reference genome for the two species
574 used in this study at the time the analyses were performed; therefore, we chose to build a de novo
575 transcriptome assembly with the pre-processed reads using the NCGAS pipeline, which combines
576 multiple assemblers and then uses an evidence-based gene modeler to decide on the final set. For
577 each species, we combined forward and reverse reads from all samples and normalized them using a
578 normalization script in Trinity v2.12.0 (97). Normalized reads were used as input to generate
579 assemblies using Velvet v1.2.10 (98), TransAbyss v2.0.1 (99) , SOAPdenovo v1.03 (100) with
580 various kmer sizes. These assemblies were combined and filtered using Evigenes v4: 2020, March
581 (101, 102), transdecoder and filtered with scripts provided in the NCGAS pipeline (103).
582
583 **Transcriptome QC, annotation and filtering:**
584 Further quality control and filtering was conducted using BUSCO (104), transrate (105), QUAST
585 (106) (v5.02), CD-HIT(v4.6.8) (107, 108)and cascaded clustering with MMseqs2 (v12) (109).
586 BUSCO scores were greater than 95%, but many genes had multiple versions from the different
587 assemblers, and because this would bias the downstream analyses we attempted to reduce the
588 redundancy at the cost of lowering BUSCO scores. Protein coding regions were predicted using
589 TransDecoder (https://github.com/TransDecoder/TransDecoder) and the cds file from was run
590 through Transrate (105), which further pulled contigs based on mapping rate and identified a "good"
591 collection of transcripts, the resulting protein-coding sequences were filtered with additional CD-HIT
592 (107, 108)and MMseqs2 (109) filtering. This resulted in lower BUSCO scores 75-85% but also much
593 lower redundancy in the gene set of about 1-3 fold. We ran the WGCNA (52) and GO enrichment
594 (110) analyses on the more stringent assemblies (BUSCO 75%, 2% genes with duplication) since

they are more sensitive to redundancy, but RNA-seq was run across the slightly lower stringency analyses to be more inclusive (BUSCO~85%, 60-70% genes with duplications).

Annotation of reference transcriptomes:

Reference genomes were annotated with eggNOGmapper and Orthofinder (54, 55). We also tried annotating the reference transcriptomes with Trinotate (111), which run against the swissprot and pfam databases using hmmscan, blastx, blastp, signalP and tmmhmm, KEGG, and GO and eggNOG (49, 110, 112–120). ORF predictions for orthofinder were obtained using *Antheraea pernyi* as a model. Orthofinder was run with well-annotated Bombycidae and Saturniidae moths, *Bombyx mori*, *Antheraea pernyi*, and *Antheraea yamamai*, to generate clusters. *Bombyx mori* was a useful reference since the Orthofinder cluster for each species generated a list of mostly single-copy genes for which more Lepidoptera annotations were available from SilkDBv3 (40). While this was not as complete as trinotate, the annotations provided from this procedure were more accurate than Trinotate. Trinotate annotation had many more human/ vertebrate hits than insect matches than eggNOGmapper and Orthofinder transferred annotations, so while the files are provided as a reference, they were not used in downstream analyses.

Differential gene expression (DGE):

Reads were mapped using Salmon (121) and differential gene expression analysis was conducted using EdgeR (122) and DESeq2 (123). EdgeR(v 3.38.1) was used to normalize and test for significantly expressed genes. DESeq2(v1.36.0) was also used to normalize and to generate PCAs and other summary statistics. We used ’Day’ as the treatment and ’Night’ as the control for all groups with EdgeR, although the order of fold change switched for DESeq2. The p-values obtained were adjusted to account for multiple hypothesis testing with FDR for EdgeR and adjusted p-value from DESeq2. Genes with FDR or adjusted p-values <0.05 and $|FC| > 2$ were used as a criteria to identify significantly expressed genes.

Gene enrichment analysis:

We performed gene enrichment analysis using GO terms with the tools TopGO v2.48.0 (124) and ShinyGo v0.75c (125) (http://bioinformatics.sdstate.edu/goc/). This analysis involved comparing the

625 selected genes of interest to a gene universe with GO term annotations to determine if there were
626 overrepresented or underrepresented GO terms. To obtain genes with GO annotations, we utilized
627 similarity clusters to map annotations from *Bombyx mori*.
628 In TopGO, we employed a more stringent approach by using filtered transcriptomes to
629 mitigate bias from gene duplications. We selected genes of interest based on a False Discovery Rate
630 (FDR) threshold of less than 0.05. We ran two different enrichment algorithms, 'classic' and 'elim',
631 and used two test statistics, Fisher's exact test and a Kolmogorov-Smirnov-like test. The tests were
632 performed for both 'Biological Process' (BP) and 'Molecular Function' (MF) categories. While we
633 generated tables of significantly enriched GO terms, obtaining individual genes within each group
634 was limited due to the constraints of TopGO with custom annotations.
635 For ShinyGO, we utilized the custom v0.75c which included the updated *Bombyx mori*
636 genome stringent. For all analysis, we used the following settings (pathway size: min.=1,
637 max=2000). The settings for number of pathways to show and FDR cut-offs were chosen to represent
638 the entire list of top 100 genes in the final datasets, although in some cases a fewer number are
639 represented in the visualizations. We focused on the 'Biological Process' (BP) category for the
640 various tests. Tables of GO terms, gene information, and graphs summarizing the significantly
641 enriched GO terms were generated. We repeated several different analyses **1)** A less stringent
642 analysis was run where genes with FC<|2| and padj. <0.05. This was useful to visualize the genes up
643 and downregulated functionally and the background used was all the genes recovered from the DEA
644 analysis that had orthologs. An FDR cut-off <0.1-0.3. For the background gene set, we used the
645 species' de-novo transcriptome matching *Bombyx* ortholog set.
646 **2)** A more stringent criteria was used to examine highly upregulated genes a FC<|5| and padj.
647 <0.05 cutoff was used, and filtering only genes that occurred in both EdgeR and DESeq2 analyses.
648 We selected a single representative when multiple Bombyx orthologs were recovered, using the
649 ortholog with the most complete annotation. An FDR cut-off <0.17 was used. For the background
650 gene set, we used the species' de-novo transcriptome matching Bombyx ortholog set.
651 **3)** For WGCNA clusters tan and grey60, we used their respective species background
652 expressed transcript set with a FDR cutoff <= 0.1. For blue and turquoise, the background of all
653 orthologs found in *Bombyx mori* was used. An FDR cutoff<1 was used, since smaller values
654 provided insufficient genes given the larger number of genes being tested.

655

656 **Gene network analysis:**

657 Gene network analysis was undertaken with WGCNA (1.71) (52, 126). DeESeq2 was used to
658 normalize reads and data was formatted with tidyverse. WGCNA identifies modules of co-expressed
659 genes. We chose modules that showed clear day-night differences and mapped the GO terms with
660 Revigo and ShinyGo. We also combined counts from both species and tested for clusters. Here we
661 chose modules that showed species specific clusters.

662

663 **Gene annotation and filtering:**

664 DEGs, enriched genes and or co-expressed transcripts/ genes were annotated by cross referencing
665 them with the *B. mori* annotations (see above). Since *B. mori* mappings were not always 'one-to-
666 one', mappings that were 'one-to-many' were dropped, i.e., transcripts that mapped to multiple
667 *Bombyx* genes were dropped. We also improved annotation by mapping the overlapping transcripts
668 with eggnog mapper. We further filtered these genes into sub functional categories, by identifying
669 genes with GO annotations or descriptions linked to vision, smell, hearing, circadian, behavior and
670 brain the list of go terms for each group was obtained by querying amigo
671 (http://amigo.geneontology.org/) (Table 8).

672

673 **Gene Mining and in-silico evolution:**

674 We mined genes of interest from ~bombycoidea moths and closely related families that had well
675 annotated genomes on Ensembl and NCBI (Supp Table 6) and across model insects (Supp. Table 7).
676 These genomes were chosen as all Bombycoidea from Ensembl with well annotated assemblies and
677 five species Noctuidae and Geometridae each were included as close relatives. We added *Bombyx*
678 *mori* as a reference and *Antheraea pernyi* to represent Saturniidae, and their peptide files were taken
679 from their respective source papers (40–42). For insects, we chose long-read well annotated genomes
680 from Ensembl (Ensembl Gene build) and five model systems from insect base (127). We ran
681 Orthofinder (v2.5.2) (49, 55) with 'orthofinder -f bom-rel/ -S diamond -M msa -A mafft -T fasttree -a
682 1 -X -z -t 16' to recover orthologs. Each orthogroup often contained multiple sequences per species.
683 We chose the sequence with the highest identity to the reference *Bombyx mori* sequence using
684 custom python scripts. In the rare instances where there were multiple references, we took the longest

one. To model the 3D structure of each reference *Bombyx mori* sequence, we ran AlphaFold v2.1.2 (128) using Deepmind's run_alphafold.py (https://github.com/deepmind/alphafold/blob/main/run_alphafold.py) on University of Florida's HiperGator. Model database files (i.e. BFD, MAGnify, PDB70, mmCIF PDB, UniRef30, UniRef90) downloaded from Deepmind in February 2022 were used as parameter value inputs. All other parameters were used with default settings. The relaxed predicted model with highest ranked pLDDT (per-residue estimate corresponding to model's predicted score on the lDDT-Cα metric) was chosen as the final model (129).

We calculated a conservation score for each alignment using Alistat. For alignments less than 1500 amino acids and with high conservation score (Proteins larger than 1300 aa were difficult to model without high memory GPU's), we modeled conservation using Consurf (130–134). The results were displayed using Jmol first glance viewer (http://firstglance.jmol.org/) and as screenshots from the default viewer. The output for alpha fold runs and consurf is available in the (Supplementary Data). PyMol (135) was used to align the 3D structures and Aliview was used to compare alignments. We used InterProscan (https://www.ebi.ac.uk/interpro/about/interproscan/) and NetPhos3.1 (136, 137) for predicting the domains and phosphorylated sites.

References

1. Y. Xiao, Y. Yuan, M. Jimenez, N. Soni, S. Yadlapalli, Clock proteins regulate spatiotemporal organization of clock genes to control circadian rhythms. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021).

2. K. Beer, C. Helfrich-Förster, Model and Non-model Insects in Chronobiology. *Front. Behav. Neurosci.* **14**, 601676 (2020).

3. D. Székely, D. Cogălniceanu, P. Székely, M. Denoël, Adult-Juvenile interactions and temporal niche partitioning between life-stages in a tropical amphibian. *PLoS One* **15**, e0238949 (2020).

4. D. Wang, G. Yang, W. Chen, Diel and Circadian Patterns of Locomotor Activity in the Adults of Diamondback Moth (Plutella xylostella). *Insects* **12** (2021).

5. J. H. Fullard, N. Napoleone, Diel flight periodicity and the evolution of auditory defences in the Macrolepidoptera. *Anim. Behav.* **62**, 349–368 (2001).

6. R. Refinetti, The diversity of temporal niches in mammals. *Biol. Rhythm Res.* **39**, 173–192 (2008).

7. A. Y. Kawahara, *et al.*, Diel behavior in moths and butterflies: a synthesis of data illuminates the evolution of temporal activity. *Organisms Diversity and Evolution* **18**, 13–27 (2018).

716 8. D. T. C. Cox, A. S. Gardner, K. J. Gaston, Diel niche variation in mammals associated with
717 expanded trait space. *Nat. Commun.* **12**, 1753 (2021).

718 9. Y. Basset, N. D. Springate, Diel activity of arboreal arthropods associated with a rainforest tree. *J.*
719 *Nat. Hist.* **26**, 947–952 (1992).

720 10. P. J. Devries, G. T. Austin, N. H. Martin, Diel activity and reproductive isolation in a diverse
721 assemblage of Neotropical skippers (Lepidoptera: Hesperiidae). *Biol. J. Linn. Soc. Lond.* **94**, 723–
722 736 (2008).

723 11. S. R. Anderson, J. J. Wiens, Out of the dark: 350 million years of conservatism and evolution in diel
724 activity patterns in vertebrates. *Evolution* **71**, 1944–1959 (2017).

725 12. S. Frey, J. T. Fisher, A. C. Burton, J. P. Volpe, Investigating animal activity patterns and temporal
726 niche partitioning using camera-trap data: challenges and opportunities. *Remote Sens. Ecol. Conserv.*
727 **3**, 123–132 (2017).

728 13. R. Steen, Diel activity, frequency and visit duration of pollinators in focal plants: in situ automatic
729 camera monitoring and data processing. *Methods Ecol. Evol.* **8**, 203–213 (2017).

730 14. H. K. Gill, G. Goyal, R. McSorley, Diel Activity of Fauna in Different Habitats Sampled at the
731 Autumnal Equinox. *Fla. Entomol.* **95**, 319–325 (2012).

732 15. B. Murugavel, A. Kelber, H. Somanathan, Light, flight and the night: effect of ambient light and
733 moon phase on flight activity of pteropodid bats. *J. Comp. Physiol. A Neuroethol. Sens. Neural*
734 *Behav. Physiol.* **207**, 59–68 (2021).

735 16. H. Somanathan, *et al.*, Nocturnal bees feed on diurnal leftovers and pay the price of day – night
736 lifestyle transition. *Front. Ecol. Evol.* **8** (2020).

737 17. R. M. Borges, H. Somanathan, A. Kelber, Patterns and Processes in Nocturnal and Crepuscular
738 Pollination Services. *Q. Rev. Biol.* **91**, 389–418 (2016).

739 18. S. M. Tierney, *et al.*, Consequences of evolutionary transitions in changing photic environments.
740 *Austral Entomology* **56**, 23–46 (2017).

741 19. W. T. Wcislo, *et al.*, The evolution of nocturnal behaviour in sweat bees, Megalopta genalis and M.
742 ecuadoria (Hymenoptera: Halictidae): an escape from competitors and enemies? *Biol. J. Linn. Soc.*
743 *Lond.* **83**, 377–387 (2004).

744 20. A. L. Stockl, W. A. Ribi, E. J. Warrant, Adaptations for Nocturnal and Diurnal Vision in the
745 Hawkmoth Lamina. *J. Comp. Neurol.* **524**, 160–175 (2016).

746 21. Y. Wu, *et al.*, Retinal transcriptome sequencing sheds light on the adaptation to nocturnal and
747 diurnal lifestyles in raptors. *Sci. Rep.* **6**, 1–12 (2016).

748 22. T. Akiyama, H. Uchiyama, S. Yajima, K. Arikawa, Y. Terai, Parallel evolution of opsin visual
749 pigments in hawkmoths by tuning of spectral sensitivities during transition from a nocturnal to a
750 diurnal ecology. *J. Exp. Biol.* **225**, jeb244541 (2022).

751 23. Y. Sondhi, E. Elis, S. Bybee, J. Theobald, A. Kawahara, Light environment drives evolution of color
752 vision genes in butterflies and moths (2021) https:/doi.org/10.5061/dryad.gmsbcc2kr (May 13,
753 2022).

754 24. M. Izutsu, A. Toyoda, A. Fujiyama, K. Agata, N. Fuse, Dynamics of Dark-Fly Genome Under
755 Environmental Selections. *G3* **6**, 365–376 (2015).

756 25. J. E. Yack, J. H. Fullard, Ultrasonic hearing in nocturnal butterflies. *Nature* **403**, 265–266 (2000).

757 26. F. Sandrelli, R. Costa, C. P. Kyriacou, E. Rosato, Comparative analysis of circadian clock genes in
758 insects. *Insect Mol. Biol.* **17**, 447–463 (2008).

759 27. R. Závodská, I. Sauman, F. Sehnal, Distribution of PER protein, pigment-dispersing hormone,
760 prothoracicotropic hormone, and eclosion hormone in the cephalic nervous system of insects. *J. Biol.*
761 *Rhythms* **18**, 106–122 (2003).

762 28. C. Hermann, *et al.*, The circadian clock network in the brain of different Drosophila species. *J.*
763 *Comp. Neurol.* **521**, 367–388 (2013).

764 29. C. A. Hamilton, *et al.*, Phylogenomics resolves major relationships and reveals significant
765 diversification rate shifts in the evolution of silk moths and relatives. *BMC Evol. Biol.* (2019)
766 https:/doi.org/10.1186/s12862-019-1505-1.

767 30. A. Kobelková, R. Závodská, I. Sauman, O. Bazalová, D. Dolezel, Expression of clock genes period
768 and timeless in the central nervous system of the Mediterranean flour moth, Ephestia kuehniella. *J.*
769 *Biol. Rhythms* **30**, 104–116 (2015).

770 31. D. Brady, A. Saviane, S. Cappellozza, F. Sandrelli, The Circadian Clock in Lepidoptera. *Front.*
771 *Physiol.* **12**, 776826 (2021).

772 32. S. Iwai, Y. Fukui, Y. Fujiwara, M. Takeda, Structure and expressions of two circadian clock genes,
773 period and timeless in the commercial silkmoth, Bombyx mori. *J. Insect Physiol.* **52**, 625–637
774 (2006).

775 33. A. Macias-Muñoz, A. G. Rangel Olguin, A. D. Briscoe, Evolution of Phototransduction Genes in
776 Lepidoptera. *Genome Biol. Evol.* **11**, 2107–2124 (2019).

777 34. X.-C. Jiang, *et al.*, Identification of Olfactory Genes From the Greater Wax Moth by Antennal
778 Transcriptome Analysis. *Front. Physiol.* **12**, 663040 (2021).

779 35. A. de Fouchier, *et al.*, Functional evolution of Lepidoptera olfactory receptors revealed by
780 deorphanization of a moth repertoire. *Nat. Commun.* **8**, 15709 (2017).

781 36. J. W. Truman, L. M. Riddiford, Neuroendocrine control of ecdysis in silkmoths. *Science* **167**, 1624–
782 1626 (1970).

783 37. I. Sauman, S. M. Reppert, Circadian clock neurons in the silkmoth Antheraea pernyi: novel
784 mechanisms of Period protein regulation. *Neuron* **17**, 889–900 (1996).

785 38. P. M. Tuskes, J. P. Tuttle, M. M. Collins, *The Wild Silk Moths of North America: A Natural History*
786 *of the Saturniidae of the United States and Canada* (Cornell University Press, 1996).

787 39. M. M. Collins, P. M. Tuskes, REPRODUCTIVE ISOLATION IN SYMPATRIC SPECIES OF
788 DAYFLYING MOTHS (HEMILEUCA: SATURNIIDAE). *Evolution* **33**, 728–733 (1979).

789 40. F. Lu, *et al.*, SilkDB 3.0: visualizing and exploring multiple levels of data for silkworm. *Nucleic*
790 *Acids Res.* **48**, D749–D755 (2020).

791 41. S.-R. Kim, *et al.*, Genome sequence of the Japanese oak silk moth, Antheraea yamamai: the first
792 draft genome in the family Saturniidae. *Gigascience* **7**, 1–11 (2018).

793 42. J. Duan, *et al.*, A chromosome-scale genome assembly of Antheraea pernyi (Saturniidae,
794 Lepidoptera). *Mol. Ecol. Resour.* **20**, 1372–1383 (2020).

795 43. H. Langer, G. Schmeinck, F. Anton-Erxleben, Identification and localization of visual pigments in
796 the retina of the moth, Antheraea polyphemus (Insecta, Saturniidae). *Cell Tissue Res.* **245**, 81–89
797 (1986).

798 44. I. Kitching, *et al.*, A global checklist of the Bombycoidea (Insecta: Lepidoptera). *Biodiversity Data*
799 *Journal* **6**, e22236 (2018).

800 45. M. J. Scoble, *The Lepidoptera. Form, function and diversity* (Oxford University Press, 1992).

801 46. Lemaire, Minet, 18. The Bombycoidea and their Relatives. *Volume 1: Evolution, Systematics, and*
802 (2013).

803 47. R. Rougerie, *et al.*, Phylogenomics Illuminates the Evolutionary History of Wild Silkmoths in Space
804 and Time (Lepidoptera: Saturniidae). *bioRxiv*, 2022.03.29.486224 (2022).

805 48. Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev.*
806 *Genet.* **10**, 57–63 (2009).

807 49. D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons
808 dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).

809 50. D. Li, *et al.*, An evaluation of RNA-seq differential analysis methods. *PLoS One* **17**, e0264246
810 (2022).

811 51. N. Sánchez-Baizán, L. Ribas, F. Piferrer, Improved biomarker discovery through a plot twist in
812 transcriptomic data analysis. *BMC Biol.* **20**, 208 (2022).

813 52. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC*
814 *Bioinformatics* **9**, 559 (2008).

815 53. Lafayette Lafayette Press, *A Survey of Best Practices for RNA-Seq Data Analysis* (CreateSpace
816 Independent Publishing Platform, 2016).

817 54. C. P. Cantalapiedra, A. Hernández-Plaza, I. Letunic, P. Bork, J. Huerta-Cepas, eggNOG-mapper v2:

818 Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale.
819 *Mol. Biol. Evol.* **38**, 5825–5829 (2021).

820 55. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics.
821 *Genome Biol.* **20**, 238 (2019).

822 56. P. E. Hardin, J. C. Hall, M. Rosbash, Behavioral and molecular analyses suggest that circadian
823 output is disrupted by disconnected mutants in D. melanogaster. *EMBO J.* **11**, 1–6 (1992).

824 57. M. S. Dushay, M. Rosbash, J. C. Hall, The disconnected visual system mutations in Drosophila
825 melanogaster drastically disrupt circadian rhythms. *J. Biol. Rhythms* **4**, 1–27 (1989).

826 58. K. J. Lee, M. Mukhopadhyay, P. Pelka, A. R. Campos, H. Steller, Autoregulation of the Drosophila
827 disconnected gene in the developing visual system. *Dev. Biol.* **214**, 385–398 (1999).

828 59. V. Suri, Z. Qian, J. C. Hall, M. Rosbash, Evidence that the TIM light response is relevant to light-
829 induced phase shifts in Drosophila melanogaster. *Neuron* **21**, 225–234 (1998).

830 60. B. K. Dey, X.-L. Zhao, E. Popo-Ola, A. R. Campos, Mutual regulation of the Drosophila
831 disconnected (disco) and Distal-less (Dll) genes contributes to proximal-distal patterning of antenna
832 and leg. *Cell Tissue Res.* **338**, 227–240 (2009).

833 61. H. Somanathan, A. Kelber, R. M. Borges, R. Wallén, E. J. Warrant, Visual ecology of Indian
834 carpenter bees II: adaptations of eyes and ocelli to nocturnal and diurnal lifestyles. *J. Comp. Physiol.*
835 *A Neuroethol. Sens. Neural Behav. Physiol.* **195**, 571–583 (2009).

836 62. A. L. Stöckl, D. O'Carroll, E. J. Warrant, Higher-order neural processing tunes motion neurons to
837 visual ecology in three species of hawkmoths. *Proceedings of the Royal Society B: Biological*
838 *Sciences* **284**, 20170880 (2017).

839 63. I. Shimizu, Y. Yamakawa, Y. Shimazaki, T. Iwasa, Molecular cloning of Bombyx cerebral opsin
840 (Boceropsin) and cellular localization of its expression in the silkworm brain. *Biochem. Biophys.*
841 *Res. Commun.* **287**, 27–34 (2001).

842 64. K. Moses, M. C. Ellis, G. M. Rubin, The glass gene encodes a zinc-finger protein required by
843 Drosophila photoreceptor cells. *Nature* **340**, 531–536 (1989).

844 65. R. Tearle, Tissue specific effects of ommochrome pathway mutations in Drosophila melanogaster.
845 *Genet. Res.* **57**, 257–266 (1991).

846 66. H. K. Shamloula, *et al.*, rugose (rg), a Drosophila A kinase anchor protein, is required for retinal
847 pattern formation and interacts genetically with multiple signaling pathways. *Genetics* **161**, 693–710
848 (2002).

849 67. A. Stöckl, *et al.*, Differential investment in visual and olfactory brain areas reflects behavioural
850 choices in hawk moths. *Sci. Rep.* **6** (2016).

851 68. A. Sourakov, R. W. Chadd, *The Lives of Moths: A Natural History of Our Planet's Moth Life*
852 (Princeton University Press, 2022).

69. A. Y. Kawahara, *et al.*, Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22657–22663 (2019).

70. J. J. Rubin, *et al.*, The evolution of anti-bat sensory illusions in moths. *Science Advances* (2018) https:/doi.org/10.1126/sciadv.aar7428.

71. J. E. Yack, L. D. Otero, J. W. Dawson, A. Surlykke, J. H. Fullard, Sound production and hearing in the blue cracker butterfly Hamadryas feronia (Lepidoptera, nymphalidae) from Venezuela. *J. Exp. Biol.* **203**, 3689–3702 (2000).

72. J. R. Barber, *et al.*, Anti-bat ultrasound production in moths is globally and phylogenetically widespread. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2117485119 (2022).

73. G. A. Jacobs, J. P. Miller, Z. Aldworth, Computational mechanisms of mechanosensory processing in the cricket. *J. Exp. Biol.* **211**, 1819–1828 (2008).

74. J. M. Camhi, Locust wind receptors: I. Transducer mechanics and sensory response. *J. Exp. Biol.* **50**, 335–348 (1969).

75. J. M. Camhi, Locust wind receptors. 3. Contribution to flight initiation and lift control. *J. Exp. Biol.* **50**, 363–373 (1969).

76. J. C. Tuthill, R. I. Wilson, Mechanosensation and Adaptive Motor Control in Insects. *Curr. Biol.* **26**, R1022–R1038 (2016).

77. J. D. Baker, S. Adhikarakunnathu, M. J. Kernan, Mechanosensory-defective, male-sterile unc mutants identify a novel basal body protein required for ciliogenesis in Drosophila. *Development* **131**, 3411–3422 (2004).

78. D. D. Dell'Aglio, W. O. McMillan, S. H. Montgomery, Shifting balances in the weighting of sensory modalities are predicted by divergence in brain morphology in incipient species of Heliconius butterflies. *Anim. Behav.* **185**, 83–90 (2022).

79. S. H. Montgomery, R. M. Merrill, S. R. Ott, Brain composition in Heliconius butterflies, posteclosion growth and experience-dependent neuropil plasticity. *J. Comp. Neurol.* **524**, 1747–1769 (2016).

80. S. Anton, W. Rössler, Plasticity and modulation of olfactory circuits in insects. *Cell Tissue Res.* **383**, 149–164 (2021).

81. S. Halali, *et al.*, Predictability of temporal variation in climate and the evolution of seasonal polyphenism in tropical butterfly communities. *J. Evol. Biol.* **34**, 1362–1375 (2021).

82. W. Kuenzinger, *et al.*, Innate colour preferences of a hawkmoth depend on visual context. *Biol. Lett.* **15**, 20180886 (2019).

83. A. Kelber, Pattern discrimination in a hawkmoth: innate preferences, learning performance and ecology. *Proc. Biol. Sci.* **269**, 2573–2577 (2002).

887 84. G. T. Broadhead, T. Basu, M. von Arx, R. A. Raguso, Diel rhythms and sex differences in the
888 locomotor activity of hawkmoths. *J. Exp. Biol.* **220**, 1472–1480 (2017).

889 85. M. Bischoff, R. A. Raguso, A. Jürgens, D. R. Campbell, Context-dependent reproductive isolation
890 mediated by floral scent and color. *Evolution* **69**, 1–13 (2015).

891 86. S. Jaeger, C. Girvin, N. Demarest, E. LoPresti, Secondary pollinators contribute to reproductive
892 success of a pink-flowered sand verbena population. *Ecology* **104**, e3977 (2023).

893 87. W. V. So, *et al.*, takeout, a novel Drosophila gene under circadian clock transcriptional regulation.
894 *Mol. Cell. Biol.* **20**, 6935–6944 (2000).

895 88. T. Kikawada, *et al.*, Trehalose transporter 1, a facilitated and high-capacity trehalose transporter,
896 allows exogenous trehalose uptake into cells. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11585–11590
897 (2007).

898 89. H. Steller, K. F. Fischbach, G. M. Rubin, Disconnected: a locus required for neuronal pathway
899 formation in the visual system of Drosophila. *Cell* **50**, 1139–1153 (1987).

900 90. C. Helfrich-Förster, Robust circadian rhythmicity of Drosophila melanogaster requires the presence
901 of lateral neurons: a brain-behavioral study of disconnected mutants. *J. Comp. Physiol. A* **182**, 435–
902 453 (1998).

903 91. L. R. Sanders, M. Patel, J. W. Mahaffey, The Drosophila gap gene giant has an anterior segment
904 identity function mediated through disconnected and teashirt. *Genetics* **179**, 441–453 (2008).

905 92. E. Blanchardon, *et al.*, Defining the role of Drosophila lateral neurons in the control of circadian
906 rhythms in motor activity and eclosion by targeted genetic ablation and PERIOD protein
907 overexpression. *Eur. J. Neurosci.* **13**, 871–888 (2001).

908 93. A. D. Keller, T. Maniatis, Only two of the five zinc fingers of the eukaryotic transcriptional
909 repressor PRDI-BF1 are required for sequence-specific DNA binding. *Mol. Cell. Biol.* **12**, 1940–
910 1949 (1992).

911 94. S. Johnson, N. Boob, A strange pair. *News of the Lepidopterists' Society* **55**, 116 (2013).

912 95. C. J. van der Kooi, D. G. Stavenga, K. Arikawa, G. Belušič, A. Kelber, Evolution of Insect Color
913 Vision: From Spectral Sensitivity to Visual Ecology. *Annu. Rev. Entomol.* **66**, 435–461 (2021).

914 96. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads.
915 *EMBnet.journal* **17**, 10–12 (2011).

916 97. M. G. Grabherr, *et al.*, Full-length transcriptome assembly from RNA-Seq data without a reference
917 genome. *Nat. Biotechnol.* **29**, 644–652 (2011).

918 98. D. R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs.
919 *Genome Res.* **18**, 821–829 (2008).

920 99. G. Robertson, *et al.*, De novo assembly and analysis of RNA-seq data. *Nat. Methods* **7**, 909–912

921 (2010).

922 100. R. Luo, *et al.*, SOAPdenovo2: an empirically improved memory-efficient short-read de novo
923 assembler. *Gigascience* **1**, 18 (2012).

924 101. D. G. Gilbert, Longest protein, longest transcript or most expression, for accurate gene
925 reconstruction of transcriptomes? *bioRxiv*, 829184 (2019).

926 102. D. Gilbert, Gene-omes built from mRNA seq not genome DNA (2016)
927 https:/doi.org/10.7490/f1000research.1112594.1 (June 6, 2022).

928 103. , *https://github.com/NCGAS/de-novo-transcriptome-assembly-pipeline* (Github) (June 6, 2022).

929 104. M. Manni, M. R. Berkeley, M. Seppey, F. A. Simão, E. M. Zdobnov, BUSCO Update: Novel and
930 Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of
931 Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).

932 105. R. Smith-Unna, C. Boursnell, R. Patro, J. M. Hibberd, S. Kelly, TransRate: reference-free quality
933 assessment of de novo transcriptome assemblies. *Genome Res.* **26**, 1134–1144 (2016).

934 106. A. Mikheenko, A. Prjibelski, V. Saveliev, D. Antipov, A. Gurevich, Versatile genome assembly
935 evaluation with QUAST-LG. *Bioinformatics* **34**, i142–i150 (2018).

936 107. W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or
937 nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

938 108. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation
939 sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).

940 109. M. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis
941 of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).

942 110. Gene Ontology Consortium, The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids*
943 *Res.* **49**, D325–D334 (2021).

944 111. D. M. Bryant, *et al.*, A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of
945 Limb Regeneration Factors. *Cell Rep.* **18**, 762–776 (2017).

946 112. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: interactive sequence similarity
947 searching. *Nucleic Acids Research* **39**, W29–W37 (2011).

948 113. M. Punta, *et al.*, The Pfam protein families database. *Nucleic Acids Res.* **40**, D290–301 (2012).

949 114. R. D. Finn, *et al.*, The Pfam protein families database: towards a more sustainable future. *Nucleic*
950 *Acids Res.* **44**, D279–85 (2016).

951 115. T. N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides
952 from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).

953 116. A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, Predicting transmembrane protein
954 topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580
955 (2001).

956 117. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool.
957 *J. Mol. Biol.* **215**, 403–410 (1990).

958 118. M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, M. Tanabe, KEGG for integration and
959 interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–14 (2012).

960 119. M. Ashburner, *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology
961 Consortium. *Nat. Genet.* **25**, 25–29 (2000).

962 120. S. Powell, *et al.*, eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different
963 taxonomic ranges. *Nucleic Acids Res.* **40**, D284–9 (2012).

964 121. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-
965 aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

966 122. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential
967 expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

968 123. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-
969 seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

970 124. Alexa, Rahnenfuhrer, topGO: Enrichment analysis for Gene Ontology. R package version 2.28. 0.
971 *Cranio* (2016).

972 125. S. X. Ge, D. Jung, R. Yao, ShinyGO: a graphical gene-set enrichment tool for animals and plants.
973 *Bioinformatics* **36**, 2628–2629 (2020).

974 126. P. Langfelder, S. Horvath, Fast R Functions for Robust Correlations and Hierarchical Clustering.
975 *J. Stat. Softw.* **46** (2012).

976 127. Y. Mei, *et al.*, InsectBase 2.0: a comprehensive gene resource for insects. *Nucleic Acids Res.* **50**,
977 D1040–D1045 (2022).

978 128. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–
979 589 (2021).

980 129. V. Mariani, M. Biasini, A. Barbato, T. Schwede, lDDT: a local superposition-free score for
981 comparing protein structures and models using distance difference tests. *Bioinformatics* **29**, 2722–
982 2728 (2013).

983 130. B. Yariv, *et al.*, Using evolutionary data to make sense of macromolecules with a "face-lifted"
984 ConSurf. *Protein Sci.* **32**, e4582 (2023).

985 131. H. Ashkenazy, *et al.*, ConSurf 2016: an improved methodology to estimate and visualize
986 evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–50 (2016).

132. G. Celniker, *et al.*, ConSurf: Using evolutionary data to raise testable hypotheses about protein function. *Isr. J. Chem.* **53**, 199–206 (2013).

133. H. Ashkenazy, E. Erez, E. Martz, T. Pupko, N. Ben-Tal, ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–33 (2010).

134. M. Landau, *et al.*, ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–302 (2005).

135. Schrödinger, LLC, The PyMOL Molecular Graphics System, Version 1.8 (2015).

136. N. Blom, S. Gammeltoft, S. Brunak, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**, 1351–1362 (1999).

137. N. Blom, T. Sicheritz-Pontén, R. Gupta, S. Gammeltoft, S. Brunak, Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633–1649 (2004).