Manuscript type: Research Article

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

Authors: Yash Sondhi1[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: Insects have been longstanding models for understanding circadian and sensory biology. The diversity of diel-activity patterns of insects is tremendous, yet our understanding of the genetic and evolutionary processes leading to this diversity is limited. Light environment is a powerful circadian entrainer that drives diel-niche and sensory evolution. To investigate the evolutionary impact of highly shifted light environments on animals with otherwise similar ecologies, we compared gene expression in closely related day- and night-active wild silkmoths.

Genes linked to eye development, neural plasticity and energy cellular maintenance showed expression patterns linked to diel activity. Notably, *disco*, a gene encoding a zinc finger transcription factor involved in fly (*Drosophila melanogaster)* optic lobe and antennal development in the pupa, 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:

Circadian rhythms drive many biological patterns, such as activity periods. The temporal partitioning that results can reduce predation, minimize competition, or enable new resource utilization. It can also drive the evolution of sensory systems, such as the highly specialized antennae with which male moths find mates, and the visual specializations with which butterflies escape from birds. However, the mechanisms that underlie such niche-shift adaptations are poorly understood. Many model organisms in circadian biology are too distantly related from one another to elucidate the genetic drivers of diel-niche transitions. To address this, we examined gene expression patterns between two closely related wild silk moth species (Saturniidae), that utilize similar host plants and habitats, but 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. We also found clusters of circadian, sensory, and brain development genes co-expressing with diel-activity. Eight genes showed a clear and statistically significant correlation between expression and diel activity in both species, and were involved in vision, olfaction, brain development, neural plasticity, energy utilization, and cellular maintenance. Disconnected (*disco)*, a zinc-finger transcription factor involved in antennal development, circadian activity, and optic lobe brain development in flies, was repeatedly recovered in multiple analyses. While *disco* mutants have circadian arrhythmia, most studies have attributed this to improper clock neuron development, and not directly to adult circadian maintenance. Comparing predicted 3D protein structure across moth (*Bombyx mori)* and fly (*Drosophila melanogaster)* genetic models revealed *disco* has likely retained its developmental function with a conserved zinc finger domain, but it has also likely gained novel functional zinc

finger domains, highly conserved in moths, but absent in *D. melanogaster.* These regions had several mutations between nocturnal and diurnal study species that co-occur with higher levels of predicted phosphorylation sites. Together, robust circadian gene expression, functional nocturnal and diurnal mutations with high phosphorylation, structural and sequence conservation, leads us to hypothesize that *disco* may be a master regulator contributing to the diversification of adult diel-activity patterns 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 silkmoths: 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 sequenced head tissue at different time periods to identify expression in sensory, circadian, and neural genes that correlate with diel activity. Expression in eight genes clearly and significantly correlates with diel activity, with functions in vision, olfaction, brain development, neural plasticity, energy utilization, and cellular maintenance. Of these, a single gene emerges consistently across analyses: *disco*, which encodes a zinc-finger transcription factor. In moths, it has likely retained functions also found in flies (eye and brain development during the pupal stage) through a conserved zinc finger domain. But it has also gained an extra zinc-finger domain, surrounded by phosphorylated sites and with mutations in both nocturnal and diurnal species, which may contribute to adult moth circadian regulation.

Results

We compared gene expression across two wild silk moth species, *Anisota pellucida* and *Dryocampa rubicunda*, whose males are diurnal and nocturnal, respectively (Fig. 1, Table 1). We generated head (eyes, brain) transcriptomes from moths collected and flash frozen at midday and midnight, referred to as ‘day’ and ‘night’ hereafter. Using multiple programs to assemble high-quality *de novo* assemblies (Table 2), we characterized the level of gene (mRNA) expression with quasi-read mapping, which can be used to correlate protein expression (48).

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

We found 350 & 393 significantly differentially expressed genes (DEGs) when comparing day and night treatments for each species (Table 3, Fig. 2A). *Anisota* had more day-upregulated genes (56%) and *Dryocampa* was slightly more night-upregulated (53%). In order to compare DEG sets between species, we mapped our DEGs to *Bombyx mori* orthologs using the Orthofinder software (49) (Supplementary Data 1). Approximately 60% of DEGs from each species had identifiable orthologs in *B. mori* (Table 3), and only a small number of DEGs (6-8 genes) overlapped between both species (Fig. 2B). We also replicated this analysis using other software to assure that our results were robust, despite different normalization methods (50). With DESeq2, we found 498 & 697 DEGs (Fig. 2C, Table 3, Supplementary Data 2), with similar *Bombyx* annotation rates (61%), although the proportion of day upregulated genes increased considerably in the *Anisota* (79%) compared to being more even split in *Dryocampa* (50%) (Table 3). The total number of overlapping genes increased (19-26, Fig. 2D) when using DESeq2. A comparison of the two methods revealed that 174 and 216 genes were shared between *Anisota* and *Dryocampa*, respectively.

Divergently expressed genes are linked to brain optic lobe, antennal and neural development

To identify important regulators involved in diel-niche evolution, we applied two filtering criteria to our gene expression data. First, we selected genes that exhibited highly significant differential expression in both species. Second, we focused on genes that displayed upregulation patterns consistent with the natural diel-activity of each species. Our rationale was that this subset of genes was more likely to contain key regulators. To compare DEG overlap between the two species, we grouped the transcripts to their matching orthologs from *Bombyx mori*; if two transcripts from different species mapped to the same ortholog, we treated them as being the same. This allowed us to examine overlapping genes between the species to see if any genes switched fold-change sign from positive to negative or vice versa. (Fig. 2, Supplementary Data 3). We found 51 overlapping DEG transcripts that mapped to 28 unique *Bombyx* genes. Nine genes showed flipped patterns of expression between the two species, and eight of them coincided with known diel activity patterns (Table 7). Examining gene ontology (GO) annotations and comparing orthologs from flybase (https://flybase.org/), we found genes linked to optic lobe and antennal development (*disco*),

locomotion and energy use (*SLC2A6, SLC17A5*), brain and neural development (*TUBG1*) and other essential biological processes like transcription, ribosomal translation, protein processing, mitochondrial maintenance (*RpS4, PARL, Mrps5*) and wound response (*PRP2*) (Table 7, Fig. 2B, D). Of these, only *disco* was recovered with both methods.

Gene network analysis identifies diel activity and species-specific co-expressed clusters

Identifying highly expressed genes helps understand which genes are activated during particular biological processes. However, determining only those that are highly expressed can often overlook genes with important biological functions (51). We examined co-expressed genes that may be correlated with diel-niche or RNA collection time. We used WGCNA, a weighted correlation network analysis tool to cluster genes together based on their normalized counts (52). After examining co-expression patterns for each species separately, we found two modules in each species that clustered with day-night treatment (cluster-*grey60* and cluster-*tan*) (Fig. S6, Supplementary data 4). Since we were interested in species-specific differences, we reran analyses and combined reads from *Anisota* and *Dryocampa,* using only normalized counts for genes that had valid *Bombyx* annotations for both species. This approach narrowed our focus to 2000 genes. Among these, we discovered two clusters (cluster-blue and cluster-turquoise) consisting of 50 genes each that exhibited different expression patterns across species (Fig. 4).

GO enrichment of photoperiodism, circadian control, muscle and neural growth genes

We used a gene enrichment analysis to determine if gene ontology (GO) terms were significantly overrepresented in the DEG and WGCNA sets compared to the appropriate background of GO terms. Using TopGo, which allows custom gene sets, we found an overrepresentation of genes involved in several biological functions (Supplementary Table 1). Those that seemed significant for diel-niche and vision were ‘response to stimulus’ and ‘smooth muscle control’, as well as folic acid serine, glycine and retinoic acid metabolism. We also used ShinyGo to examine WGCNA clusters (Supplementary Data 5). We matched orthologous genes in the less duplicated, filtered transcriptomes to obtain identifications of the most related *Bombyx* genes (Table 4). We examined the enrichment of both tan and grey60 modules, listing the non-redundant terms using ReviGo (Fig. S7, Supplementary Table 2, 3). Gene clusters that co-expressed in the same direction together in the

day and night treatments of both species included photoperiodism, circadian control, negative phototaxis and nervous system development. We next checked for the enrichment of the modules that showed species-specific patterns (blue and turquoise, Fig. S8, Supplementary Table 2). These included genes involved in muscle proliferation and nerve growth, neural signaling, glycolysis, oxidative stress response, and basic cellular functioning such as protein processing and transcriptional regulation.

Day upregulation of vision genes in the diurnal moth

Since both EdgeR and DESeq2 analyses use different normalization methods and statistical model assumptions (50, 53), we repeated enrichment analyses by combining datasets and examining genes that appeared in both analyses. For the *Anisota,* we tested over-enrichment of a smaller subset (FC ≤ -5) of diurnally highly upregulated genes (Fig. 3, Table 5). We found gene enrichment for visual perception, excretion regulation, negative gravitaxis, synaptic plasticity, along with genes associated with other biological processes, such as RNA interference, endopeptidase activity and endocytosis. A reduction in stringency (FC ≤- 2) did not alter results considerably (Fig. S4). Night-upregulated genes (FC ≥ 2) included ocellar pigment genes, eye-photoreceptor cell development, snRNA processing, post embryonic development and neurotransmitter secretion, among a host of other processes that may be required for growth, development and metabolism (wnt signaling, tricarboxylic acid cycle, and cellular response to insulin, glucose transport) (Fig. S4).

Night upregulation of antennal and olfactory brain regions mushroom development genes

We repeated the same analyses for *Dryocampa* and tested highly nocturnally upregulated genes (FC ≥ 5). Our results show upregulation in genes known to be associated with mushroom body development, locomotor rhythm, synaptic growth, energy utilization (Sialin transport) and mitochondrial translation (Fig. 3B, Table 6). A reduction in stringency (|FC| ≥ 2) showed entrainment of the clock cycle, and antennal development genes. Genes associated with innate immune response, DNA repair, cell division, histone acetylation, circadian rhythm, retinoid cycle were upregulated during the day, possibly indicating a period of cellular repair during a time when these moths are inactive (Fig. S5).

Key sensory, circadian, eye development and behavioral genes can drive diel-niche switches

We combined results from the DEG (EdgeR and DESeq2) and gene network analyses (WGCNA) to create a cumulative list of 1700 transcripts (Supplementary Data 5, 6). Focusing on genes that were recovered across *Anisota* and *Dryocampa* reduced the set to 274 transcripts (Supplementary Data 7). Because many transcripts had poorly annotated *Bombyx* hits, we improved annotations using the program eggNog mapper (54). We tested if these genes had GO terms associated with sensory, circadian, brain and neural development, or behavioral regulatory genes (Fig. 4, Table 7, Supplementary Data 8). We found that several genes in each category had associated GO terms, with a predominance of vision and brain development genes (Supplemental Table 4).

Predicted functional regions and homology patterns identified for genes of interest

We examined protein and gene evolution for a set of genes which we found were of interest based on results from DGE, WGCNA, GO annotations (Supplemental Table 5). In order to infer functional homology from structural conservation, we downloaded high quality genomes of Bombycoidea moths and relatives from Darwin Tree of Life [https://www.darwintreeoflife.org/] (Supplemental Table 6). We assigned a reference protein sequence for each gene of interest from the *Bombyx mori* predicted proteome. We used Orth finder to identify orthologs and filtered orthogroups containing the reference sequence (49, 55). Since we had more than one gene per species, we reduced the number down to a single gene by choosing the highest identity sequence relative to the *B. mori* reference sequence. We also modelled the 3D structure of *B. mori* proteins and mapped the evolutionary conservation onto the 3D predicted structure for proteins above a certain conservation threshold (Supplementary Data 9). These analyses predict structurally and functional conserved regions of proteins (Fig S9, Supplementary Data 10). We repeated this analysis with 38 insect genomes (Supplemental Table 7) and mapped evolutionary conservation onto 3D protein structure (Supplementary Data 9,10). We include results of evolutionary conservation analyses for two regulatory candidates (*disco* and *tk*) that showed varying levels of sequence and protein evolution between insects and moths (Fig. S10).

Modeling predicts additional functional zinc finger domains for disco in Lepidoptera

Disco was recovered across multiple analyses. In *Drosophila melanogaster* it is known for its role in eye development, important for circadian maintenance, and for leg and antennal appendage

formation (56–60). To determine if *disco* was conserved between moths and *Drosophila*, we compared the primary sequence and 3D protein structure of *Drosophila* and *Bombyx mori*. In the moth, the sequence length of *disco* was nearly double that of *Drosophila* (Fig 6A). However, a region spanning over 100 amino acids was highly conserved, contained the zinc-finger domain important for its function, and showed strong 3D structural conservation (Whole protein alignment RMSD(align, super) = 37.833, 2.566, MatchAlign score (align, super) = 540 (2234 atoms), 333.1 (554 atoms) vs. alignment of conserved region RMSD=2.699,2.566, MatchAlignScore =443(671 atoms), 351 (615) atoms). This indicates the DNA binding function of *disco* has likely been conserved. However, an additional ~500 amino acid region absent in *Drosophila* is highly conserved across moths, and includes several regions predicted to be functional (Fig. 6A, Fig. S10). We hypothesize that *disco* has a novel role in moths for diel-niche regulation. To further test this, we compared *disco* sequences across *Anisota* and *Dryocampa* and found 23 mutations between them, three of which mapped to the predicted functional region (Fig. 6B, Supplementary Data 10). InterProScan predicted four zinc finger domains, three in this region, although the CATH-Gene3D databases prediction combined the two separate domains into a single predicted domain (Fig. S11). We also found many sites (53) with predicted phosphorylation potential, especially around the second zinc finger domain (18). Reducing the stringency increased the total to 142, which were still enriched around the second domain.

Discussion

We identified the genetic mechanisms of diel-niche switches by examining gene expression during the day and night in two closely related moth species. We used RNA-Seq to measure gene expression profiles of head tissue and isolated sensory, circadian and neural processes. We found between 300 and 700 genes that significantly altered day night circadian expression patterns. The diurnal species had enriched visual perception genes during the day and the nocturnal species had locomotor and olfactory genes upregulated at night.

Thirty overlapping DEGs were present in both species, with some DGEs showing divergent patterns of expression matching the species' diel-niche. Examining expression data with a sensitive clustering analysis yielded over 170 genes that showed day-night or species specific co-expressed clusters. We also used GO terms to search for genes associated with sensory, neural, or circadian patterns. We found genes in each category (Supplemental Table 4) and modeled the 3D structure and

predicted functional regions for a subset (Supplemental Table 5). The divergently expressed gene *disconnected* (*disco*) was implicated in vision, hearing, locomotion and brain development. Given the multiple lines of evidence supporting the importance of this gene in regulating diurnal and nocturnal activity in moths, we explored the evolution of *disco* by modeling fly and moth *disco* structures. This analysis revealed novel zinc-finger conserved domains in moths, which were lacking in *Drosophila,* surrounded by phosphorylated sites. Several mutations between *Anisota* and *Dryocampa* also mapped to these regions, further strengthening evidence for its role in diel-niche shifts.

Visual and olfaction

Visual systems often accompany diel and photic environment shifts (18). For example, nocturnal carpenter bees have much larger facets in their eyes than their diurnal counterparts (61). Nocturnal moths have higher photoreceptor light sensitivity and have neurons more suited to pooling than close diurnal relatives (62). Some of these trends seem to hold across the phylogeny, with diurnal species evolving more complex color visual systems, often reflected in visual opsin sequence evolution (22, 23). We did not find diel-expression patterns in color vision opsins, a result corroborated by a recent study (22). However, we discovered a cerebral opsin (ceropsin) upregulated during the day. Ceropsin has been implicated in photoperiodism and is expressed in the brain (63).

We also found several eye development genes (*TENM2, ANKRD17, EHD4, JAK2*), phototransduction genes (*PPAP2*, *RDH11*), and retina homeostasis, eye-antennal disc development, and photoreceptor cell maintenance genes (*disco, glass*) (64). Surprisingly a few visual genes, such as *garnet* and *rugose* also appeared to have different isoforms present, showing both day and night upregulation. *Garnet* is an eye color mutant gene in flies (65) and *rugose* is implicated in retinal pattern formation (66).

One of the classical differences between day flying butterflies and night flying moths is that butterflies have a clubbed antennae, while moths often have a sensilla-covered antenna; likely a function of increased investment in olfactory sensing, similarly, there is also evidence for increased investment in olfactory sensing in the brain, with relatively larger mushroom bodies in nocturnal species (67, 68). In our diel gene-expression dataset, we found several olfactory genes, including those involved in odorant binding (*Obp84a,Obp58b*), pheromone response (*tk*), mushroom body development (*DAAM2 and DST*) and antennal development (*disco*).

300

301 ***Hearing and mechanoreception***

302 Many moths have active hearing organs, and these have evolved repeatedly across Lepidoptera (69),
303 and their use varies depending on the ecological context of the moth. Several drivers for moth
304 hearing include sexual selection and as a defense against insectivorous bats (70). The primary
305 nocturnal group of butterflies, Hedylidae, has reverted to a nocturnal niche and regained hearing
306 organs (25). Thus, although rudimentary vibrational hearing exists in some diurnal moths (71) it is
307 significantly more developed and used at broader frequency ranges at night, when light is less
308 available. While Saturniidae lack hearing organs (72) they may instead still have vibration
309 mechanosensors that can be useful in evading predation similar to those in locusts and crickets (73–
310 76). Indeed, we did find diel co-expression of several mechanosensory and ear development genes,
311 including *PI4KB, KCTD15, unc-22,* and *RhoGAP92B* (77)

312

313 ***Brain and neural rewiring***

314 Finding an upregulation of neural and brain development was expected since we sequenced head
315 transcriptomes. However, we tested gene expression two days after pupal eclosion, where
316 presumably gross structures are already developed, so it is possible these genes we found expressed
317 regulate adult plasticity. Adult neural plasticity has been showed in Lepidoptera and other insects
318 (78–80), and we recovered several genes linked to axon regeneration (*APOD*), central complex
319 development (*Ten-a, TENM2, DST, ALDH3A2, OGT*), central nervous system development (*disco*,
320 *RpL4*), and neuropeptide hormonal activity (*tk*). Several genes were specific to retinal ganglion cell
321 axon guidance (*TENM2*) and mushroom body development (*DAAM2*), leading us to speculate that
322 plasticity could occur through neural wiring and plasticity are shaping sensory adaptation. Many
323 Lepidoptera have distinct phenotypic plasticity in wing patterns and coloration in different seasons
324 (81), and there is some evidence that they also have seasonal plasticity in their behavior and foraging
325 preferences (82). Research has found Lepidoptera can override their innate preferences and learning
326 preferences for new visual and olfactory cues after eclosion (83). It is possible the diel-niche and
327 circadian rhythms too have plasticity, there are reports of some moth species like *Hyles lineata*
328 showing relatively labile diel-niches possibly driven by temperature and resources (84–86). These

same mechanisms allowing flexibility within a species, might be involved in diel-niche evolution between species.

Circadian and behavioral regulators

The behavioral state of an animal to engage in any activity, e.g., feeding, flight or mating, is likely a function of circadian and behavioral regulators that respond to certain stimuli. We found differential expression of genes involved in locomotion (*unc-22, KCTD15,Tk*) and circadian or rhythmic behavior (*SREBF1/SREBF2, OGT, disco, JAK2*) in both species. We also found several key circadian regulators like *per*, *tim* although they were downregulated only in *Dryocampa*. C*lock-like* or *takeout* another gene under circadian control, (87) was expressed in both species, although without any significant up or down regulation. *Takeout* was moderately conserved in the moths we examined, but had diverged considerably among other insects, to the point where orthology searches with Orthofinder failed to recover orthologs among insects. Even among *D. melanogaster*, its closest homologs were *Jhbp5* and *takeout* which had 21-25% sequence identity (Supplementary Data 11). Surprising, , it's 3D structure was highly conserved (RMSD (align, super) = 2.556, 2.755, MatchAlign score (align, super) =199 , 642.937, Supplementary Data 11, Fig. S12), highlighting an interesting example of likely functional convergence despite primary sequence divergence.

Energy use and general cellular maintenance

In addition to the genes mentioned above, several key divergently expressed genes were involved in energy utilization. This finding makes biological sense, as once activity is initiated, energy mobilization and upregulation of basic cellular processes need to be maintained. Two of these genes were *SLC175A/MFS10* and *SLC2A6/Tret1-1*, which encode different trehalose sugar transporters. Trehalose is an important sugar present in insect hemolymph (88). Other genes found were ribosomal (*RpL4, mRpS5*), golgi (*GCC*), and mitochondrial maintenance (*PARL/rho7*) along with transcriptional regulation (*Spt5).*

Disco as putative master diel-niche regulator in adult moths

Many genes that we discovered likely play a critical role in maintaining sensory shifts, but in order to identify genes that might regulate upstream, we looked for genes that were: 1) expressed in both

359 species, 2) showed coincident expression patterns with respect to diel-niche, and 3) played a role in
360 sensory and neural functioning and circadian control. Only *disconnected* (*disco*) fit these criteria.
361 *Disco* is a key developmental and patterning gene, first discovered for its role in neural migration
362 from the disconnected optic lobe mutant phenotype (89). This gene is now considered an appendage
363 development gene, with a well characterized pupal role in antennal and leg distal patterning in
364 *Drosophila melanogaster (60)*. Mutants also have a disrupted circadian locomotor rhythm due to the
365 improper formation of neurons that express clock genes (90, 91). As a gene that is involved both in
366 optic lobe, antennal formation, and in neural wiring and circadian control, it is a very strong
367 candidate for driving behavioral diel-niche shifts and sensory adaptations (56, 57, 92). *Disco*
368 expression data from *B. mori* suggests strong adult head, antennal, and nucleus expression,
369 suggesting that it still acts as a transcription factor, and can regulate other genes (40).
370 *Disco* encodes the acid zinc finger transcription factor and is approximately 500 and 1000
371 amino acids in length in *Drosophila* and *B. mori,* respectively. Our modeling showed that a ~150
372 amino acid region is conserved structurally at the sequence level, and this region also contains zinc-
373 finger motifs associated with DNA binding (93). This indicates that *disco* likely has retained its DNA
374 binding and likely pupal and appendage patterning function in moths. We predicted the functional
375 regions of *disco* in *Bombyx* based on evolutionary conservation modelling, and found an additional
376 500 amino acids, absent in *Drosophila*, were predicted to be functional (conserved and exposed).
377 Domain and family level modeling predicted at least two additional zinc finger domains in this
378 region (Fig. S10,Fig. S11). We also found many phosphorylated sites surrounding the zinc finger
379 domains. Examining mutations between *Anisota* and *Dryocampa* revealed three mutations mapped to
380 these predicted functional regions (Fig. 6B).
381 Given *disco's* adult diel-specific expression in moths, additional zinc-finger DNA binding
382 domains, and the high number of phosphorylated sites, we propose it as a candidate transcriptional
383 regulator for diel-regulation in adult moths.
384
385 ***Study species and their evolution and variation in life histories***
386 *Anisota* and *Dryocampa* are closely related and ecologically similar saturniids of the subfamily
387 Ceratocampinae, thought to have diverged ~3.81 Mya (47). They largely occupy the same range and
388 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:

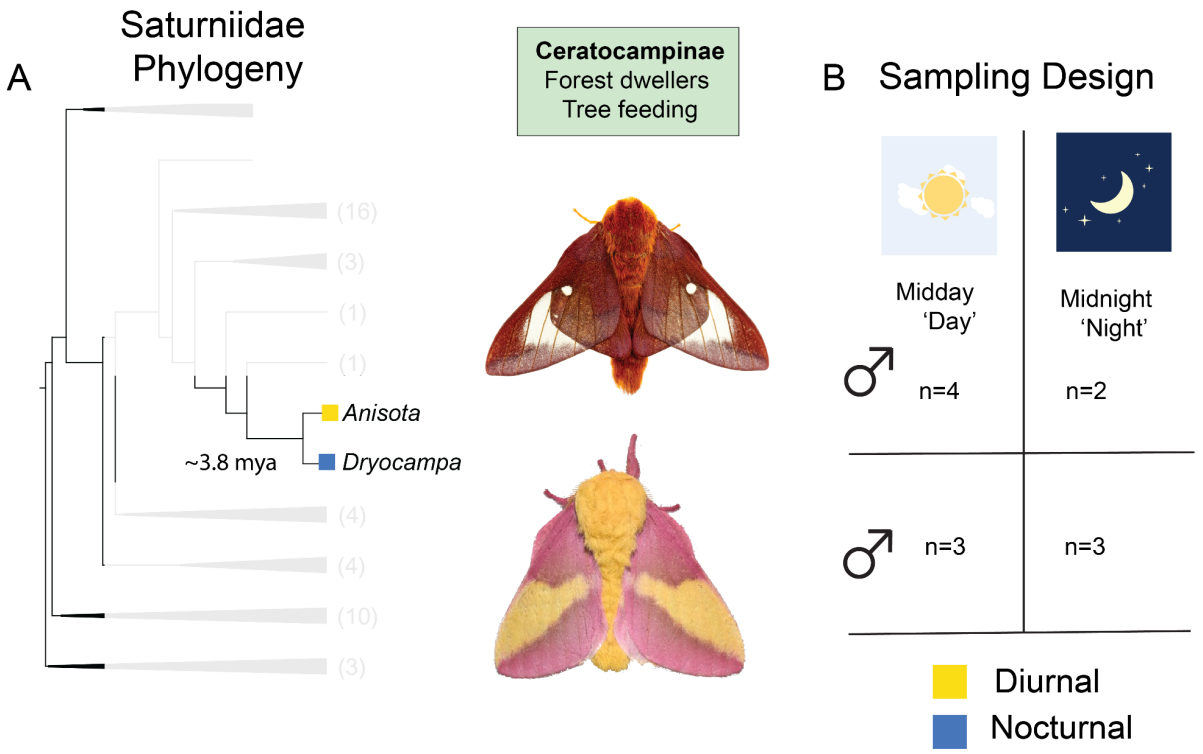


Figure 1: Nocturnal and diurnal moths on a phylogeny with RNA-seq sampling design.

(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;

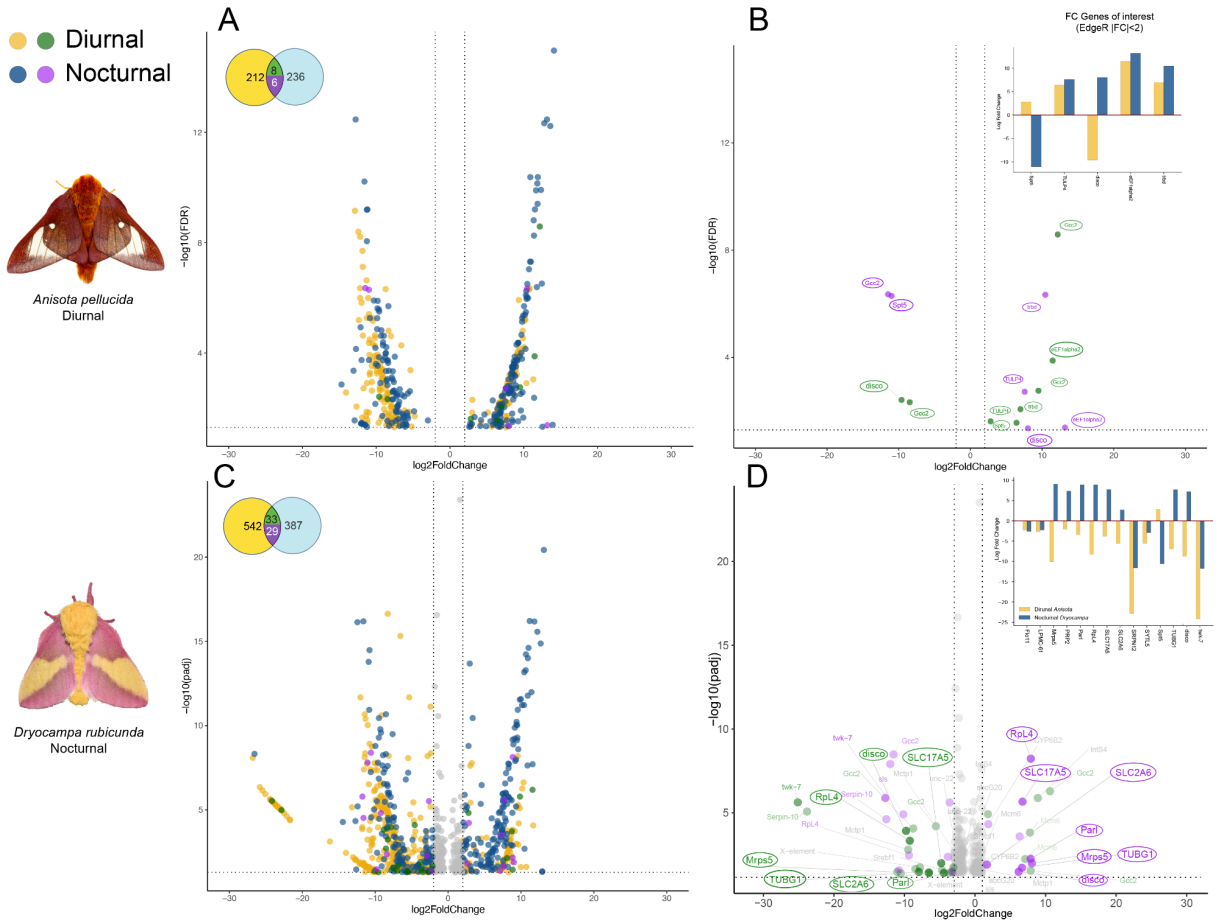


Figure 2: Nocturnal and diurnal species show divergent patterns of gene expression across two different analyses software (EdgeR and DESeq2)

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.*

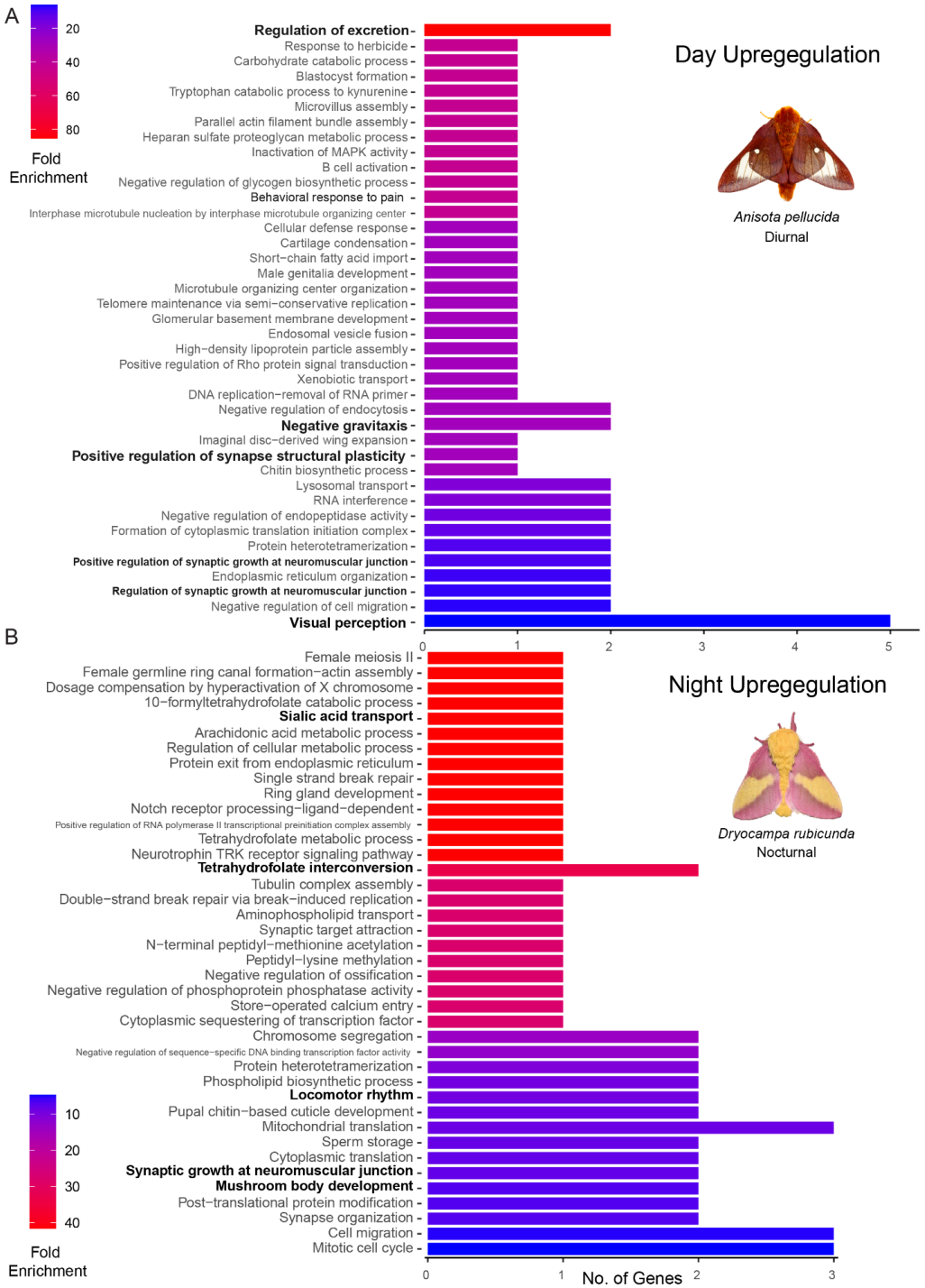


Figure 3: Visual genes are upregulated in the diurnal species during the day, and energy utilization, brain olfactory region and locomotion in the nocturnal species during the night.
Go enrichment of highly upregulated genes recovered from both DEG analyses coinciding with the species. highlighting the 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.

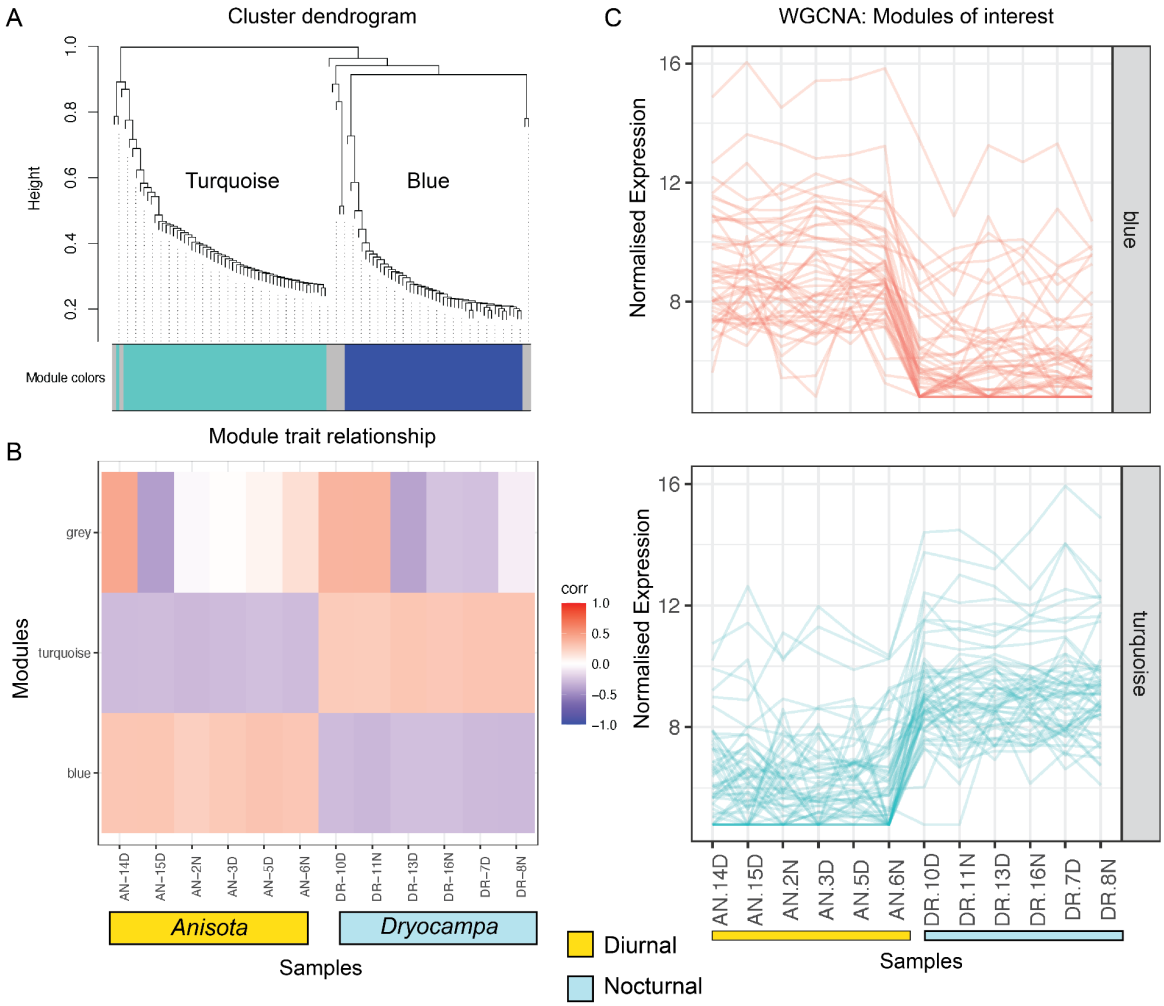


Figure 4: Modules of clustered co-expressed genes grouped using normalized expression highlighting the two modules that showed species specific clustering patterns.
A: Cluster dendrogram shows WGCNA clusters. B: Shows how patterns of gene expression correlate

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

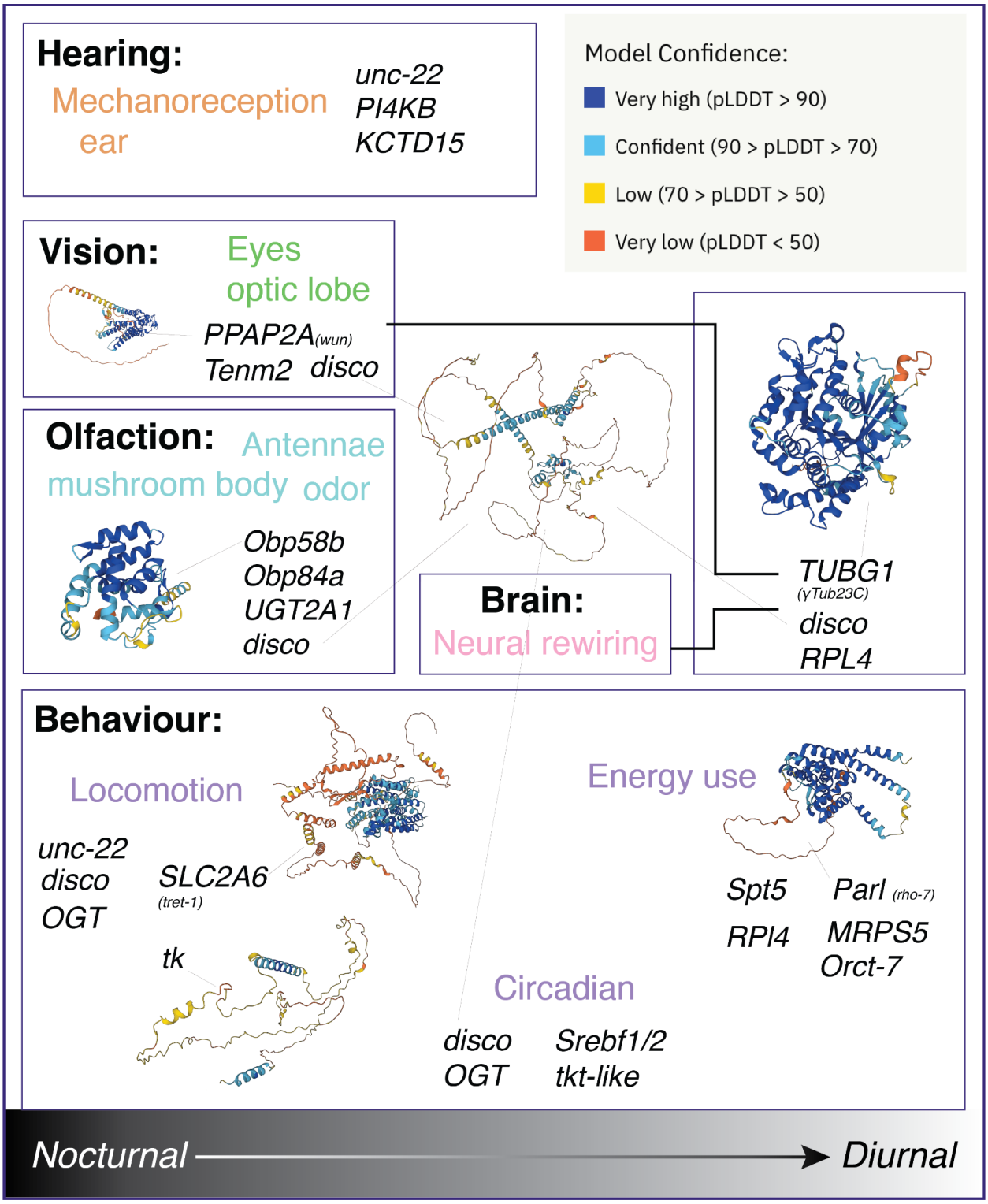


Figure 5: Model for diel-niche shift from and genes of interest obtained from GO searches and annotating divergently expressed genes. For representative purposes, homologs from *Drosophila*

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.

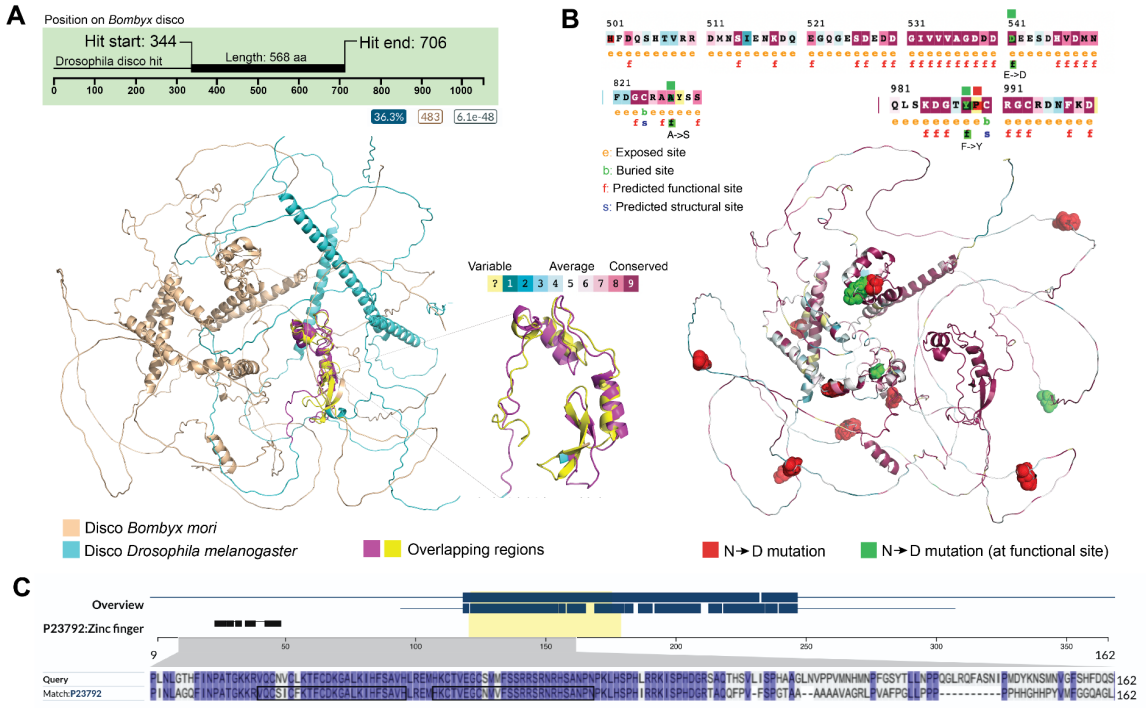


Figure 6: *Disco* has 150 amino acid long highly conserved region across *Bombyx* (1054 aa) & *Drosophila* (568 aa) involved in its role as a transcription factor, but it also has other roles in Lepidoptera due to the presence of other highly conserved and predicted functional regions, that are not present 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.

Supplemental Methods

Insect rearing:

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

Sampling design:

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

RNA extraction and quality control:

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

fluorometer and used Agilent tapestation or Bioanalyzer to characterize extraction quality, weight, and size distributions. Where possible, a concentration of more than 10 ng/µL was used for library preparation. Yields below 10 ng/µL were prone to higher loss when purifying with a DNAse agent, which often caused 20-30 % loss in RNA yield, approaching 5ng/µL is below the Qubit detection threshold (Table S1).

Library preparation and NGS sequencing:

Libraries were prepared by sequencing cores at NBAF-Liverpool using the NEB polyA selection and NEBNext Ultra II directional stranded kit, suitable for low input yields. 12 samples were run on one lane of an Illumina NovaSeq using SP chemistry (Paired-end, 2x150 bp sequencing). Samples were shipped overnight from MCGL to the NERC-NBAF after dehydrating in a biosafety chamber using GenTegra RNA tubes.

Read trimming and cleanup:

For *Anisota pellucida* and *Dryocampa rubicunda*, trimming was undertaken by the NERC-NBAF core and the raw Fastq files were trimmed for the presence of Illumina adapter sequences using Cutadapt version 1.2 (96). The option -O 3 was used, so the 3' end of any reads that match the adapter sequence for 3 bp or more were trimmed. Trimming was also done with trimmomatic. The reads are further trimmed using Sickle version 1.200 with a minimum window quality score of 20. Reads shorter than 15 bp after trimming were removed.

Transcriptome library sizes, QC, and PCA:

Quality Control (QC) was conducted on the trimmed reads, library size varied for each species (Fig. S1). We examined expression data and removed genes with TPM < 1. PCA results showed that *Anisota* and *Dryocampa* did not separate with diel-niche (Fig. S3).

***De novo* assemblies:**

We combined reads from multiple samples and generated several reference de novo transcriptomes using different assemblers, and then combining them, compared their quality using BUSCO scores and number of single-copy BUSCO genes across the different versions (see Table 2 for the BUSCO

scores, redundancy and duplication for the various assemblies). We found that unfiltered assemblies were highly redundant, but had the highest BUSCO scores. These assemblies were too overrepresented to be used in any downstream analysis so we used Transrate, CD-HIT, MMeqs2, as well as Transdecoder, to filter the assemblies. We chose v5 and v6 to use in downstream analysis, listed as the last two assemblies for each species (Table 3) and repeated certain analyses with both assemblies to see how stringent or a less stringent filter would affect the results.

Multi assembler *de novo* transcriptome assembly:

To our knowledge, there is no publicly available annotated reference genome for the two species used in this study at the time the analyses were performed; therefore, we chose to build a de novo transcriptome assembly with the pre-processed reads using the NCGAS pipeline, which combines multiple assemblers and then uses an evidence-based gene modeler to decide on the final set. For each species, we combined forward and reverse reads from all samples and normalized them using a normalization script in Trinity v2.12.0 (97). Normalized reads were used as input to generate assemblies using Velvet v1.2.10 (98), TransAbyss v2.0.1 (99) , SOAPdenovo v1.03 (100) with various kmer sizes. These assemblies were combined and filtered using Evigenes v4: 2020, March (101, 102), transdecoder and filtered with scripts provided in the NCGAS pipeline (103).

Transcriptome QC, annotation and filtering:

Further quality control and filtering was conducted using BUSCO (104), transrate (105), QUAST (106) (v5.02), CD-HIT(v4.6.8) (107, 108)and cascaded clustering with MMseqs2 (v12) (109). BUSCO scores were greater than 95%, but many genes had multiple versions from the different assemblers, and because this would bias the downstream analyses we attempted to reduce the redundancy at the cost of lowering BUSCO scores. Protein coding regions were predicted using TransDecoder (https://github.com/TransDecoder/TransDecoder) and the cds file from was run through Transrate (105), which further pulled contigs based on mapping rate and identified a “good” collection of transcripts, the resulting protein-coding sequences were filtered with additional CD-HIT (107, 108)and MMseqs2 (109) filtering. This resulted in lower BUSCO scores 75-85% but also much lower redundancy in the gene set of about 1-3 fold. We ran the WGCNA (52) and GO enrichment (110) analyses on the more stringent assemblies (BUSCO 75%, 2% genes with duplication) since

607 they are more sensitive to redundancy, but RNA-seq was run across the slightly lower stringency
608 analyses to be more inclusive (BUSCO~85%, 60-70% genes with duplications).
609
610 **Annotation of reference transcriptomes:**
611 Reference genomes were annotated with eggNOGmapper and Orthofinder (54, 55). We also tried
612 annotating the reference transcriptomes with Trinotate (111), which run against the swissprot and
613 pfam databases using hmmscan, blastx, blastp, signalP and tmmhmm, KEGG, and GO and eggNOG
614 (49, 110, 112–120). ORF predictions for orthofinder were obtained using *Antheraea pernyi* as a
615 model. Orthofinder was run with well-annotated Bombycidae and Saturniidae moths, *Bombyx mori*,
616 *Antheraea pernyi*, and *Antheraea yamamai*, to generate clusters. *Bombyx mori* was a useful reference
617 since the Orthofinder cluster for each species generated a list of mostly single-copy genes for which
618 more Lepidoptera annotations were available from SilkDBv3 (40). While this was not as complete as
619 trinotate, the annotations provided from this procedure were more accurate than Trinotate. Trinotate
620 annotation had many more human/ vertebrate hits than insect matches than eggNOGmapper and
621 Orthofinder transferred annotations, so while the files are provided as a reference, they were not used
622 in downstream analyses.
623
624 **Differential gene expression (DGE):**
625 Reads were mapped using Salmon (121) and differential gene expression analysis was conducted
626 using EdgeR (122) and DESeq2 (123). EdgeR(v 3.38.1) was used to normalize and test for
627 significantly expressed genes. DESeq2(v1.36.0) was also used to normalize and to generate PCAs
628 and other summary statistics. We used 'Day' as the treatment and 'Night' as the control for all
629 groups with EdgeR, although the order of fold change switched for DESeq2. The p-values obtained
630 were adjusted to account for multiple hypothesis testing with FDR for EdgeR and adjusted p-value
631 from DESeq2. Genes with FDR or adjusted p-values <0.05 and $|FC| > 2$ were used as a criteria to
632 identify significantly expressed genes.
633
634 **Gene enrichment analysis:**
635 We performed gene enrichment analysis using GO terms with the tools TopGO v2.48.0 (124) and
636 ShinyGo v0.75c (125) (http://bioinformatics.sdstate.edu/goc/). This analysis involved comparing the

selected genes of interest to a gene universe with GO term annotations to determine if there were overrepresented or underrepresented GO terms. To obtain genes with GO annotations, we utilized similarity clusters to map annotations from *Bombyx mori*.

In TopGO, we employed a more stringent approach by using filtered transcriptomes to mitigate bias from gene duplications. We selected genes of interest based on a False Discovery Rate (FDR) threshold of less than 0.05. We ran two different enrichment algorithms, 'classic' and 'elim', and used two test statistics, Fisher's exact test and a Kolmogorov-Smirnov-like test. The tests were performed for both 'Biological Process' (BP) and 'Molecular Function' (MF) categories. While we generated tables of significantly enriched GO terms, obtaining individual genes within each group was limited due to the constraints of TopGO with custom annotations.

For ShinyGO, we utilized the custom v0.75c which included the updated *Bombyx mori* genome stringent. For all analysis, we used the following settings (pathway size: min.=1, max=2000). The settings for number of pathways to show and FDR cut-offs were chosen to represent the entire list of top 100 genes in the final datasets, although in some cases a fewer number are represented in the visualizations. We focused on the 'Biological Process' (BP) category for the various tests. Tables of GO terms, gene information, and graphs summarizing the significantly enriched GO terms were generated. We repeated several different analyses **1)** A less stringent analysis was run where genes with FC<|2| and padj. <0.05. This was useful to visualize the genes up and downregulated functionally and the background used was all the genes recovered from the DEA analysis that had orthologs. An FDR cut-off <0.1-0.3. For the background gene set, we used the species' de-novo transcriptome matching *Bombyx* ortholog set.

2) A more stringent criteria was used to examine highly upregulated genes a FC<|5| and padj. <0.05 cutoff was used, and filtering only genes that occurred in both EdgeR and DESeq2 analyses. We selected a single representative when multiple Bombyx orthologs were recovered, using the ortholog with the most complete annotation. An FDR cut-off <0.17 was used. For the background gene set, we used the species' de-novo transcriptome matching Bombyx ortholog set.

3) For WGCNA clusters tan and grey60, we used their respective species background expressed transcript set with a FDR cutoff <= 0.1. For blue and turquoise, the background of all orthologs found in *Bombyx mori* was used. An FDR cutoff<1 was used, since smaller values provided insufficient genes given the larger number of genes being tested.

667

668 **Gene network analysis:**

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

674

675 **Gene annotation and filtering:**

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

684

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

686 We mined genes of interest from ~bombycoidea moths and closely related families that had well
687 annotated genomes on Ensembl and NCBI (Supp Table 6) and across model insects (Supp. Table 7).
688 These genomes were chosen as all Bombycoidea from Ensembl with well annotated assemblies and
689 five species Noctuidae and Geometridae each were included as close relatives. We added *Bombyx*
690 *mori* as a reference and *Antheraea pernyi* to represent Saturniidae, and their peptide files were taken
691 from their respective source papers (40–42). For insects, we chose long-read well annotated genomes
692 from Ensembl (Ensembl Gene build) and five model systems from insect base (127). We ran
693 Orthofinder (v2.5.2) (49, 55) with 'orthofinder -f bom-rel/ -S diamond -M msa -A mafft -T fasttree -a
694 1 -X -z -t 16' to recover orthologs. Each orthogroup often contained multiple sequences per species.
695 We chose the sequence with the highest identity to the reference *Bombyx mori* sequence using
696 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).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

763 23. Y. Sondhi, E. Elis, S. Bybee, J. Theobald, A. Kawahara, Light environment drives evolution of color
764 vision genes in butterflies and moths (2021) https:/doi.org/10.5061/dryad.gmsbcc2kr (May 13,
765 2022).

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

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

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

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

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

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

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

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

784 32. S. Iwai, Y. Fukui, Y. Fujiwara, M. Takeda, Structure and expressions of two circadian clock genes,
785 period and timeless in the commercial silkmoth, Bombyx mori. *J. Insect Physiol.* **52**, 625–637
786 (2006).

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

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

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

793 36. J. W. Truman, L. M. Riddiford, Neuroendocrine control of ecdysis in silkmoths. *Science* **167**, 1624–
794 1626 (1970).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(2010).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

131. H. Ashkenazy, *et al.*, ConSurf 2016: an improved methodology to estimate and visualize 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).

Supplemental Figures

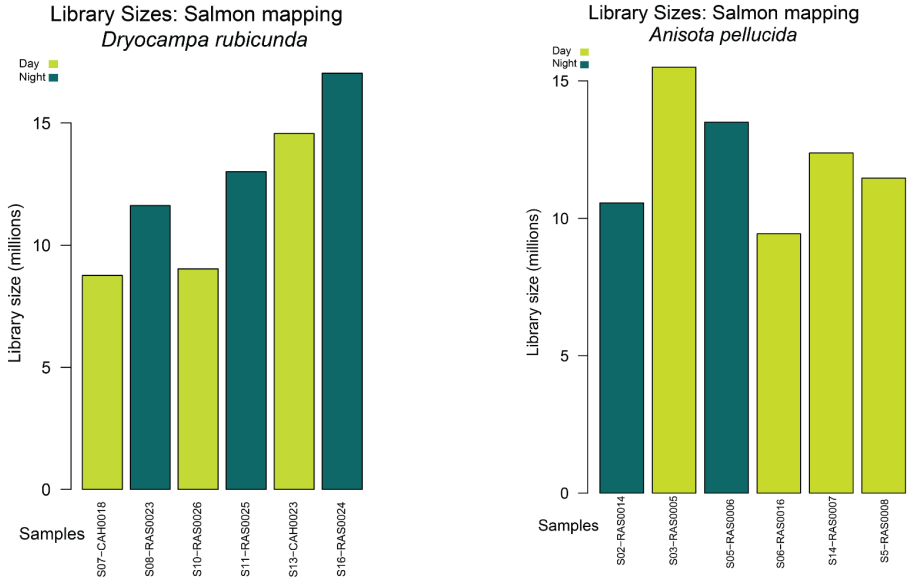


Figure S1: Variation in salmon mapped libraries for *Dryocampa rubicunda* and *Anisota pellucida*

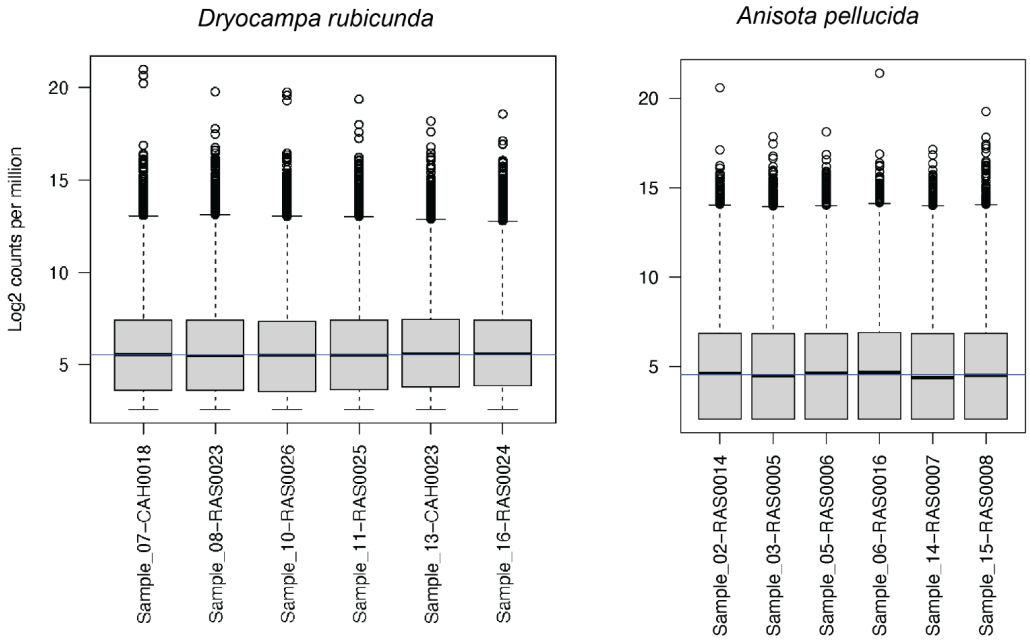
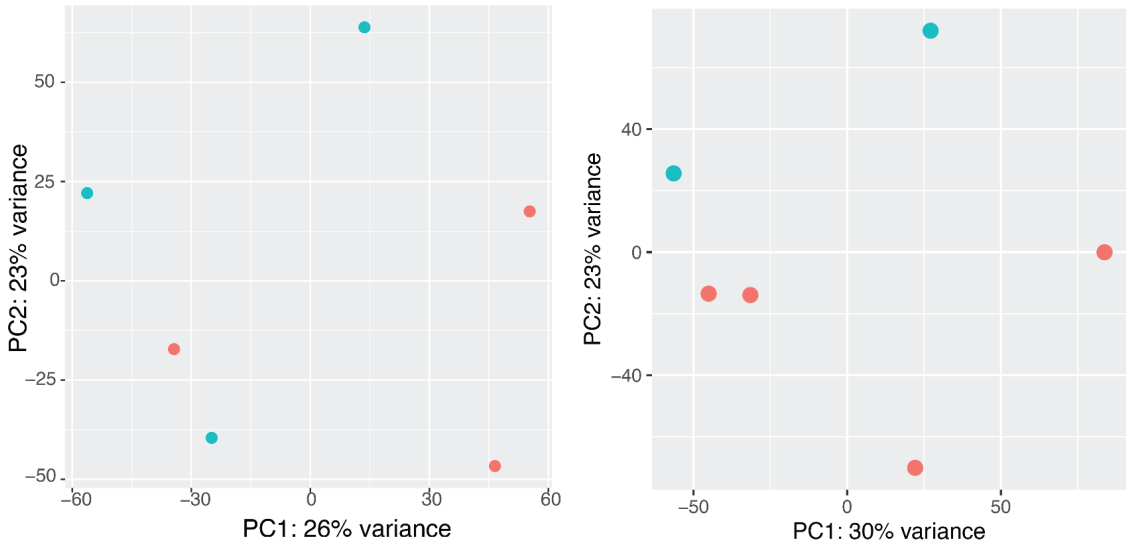


Figure S2: Variation in normalised libraries for *Dryocampa rubicunda* and *Anisota pellucida*

1039

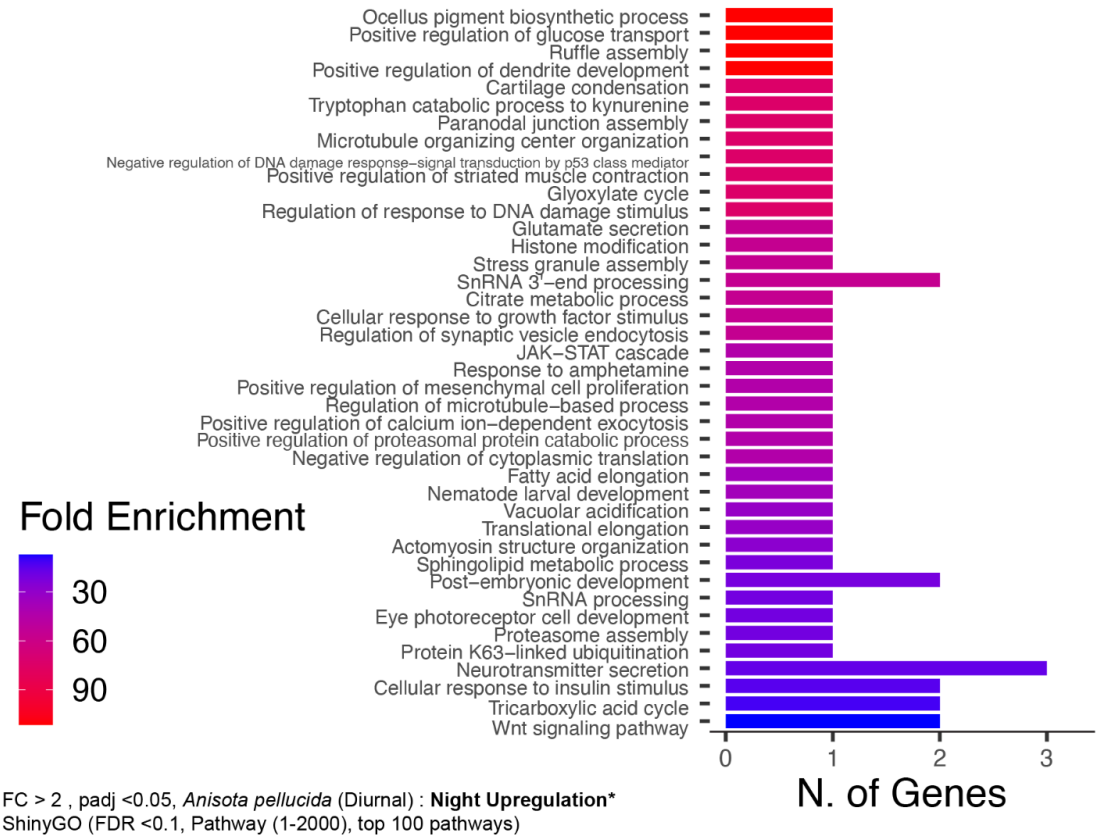


Dryocampa rubicunda *Anisota pellucida*

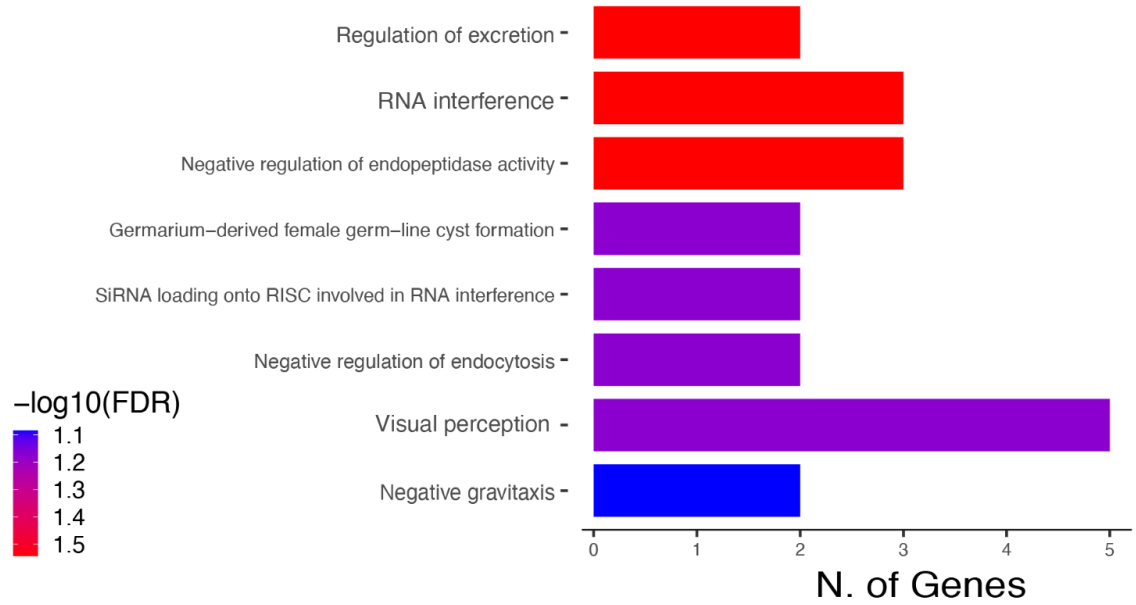
1040

1041 **Figure S3**: PC1 and PC2 of libraries labelled for *Dryocampa rubicunda* and *Anisota pellucida*.
1042 Orange and blue labels refer to midday 'Day' and midnight 'Night' collection time points,
1043 respectively. This PCA was generated from mapped salmon reads after normalising with DESeq2.

1044



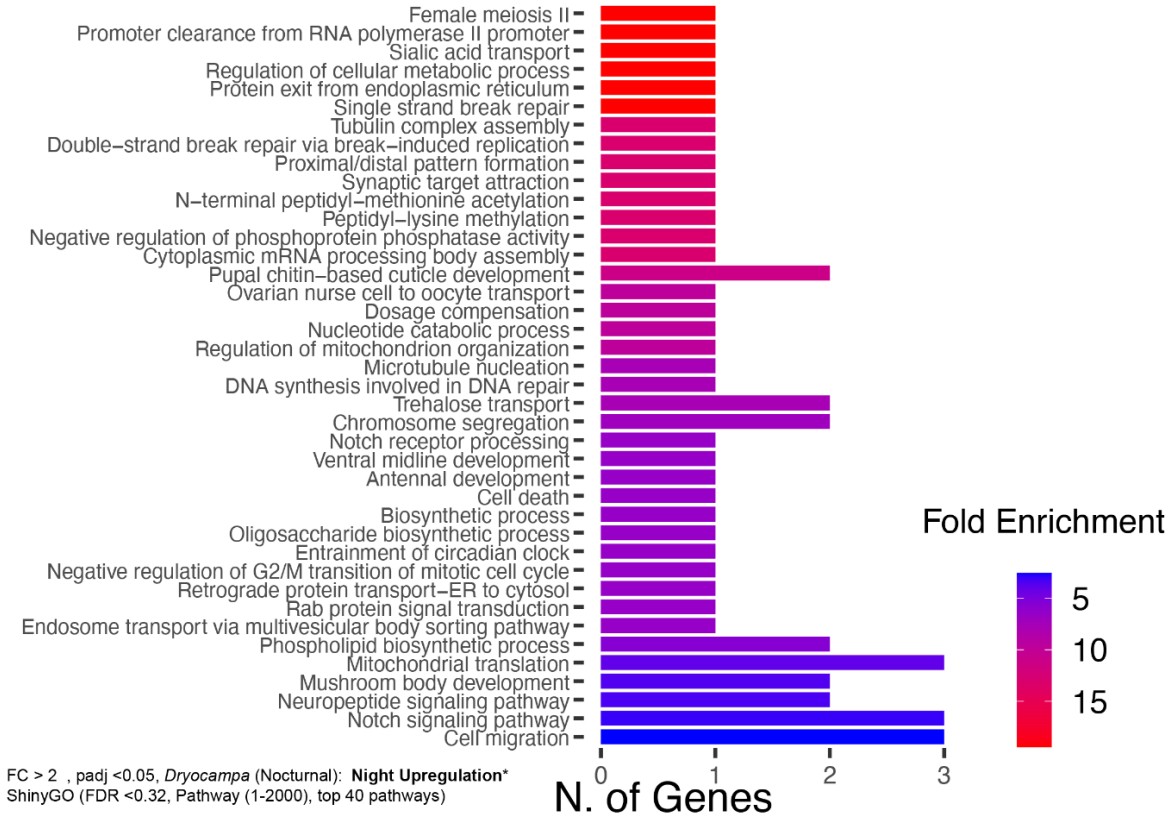
FC > 2 , padj <0.05, *Anisota pellucida* (Diurnal) : **Night Upregulation***
ShinyGO (FDR <0.1, Pathway (1-2000), top 100 pathways)



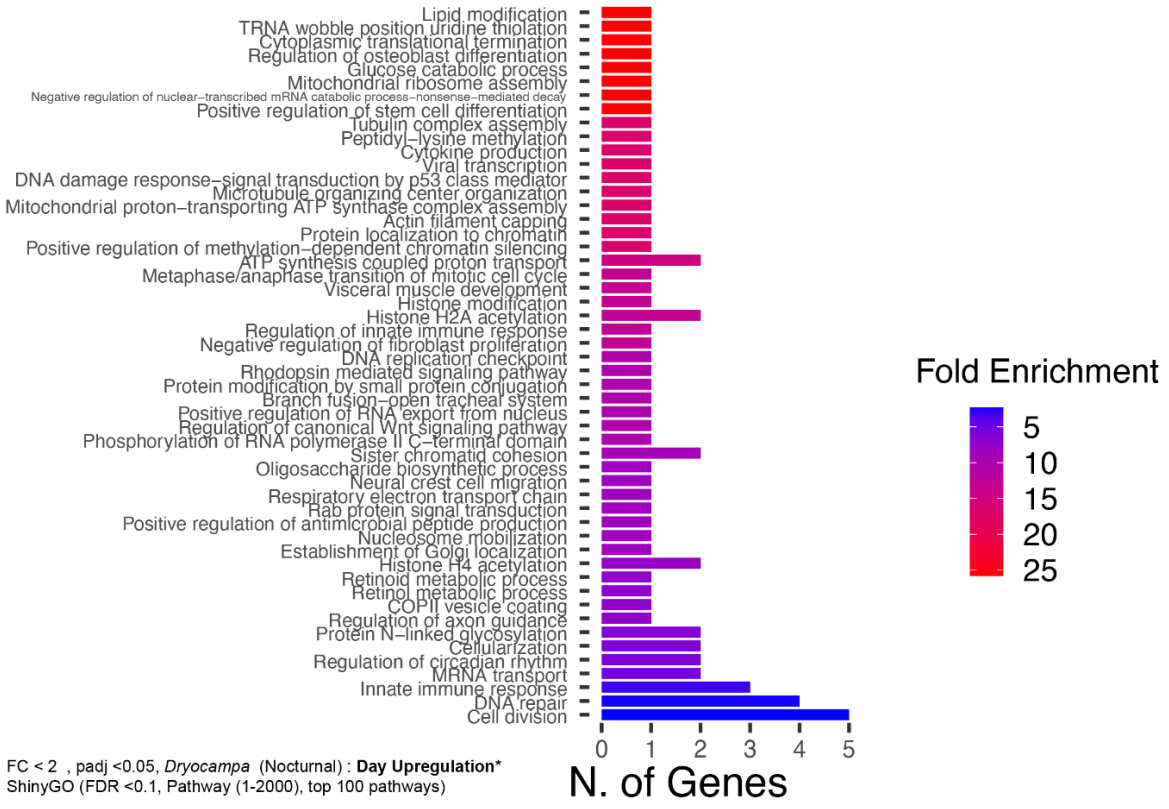
FC < -2 , padj <0.05, *Anisota pellucida* (Diurnal) : **Day Upregulation***
ShinyGO (FDR <0.1, Pathway (1-2000), top 20 pathways)

1045

Figure S4: Enriched lists of GO terms associated with significant positively and negatively expressed genes recovered from 2 or more analyses (DESeq2, WGCNA, EdgeR) for *Anisota pellucida.* ShinyGo was used to highlight terms with genes showing representation compared to the background of detectable genes values representing ShinyGo overrepresentation. Top: Enriched Day Upregulated genes (FDR>0.1) with gene expression FC>=2, adjusted p-value <0.05. Bottom: Enriched Night Upregulated genes (FDR>0.1) with gene expression FC<=2, adjusted p-value <0.05 and that showed at least 2 occurrences in the combined dataset. Values are sorted by -log FDR *Genes that showed up in at least 2 separate analyses are included, (>=2 occurrences in the combined dataset). The background used was all transcripts that were recovered in the DEG analysis and had *Bombyx* orthologs



FC > 2 , padj <0.05, *Dryocampa* (Nocturnal): **Night Upregulation***
ShinyGO (FDR <0.32, Pathway (1-2000), top 40 pathways)



FC < 2 , padj <0.05, *Dryocampa* (Nocturnal) : **Day Upregulation***
ShinyGO (FDR <0.1, Pathway (1-2000), top 100 pathways)

Figure S5 : Enriched lists of GO terms associated with significant positively and negatively expressed genes recovered from 2 or more analyses (DESeq2,WGCNA, EdgeR) for *Dryocampa rubicunda* . ShinyGo was used to highlight terms with genes showing representation compared to the background of detectable genes values representing ShinyGo overrepresentation. Top: Enriched Day Top 40 upregulated genes (FDR>0.32) with gene expression FC>=2, adjusted p-value <0.05, 77 terms enriched , but not all are displayed, see supplementary material for entire list. Bottom: Enriched Night Upregulated genes (FDR>0.3) with gene expression FC<=2, adjusted p-value <0.05 and that showed at least 2 occurrences in the combined dataset *****Genes that showed up in at least 2 separate analyses are included, (>=2 occurrences in the combined dataset). The background used was all transcripts that were recovered in the DEG analysis and had *Bombyx* orthologs

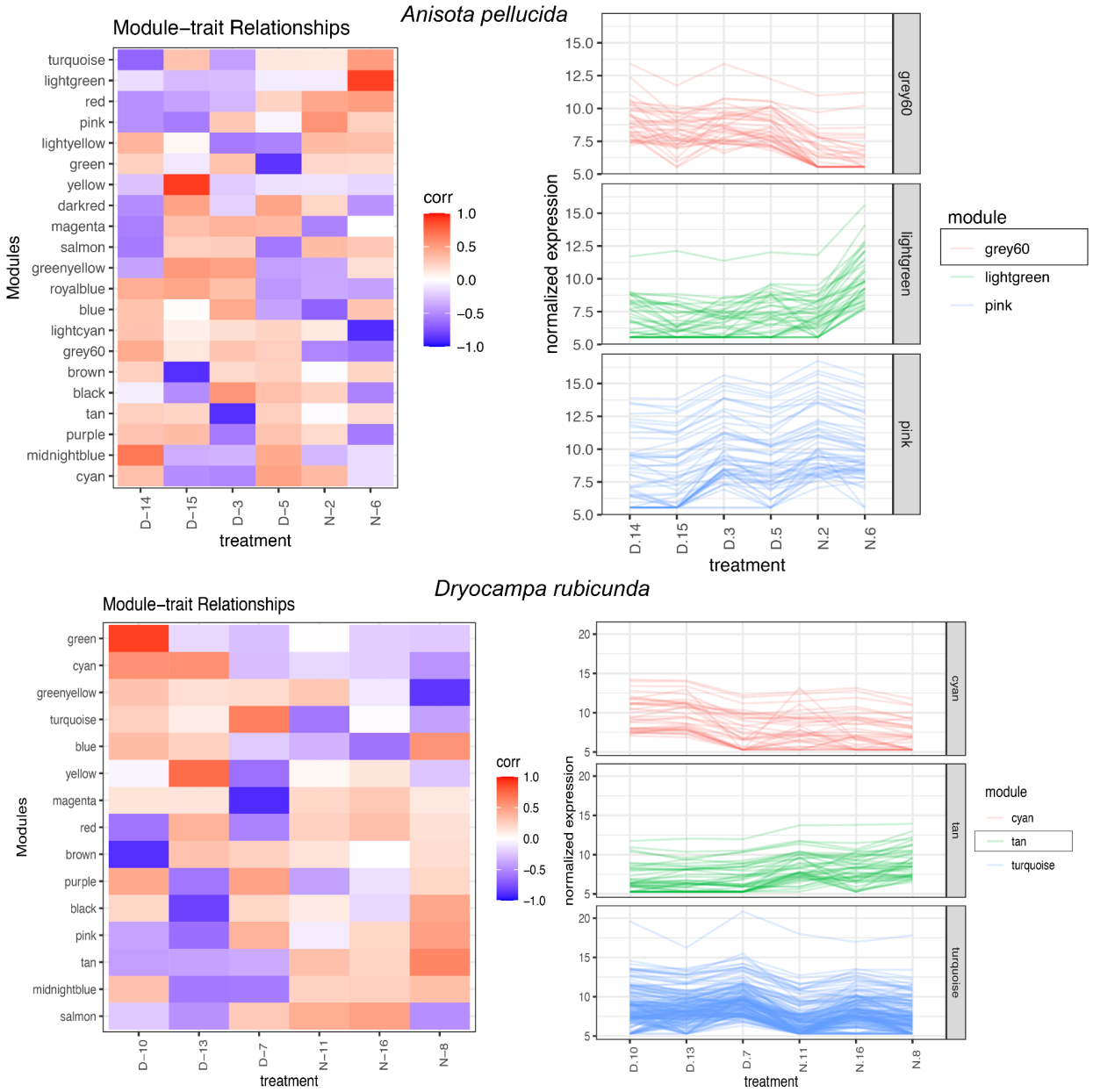


Figure S6: Modules of clustered genes grouped using normalized expression for *Anisota pellucida* (top) and *Dryocampa rubicunda* (bottom) Grey60 (38 genes) and Tan (32 genes) modules show species- time of day linked patterns. Left: Shows how patterns of gene expression correlate across samples for modules Right: Shows the normalized expression for all. Normalization was done with DESeq2 and reads were mapped to the more stringently filtered transcriptome. A soft power analysis was done and the picked power=9 for the WGCNA analysis.

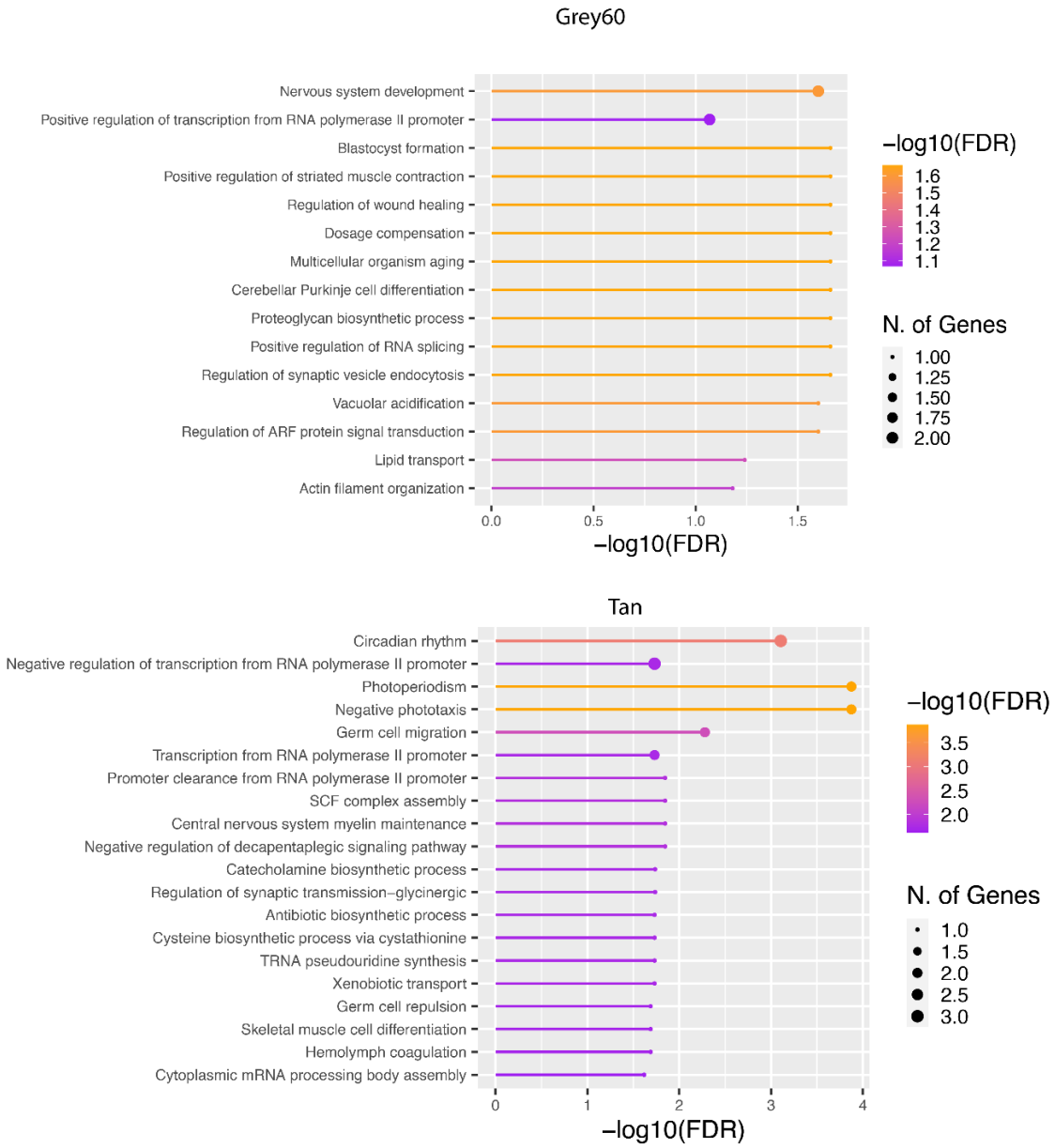


Figure S7: Shiny Go enrichment for Modules of clustered genes grouped using normalized expression for *Anisota pellucida* (top) and *Dryocampa rubicunda* (bottom) Grey60 and Tan. Background used was all *Bombyx mori* genes and FDR cut-off =0.1 min pathway =1 and GO Biological processes are displayed.

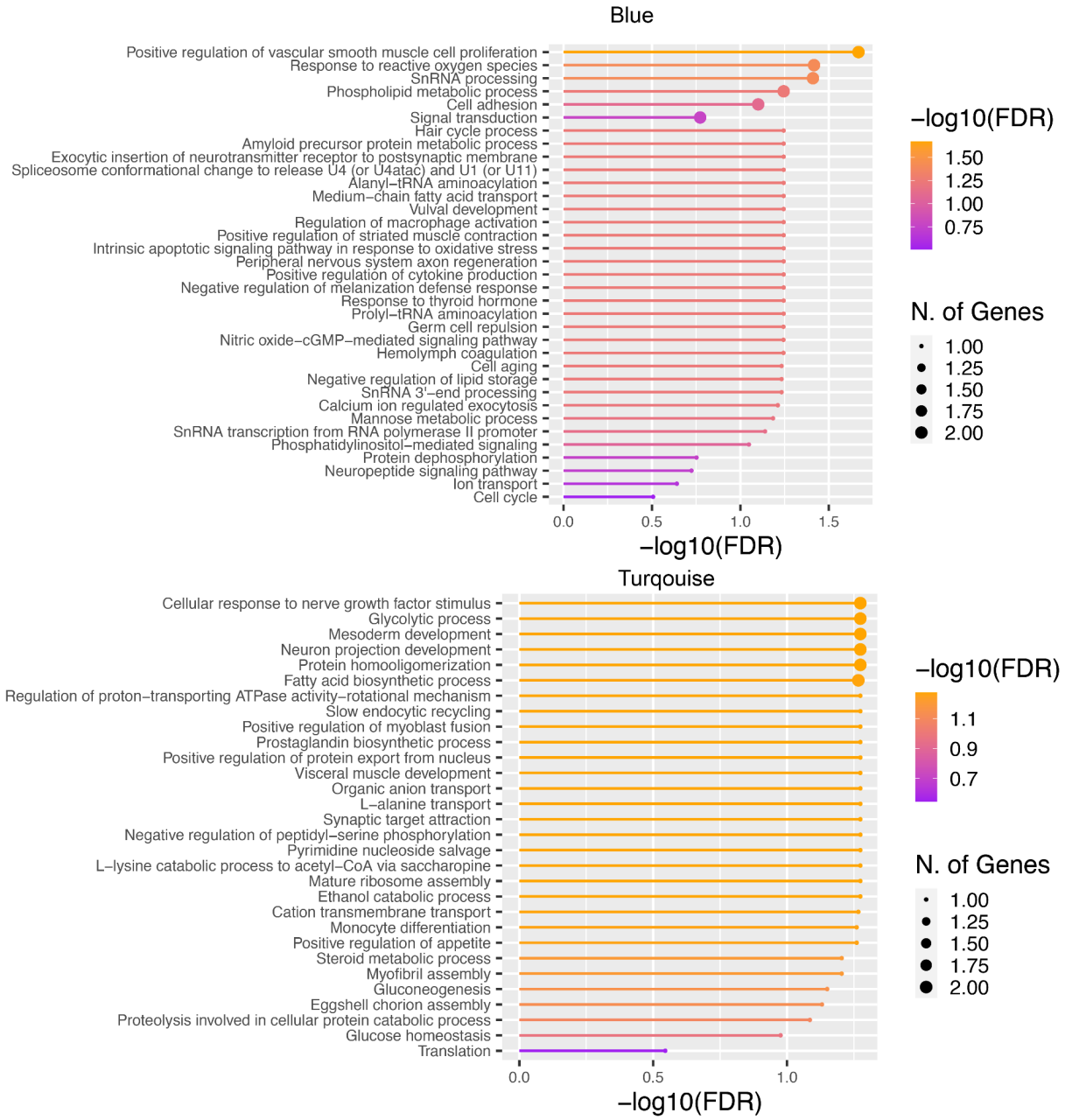


Figure S8: Shiny Go enrichment for Modules of clustered genes grouped using normalized expression for combined overlapping reads for the two species (Top) Blue and (Bottom) Turquoise. Background used was all *Bombyx mori* genes and FDR cut-off =1 min pathway =1 and GO Biological processes are displayed.

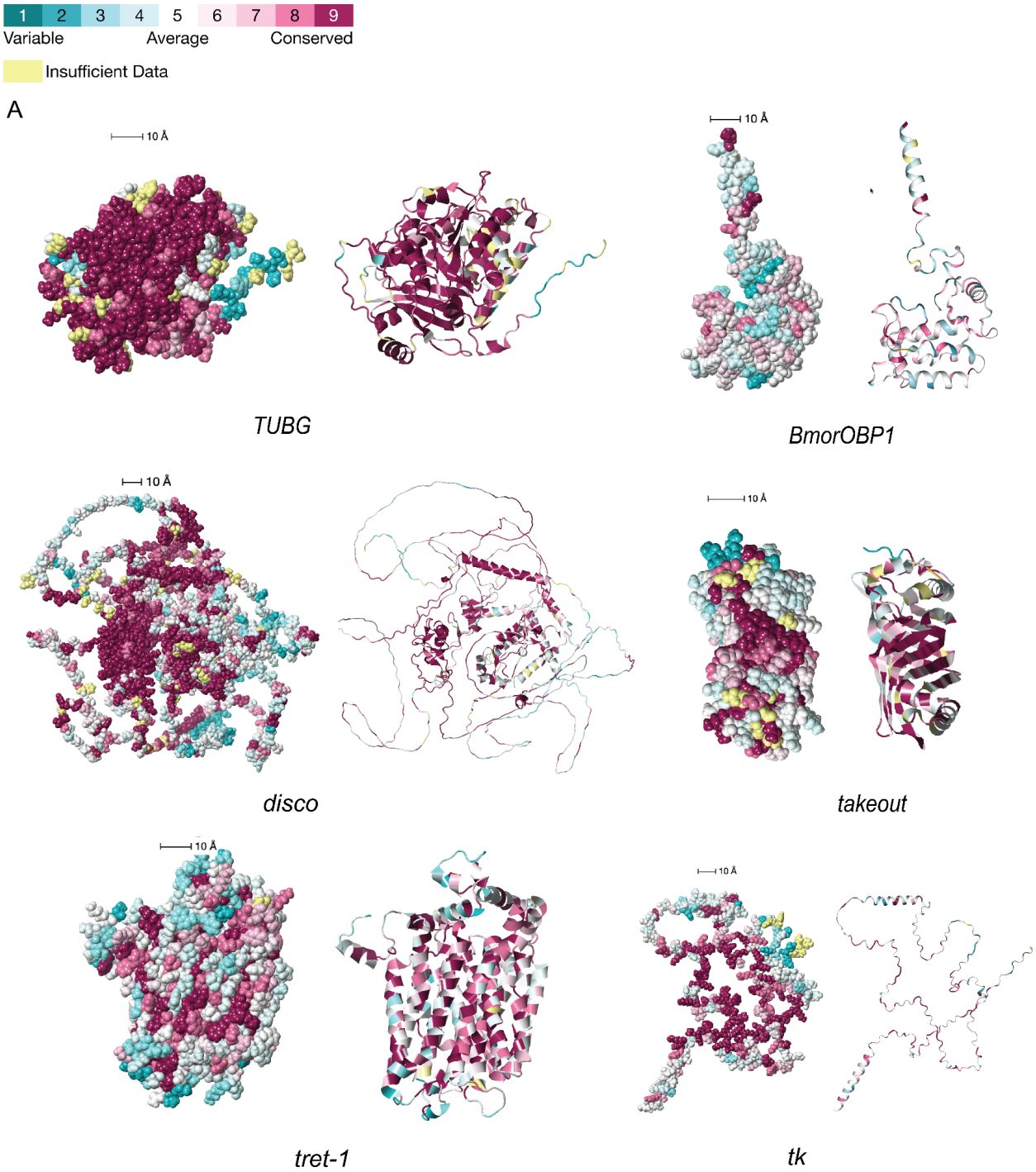


Figure S9: Variation in structural conservation across moths for predicted proteins for genes interest. Structures were obtained using AlphaFold predictions of *Bombyx mori* homologs. Left: Space filling models, Right cartoon models, colors represent conservation across the alignment from or Bombycoidea and related moth species. Conservation scores were calculated and mapped onto the

alpha fold structures using the Consurf (131) web server (For other proteins, see Supplementary data)

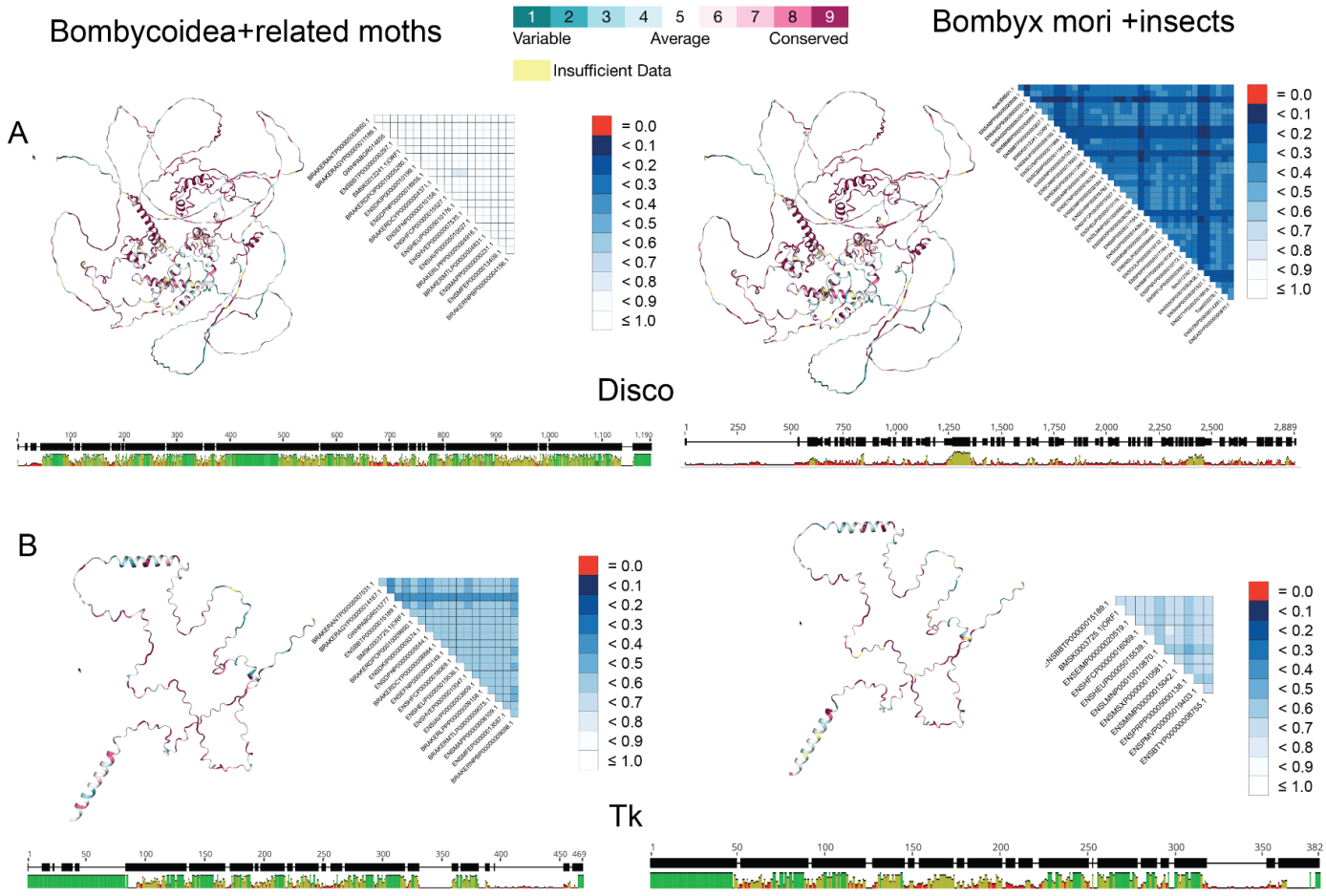


Figure S10: Illustration of conservation across the alignments and structure for **(A)** *disco* and **(B)** *tk* across Bombycoidea and relatives (Left) and insects (Right). Each inset contains the *Consurf* mapped pdb structures on the left. The *alistat* pairwise sequence conservation in the alignment scores on the right and the consensus of the protein alignments were visualized using *Geneious.* For conservation scores for other genes and lists of the species used for this modelling, see Supplementary Tales (Supp. Table: 5-7). Supplementary Dataset 10 has the associated models and output from the analyses.

Figure S11: InterPro and NetPhos predictions for *B. mori disco* protein. **(A)**InterPro predicted several zinc finger domains, disordered regions and disco and disco-related families. The numbers in black highlight the regions for the predicted domain and the numbers in red list the NetPhos predicted phosphorylated sites (threshold >0.9). **(B)** An overall distribution of the predicted phosphorylated sites with different thresholds (Left, Threshold=0.9, Right Threshold= 0.5)

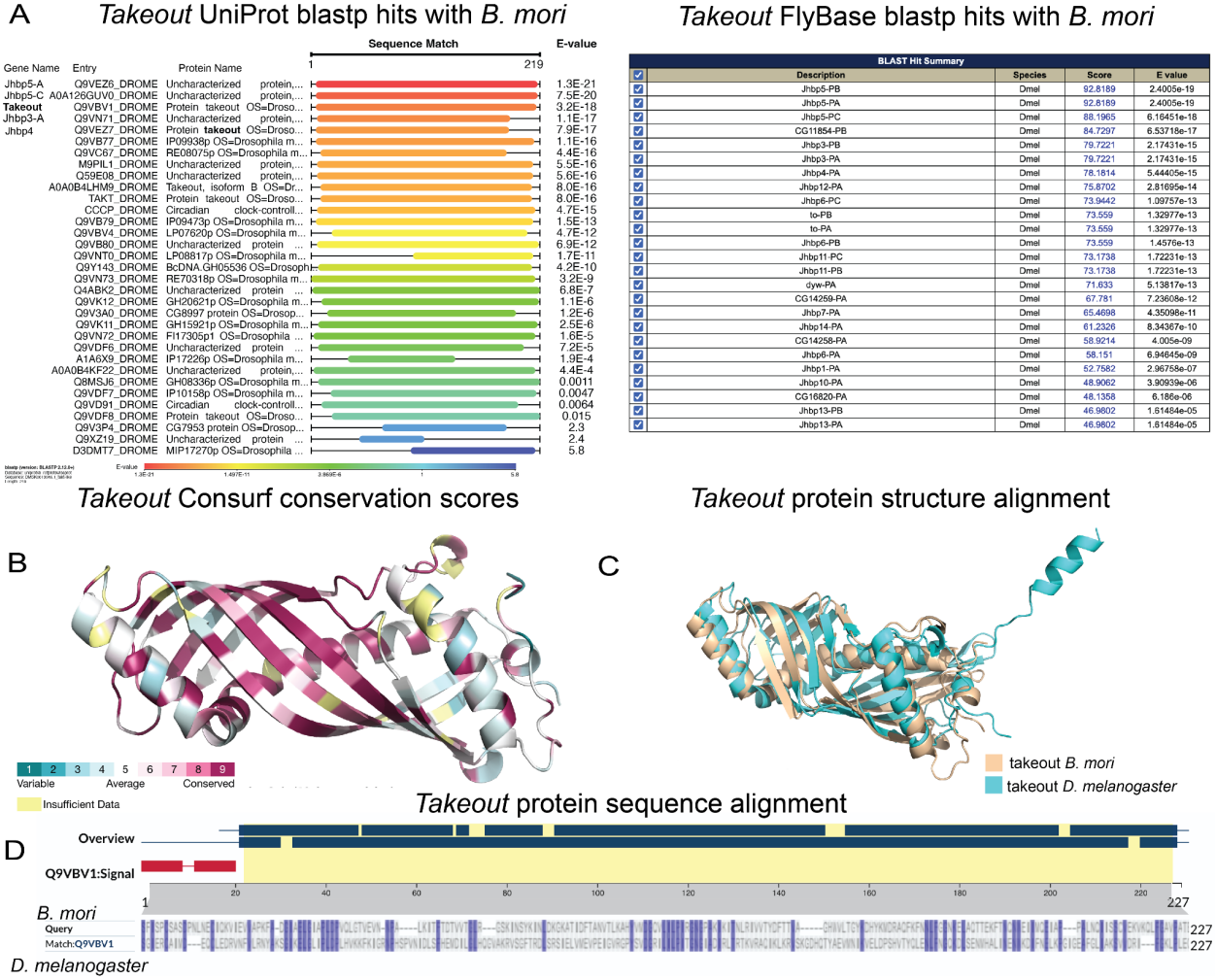


Figure S12: Lack of sequence level conservation in takeout, but high structural overlap **(A)** Right: Takeout UniProt blastp hits. (BLOSUM62, filtered for *D. melanogaster*), alignments =250, e-threshold= 10, database = uniprotkb_refprot+swissprot, list of illustrative hits is shown. Left: Similar blastp results from flybase limiting the species to *D. melanogaster.* (**B)** Consurf predicted scores for closely related moths mapped on to the *B, mori* takeout predicted protein structure. **(C)** 3D structural alignments for fly (*D. melanogaster)* and moth (*B. mori)* takeout 3D models. **(D)** Uniprot protein sequence alignment for fly and moth takeout proteins.

Tables

Species	Diel-niche	Collection Time	Replicates (n)
Anisota pellucida	Diurnal	Day	4
		Night	2
Dryocampa rubicunda	Nocturnal	Day	3
		Night	3

Table 1: Sampling design showing species, collection time and number of replicates in each treatment. All specimens were male.

Species	Assembler	kmer-size	Max- BUSCO groups	Lep BUSCO Score(%)	max-N50
Ani_pel	Velvet	35-85	4703 (k55)	88	3918 (k55)
Ani_pel	Soap	35-85	4682 (k35)	88.5	1661 (k35)
Ani_pel	Transabyss	35-85	4216 (k35)	79	1332 (k35)
Ani_pel	Trinity	25	4948	93	1608
Dry_rub	Velvet	45-85	4768 (k55)	90	3957 (k55)
Dry_rub	Soap	35-85	4753 (k35)	89	1915 (k75)
Dry_rub	Transabyss	55-85	3159 (k55)	59	1058 (k55)
Dry_rub	Trinity	25	4939	93	1595

Table 2: Assembly variation using different assemblers and kmer settings.

Key: Dry_rub: *Dryocampa rubicunda*, Ani_pel: *Anisota pellucida.* All models included day and night samples.

Species	Method	Total DEGs	DEGs (Up)	DEGs (Down)	Annotated DEGs	Unannotated %
A. pellucida D	EdgeR	350	154	196	211	39.71
D, rubicunda N	EdgeR	393	211	182	236	39.95
A. pellucida D	DESeq2	697	141	556	420	39.7
D. rubicunda N	DESeq2	498	253	245	300	39.7

Table 3: DEG recovery and gene annotation for EdgeR and DESeq2 analyses along with percent annotation using *Bombyx.* Up=FC>2, Implies night expression is greater than day expression. Down=FC<2, implies day expression is greater (control=Day, Night = Treatment)

Species	Assembly annotation	BUSCO score	BUSCO duplication	No. of contigs*	Redundancy (/15k)
Dry_rub	Combined	94.3	90.8	281790	18.8x
Dry_rub	Combined_cds_cdhit_95	90.1	80.2	184808	12.3x
Dry_rub	Combined_cds_filtered (v5)	85.2	59.6	41712	2.8x
Dry_rub	Combined_cds_stringent_filter (v6)	75.4	4.9	19877	1.3x
Ani_pel	Combined	94.3	90.8	281790	18.78
Ani_pel	Combined_cds	93.9	90.3	233097	13.3x
Ani_pel	Combined_cds_filtered_cds (v5)	85.2	59.2	41712	2.8x
Ani_pel	Combined_cds_stringent_filter (v6)	79.3	4.7	19087	1.3x

1152

1153 **Table 4:** BUSCO and contig distribution statistics for different species' *de-novo* assemblies with
1154 and without filtering. Increased filtering reduces redundancy and duplication but comes at the
1155 cost of reducing BUSCO scores. *All statistics are based on contigs of size >= 500 bp.

1156

nGenes	Fold Enrichment	Pathway	GO_ID	Gene Names
2	85.53	**Regulation of excretion**	GO:0044062	NPHS1,SLC9A3R1
1	42.76	Response to herbicide	GO:0009635	Alad
1	42.76	Carbohydrate catabolic process	GO:0016052	BMSK0015853
1	42.76	Blastocyst formation	GO:0001825	Actl6a
1	42.76	Tryptophan catabolic process to kynurenine	GO:0019441	KFase
1	42.76	Microvillus assembly	GO:0030033	SLC9A3R1
1	42.76	Parallel actin filament bundle assembly	GO:0030046	jv
1	42.76	Heparan sulfate proteoglycan metabolic process	GO:0030201	botv
1	42.76	Inactivation of MAPK activity	GO:0000188	Spred1
1	42.76	B cell activation	GO:0042113	AKAP17A
1	42.76	Negative regulation of glycogen biosynthetic process	GO:0045719	Gfpt1
1	42.76	Behavioral response to pain	GO:0048266	pain
1	42.76	Interphase microtubule nucleation by interphase microtubule organizing center	GO:0051415	Grip91
1	28.51	Cellular defense response	GO:0006968	PNLIPRP2
1	28.51	Cartilage condensation	GO:0001502	COL11A1
1	28.51	Short-chain fatty acid import	GO:0015913	SLC5A8
1	28.51	Male genitalia development	GO:0030539	ken
1	28.51	Microtubule organizing center organization	GO:0031023	Gcc2
1	28.51	Telomere maintenance via semi-conservative replication	GO:0032201	RFC3
1	28.51	Glomerular basement membrane development	GO:0032836	NPHS1
1	28.51	Endosomal vesicle fusion	GO:0034058	Ankfy1
1	28.51	High-density lipoprotein particle assembly	GO:0034380	ABCA7

1	28.51	Positive regulation of Rho protein signal transduction	GO:0035025	MCF2L
1	28.51	Xenobiotic transport	GO:0042908	Mdr49
1	28.51	DNA replication-removal of RNA primer	GO:0043137	Rnaseh1
2	28.51	Negative regulation of endocytosis	GO:0045806	ABCA7,Mctp1
2	28.51	**Negative gravitaxis**	GO:0048060	pain,pyx
1	28.51	Imaginal disc-derived wing expansion	GO:0048526	Spn88Ea
1	28.51	**Positive regulation of synapse structural plasticity**	GO:0051835	FRMPD4
1	28.51	Chitin biosynthetic process	GO:0006031	chs-2
2	19.01	Lysosomal transport	GO:0007041	Vps53,g
2	19.01	RNA interference	GO:0016246	Dcr-1,bel
2	15.55	Negative regulation of endopeptidase activity	GO:0010951	Spn88Ea,Serpinb1a
2	13.16	Formation of cytoplasmic translation initiation complex	GO:0001732	eIF3-S5, BMSK0008517
2	10.69	Protein heterotetramerization	GO:0051290	pyx,Agrn
2	10.06	**Positive regulation of synaptic growth at neuromuscular junction**	GO:0045887	Agrn,Sin1
2	8.55	Endoplasmic reticulum organization	GO:0007029	atl,Lman1
2	7.78	**Regulation of synaptic growth at neuromuscular junction**	GO:0008582	atl,hiw
2	7.13	Negative regulation of cell migration	GO:0030336	SLC9A3R1,Mctp1
5	5.94	**Visual perception**	GO:0007601	RDH11,COL11A1, OP1,WDR36,disco

Table 5: Top significantly enriched GO terms for *A. pellucida* day upregulation

No. Genes	Fold Enrichment	Pathway	GO_ID	Gene Names
1	41.61	Female meiosis II	GO:0007147	polo
1	41.61	Female germline ring canal formation-actin assembly	GO:0008302	Src64B
1	41.61	Dosage compensation by hyperactivation of X chromosome	GO:0009047	mle
1	41.61	10-formyltetrahydrofolate catabolic process	GO:0009258	aldh1l1
1	41.61	**Sialic acid transport**	**GO:0015739**	**SLC17A5**

1	41.61	Arachidonic acid metabolic process	GO:0019369	FAAH2
1	41.61	Regulation of cellular metabolic process	GO:0031323	melt
1	41.61	Protein exit from endoplasmic reticulum	GO:0032527	SEC16A
1	41.61	Single strand break repair	GO:0000012	gkt
1	41.61	Ring gland development	GO:0035271	gl
1	41.61	Notch receptor processing-ligand-dependent	GO:0035333	Nct
1	41.61	Positive regulation of RNA polymerase II transcriptional preinitiation complex assembly	GO:0045899	PSMC6
1	41.61	Tetrahydrofolate metabolic process	GO:0046653	MTHFS
1	41.61	Neurotrophin TRK receptor signaling pathway	GO:0048011	Bcar1
1	41.61	Glucose catabolic process	GO:0006007	agp
1	41.61	Glutaminyl-tRNAGln biosynthesis via transamidation	GO:0070681	gatA
1	41.61	Error-free translesion synthesis	GO:0070987	POLH
1	41.61	Protein deubiquitination involved in ubiquitin-dependent protein catabolic process	GO:0071947	trbd
1	41.61	Negative regulation of release of cytochrome c from mitochondria	GO:0090201	Parl
1	41.61	Regulation of cardiac conduction	GO:1903779	nkx-2.5
1	41.61	Methionyl-tRNA aminoacylation	GO:0006431	MetRS-m
2	33.29	Tetrahydrofolate interconversion	GO:0035999	Sardh,MTHFS
1	27.74	Tubulin complex assembly	GO:0007021	Tbcd
1	27.74	Double-strand break repair via break-induced replication	GO:0000727	Cdc45
1	27.74	Aminophospholipid transport	GO:0015917	ATP11B
1	27.74	**Synaptic target attraction**	**GO:0016200**	**Ten-a**
1	27.74	N-terminal peptidyl-methionine acetylation	GO:0017196	Naa60
1	27.74	Peptidyl-lysine methylation	GO:0018022	Eef1akmt2
1	27.74	Negative regulation of ossification	GO:0030279	LRP4
1	27.74	Negative regulation of phosphoprotein phosphatase activity	GO:0032515	Bod1
1	27.74	Store-operated calcium entry	GO:0002115	olf186-F
1	27.74	Cytoplasmic sequestering of transcription factor	GO:0042994	Sufu
1	27.74	Melanin biosynthetic process from tyrosine	GO:0006583	BMSK0009475
1	20.80	Cytoplasmic mRNA processing body assembly	GO:0033962	patl1

1	20.80	Positive regulation of transcription from RNA polymerase I promoter	GO:0045943	HEATR1
1	20.80	Negative regulation of insulin secretion	GO:0046676	HADH
1	16.64	Microtubule nucleation	GO:0007020	TUBG1
1	16.64	**Eclosion rhythm**	**GO:0008062**	**disco**
1	16.64	Late endosome to vacuole transport	GO:0045324	chmp3
1	16.64	**Negative regulation of axon extension involved in axon guidance**	**GO:0048843**	**tap**
1	16.64	Glucose transmembrane transport	GO:1904659	SLC2A3
1	16.64	Nucleocytoplasmic transport	GO:0006913	Anp32a
2	13.87	Chromosome segregation	GO:0007059	PPP2R1A,Naa60
1	13.87	Nucleotide catabolic process	GO:0009166	BMSK0009399
1	13.87	Phosphatidylethanolamine biosynthetic process	GO:0006646	PCYT2
2	12.80	Negative regulation of sequence-specific DNA binding transcription factor activity	GO:0043433	melt,Sufu
2	10.40	Protein heterotetramerization	GO:0051290	pyx,LRP4
2	9.25	Phospholipid biosynthetic process	GO:0008654	Mboat1,PCYT2
2	9.25	**Locomotor rhythm**	**GO:0045475**	**Ork1,disco**
2	8.76	Pupal chitin-based cuticle development	GO:0008364	Gld,Gld
3	8.61	Mitochondrial translation	GO:0032543	BMSK0007759,MRPS35,gatA
2	8.32	Sperm storage	GO:0046693	Gld,Gld
2	7.93	Cytoplasmic translation	GO:0002181	RpL4,
2	7.93	Synaptic growth at neuromuscular junction	GO:0051124	LRP4,Ten-a
2	7.56	Negative regulation of canonical Wnt signaling pathway	GO:0090090	PSMC6,LRP4
2	6.93	**Mushroom body development**	**GO:0016319**	**tap,Src64B**
2	6.93	Post-translational protein modification	GO:0043687	TULP4,PSMC6
2	6.93	**Synapse organization**	**GO:0050808**	**Ten-a,LRP4**
2	5.74	**Trehalose transport**	**GO:0015771**	**Tret1-1,Tret1-1**
2	5.20	Positive regulation of canonical Wnt signaling pathway	GO:0090263	PSMC6,trbd
3	5.09	Cell migration	GO:0016477	trbd,Bcar1,SPEF1
2	5.04	Response to endoplasmic reticulum stress	GO:0034976	SEC16A,Pdi

3	4.62	Mitotic cell cycle	GO:0000278	TUBG1,Tbcd,polo
2	4.62	Microtubule cytoskeleton organization	GO:0000226	TUBG1,Tbcd
3	3.37	G-protein coupled receptor signaling pathway	GO:0007186	Bcar1,Atrnl1,TRHR

Table 6: Top significantly enriched GO terms for *D. rubicunda* night upregulation

Gene Symbol	Gene Name	Function	DESeq2	EdgeR	WGCNA	Flybase ID	SILKD ID
BNC1/disco	disconnected	vision	1*	1*		FBgn0000459	BMSK0012241
Spt5	Spt5	transcription	1*	1*		FBgn0040273	BMSK0011143
1q/trbd	trabid	development		1		FBgn0037734	BMSK0015294
RpL4	Ribosomal protein L4	energy use, Ribosomal production	1*			FBgn0003279	BMSK0006059
Prp2/PPO2	Prophenoloxidase 2	Wound melanization	1*			FBgn0033367	BMSK0009475
mRpS5	mitochondrial ribosomal proteins	energy use, mitochondrial maintenance	1*			FBgn0287187	BMSK0014662
SLC2A6/Tret1-1	Solute Carrier Family 2 Member 6	locomotion, energy metabolism	1*			FBgn0050035	BMSK0003817**, BMSK0003818
TUBG1/γTub23C	tubulin gamma 1	brain development in adult	1*			FBgn0260639	BMSK0002451
PARL/rho-7	presenilin associated rhomboid like	mitochondrial maintenance	1*			FBgn0033672	BMSK0008635
SLC17A5/MFS10	Solute carrier family 17 member 5	transmembrane ion transporter	1*			FBgn0030452	BMSK0001219
titin1/sls	sallimus	locomotion	1			FBgn0086906	BMSK0000202
unc-22/unc	uncoordinated	adult locomotion, hearing	1		1	FBgn0003950	BMSK0015609**, BMSK0000066
ARHGAP17/RhoGAP92B	rho GTPase-activating protein 17	auditory, brain development			1	FBgn0038747	BMSK0009977
tk	tachykinin	locomotion and circadian rhythms			1	FBgn0037976	BMSK0003725
BmorOBP1/	Odorant Binding	odor binding			1	FBgn0034768	BMSK0013390

Obp58a	Protein 1						
PPAP2A/wun	Inorganic Pyrophosphatase 2/ wunnen	vision			1	FBgn0016078	BMSK0011305
BmorOBP2/ Obp84a	Odorant Binding Protein 2	odor binding			1	FBgn0011282	BMSK0009610
takt/JHBP	takeout-like	circadian			1	FBgn0038395	BMSK0013046

Table 7: Gene symbol from EggNog or *B. mori* annotation. Gene names, function and closest flybase ID orthologs used to collect evidence for the various listed functions.* Indicated that the gene altered direction or trend of expression across a diel pair. ** Indicates this ortholog was used for structural modeling and conservation analyses. This list is only a subset of the genes recovered from various analyses. The genes were chosen because they had either 1) multiple lines of evidence, robust divergent expression pattern, or 2) a gene ontology term for sensory or circadian function among three datasets. For exhaustive lists, see Supplementary Data

Process	**Search Query**	**Excluding**
Vision	"vision" or "eye" or "visual" or " compound eye" or "visual perception" or "light" or "ocellus" or "visual" or "photoreceptor"	- "chain" - "immunoglobulin" - "light harvesting" - "photosynthesis" - "plastid" -"plant" -actin -myosin -photoperiod - touch -smell -"light chain"
Smell	"olfaction" or "smell" or "antennal" or "olfactory" or "odor" or "pheromone"	
Hearing	"audition" or "hearing" or "sound" or "auditory" or "tympanum" or "ear"	

Circadian	"circadian" or "rhythm" or "clock" or "diurnal" or "nocturnal" or "entrainment"	
Behaviour	"locomotion" or "flight" or "behavioral"	- "cell" - "morphogenesis" - "development" or - "pheromone"
Brain	"neural plasticity" or "brain" or "neuropeptide"	

Table 8: GO query terms for functional lookup