Resource availability affects activity profiles of regulatory elements in a long-distance butterfly migrant

Karin Näsvall[1], Daria Shipilina[1], Roger Vila[2], Gerard Talavera[3], and Niclas Backstrom[1]

[1]Uppsala University
[2]Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra)
[3]Institut Botanic de Barcelona

March 8, 2023

Abstract

The oogenesis-flight syndrome reflects the temporal allocation of energy resources between dispersal / migration and reproduction and is a key concept in research on migratory behaviour in animals. In migratory butterflies, host plant abundance and quality may act as environmental cues to switch between the two states, but the mechanisms regulating this process are virtually unknown. Here, we used an experimental set-up to assess how variation in host plant abundance affected the activity of regulatory elements in the painted lady butterfly (*Vanessa cardui*), a model species for insect migratory behaviour studies. Chromatin immunoprecipitation (ChIP-seq) was used to evaluate histone tail modifications of H3K27ac and H3K4me3, as a proxy for regulatory activity. The results indicate that recently eclosed females that had access to host plants invested in reproduction at an earlier stage and that variation in host plant abundance triggered significant differences in regulatory element activity via histone tail acetylation. The functions of genes in the vicinity of differentially activated regions were primarily associated with metabolism, egg shell formation, female receptivity, muscle activity, pheromone binding and chromosome maintenance. Our results provide a first glimpse into the regulatory underpinnings of how females perceive the environment and allocate resources for either migration or reproduction and a starting point for more detailed understanding of the links between environmental variation, gene regulation and behaviour in butterflies.

Resource availability affects activity profiles of regulatory elements in a long-distance butterfly migrant

Karin Näsvall[1*], Daria Shipilina[1], Roger Vila[2], Gerard Talavera[3], Niclas Backström[1]

[1]Evolutionary Biology Program, Department of Ecology and Genetics (IEG), Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden

[2]The Butterfly Diversity and Evolution Lab, Institut de Biologia Evolutiva, Passeig Martim de la Barceloneta 37-49, 08003 Barcelona, Spain

[3]Institut Botànic de Barcelona (IBB), CSIC-Ajuntament de Barcelona, Passeig del Migdia s/n, 08038 Barcelona, Spain

Emails:

Karin Näsvall: karin.nasvall[at]ebc.uu.se

Daria Shipilina: daria.shipilina[at]ebc.uu.se

Roger Vila: roger.vila[at]csic.es

Gerard Talavera: gerard.talavera[at]csic.es

Niclas Backström: niclas.backstrom[at]ebc.uu.se

* Correspondence: karin.nasvall[at]ebc.uu.se

Key words

ChipSeq, Migration syndrome, Regulatory elements, Enhancer, Promoter,

Vanessa cardui.

Running title:

Regulatory elements, oviposition and migratory behaviour in *V. cardui* .

Abstract

The oogenesis-flight syndrome reflects the temporal allocation of energy resources between dispersal / migration and reproduction and is a key concept in research on migratory behaviour in animals. In migratory butterflies, host plant abundance and quality may act as environmental cues to switch between the two states, but the mechanisms regulating this process are virtually unknown. Here, we used an experimental set-up to assess how variation in host plant abundance affected the activity of regulatory elements in the painted lady butterfly (*Vanessa cardui*), a model species for insect migratory behaviour studies. Chromatin immunoprecipitation (ChIP-seq) was used to evaluate histone tail modifications of H3K27ac and H3K4me3, as a proxy for regulatory activity. The results indicate that recently eclosed females that had access to host plants invested in reproduction at an earlier stage and that variation in host plant abundance triggered significant differences in regulatory element activity via histone tail acetylation. The functions of genes in the vicinity of differentially activated regions were primarily associated with metabolism, egg shell formation, female receptivity, muscle activity, pheromone binding and chromosome maintenance. Our results provide a first glimpse into the regulatory underpinnings of how females perceive the environment and allocate resources for either migration or reproduction and a starting point for more detailed understanding of the links between environmental variation, gene regulation and behaviour in butterflies.

Introduction

Temporal heterogeneity in suitable habitat for reproduction is a main driver of insect adaptations. Some species cope with seasonally unfavourable conditions by entering a phase of dormancy, the diapause. Other species move between distinct geographic regions in more or less regular circuits to utilize temporally and spatially variable habitat patches. In insects, such migratory movements predominantly occur by flight and migration itself is obviously associated with individual costs (Dingle & Drake, 2007). Migratory flight can for example increase the risk for predation, and unpredictable weather conditions during migration can lead to severe deviations from the optimal flight direction and distance. In addition, migration may not necessarily lead to that the individual ends up in a more suitable habitat patch than the one it left. The energetic demands for migratory flight are also extremely high and a migrating individual has to allocate considerable resources to sustain the effort, resources that could otherwise be used for investment in reproduction (Wegener, 1996). Investment in reproduction can also come with a cost if conditions for egg laying are suboptimal and/or if host plant availability is limited. The evolution of migratory behaviour hence involves a cost-benefit trade-off, in essence reflecting the more short-term benefits of investing resources in reproduction in a potentially less suitable habitat patch versus the more long-term advantages of dispersing to, and colonizing, habitat patches with more suitable conditions for feeding and oviposition (Chapman et al., 2015).

The "oogenesis-flight syndrome" is a model for migratory insects which postulates that migratory and reproductive states are temporally separated during the life span of adult females (Chapman & Drake, 2019; Johnson, 1969). Although this generalisation does not apply to all migratory insects – observations suggest that some species are reproductively active also during the migratory phase (e.g. Jiang et al., 2010; Tigreros and Davidowitz, 2019) – there is a general consensus that the oogenesis-flight syndrome serves as a robust overarching model for insect migration studies (Bhaumik & Kunte, 2018; Chapman & Drake, 2019; Dingle, 2014). Two of the main model systems for butterfly migration are the monarch (*Danaus chrysippus*) and the

painted lady butterfly (*Vanessa cardui*). In monarchs, females fulfil the oogenesis-flight syndrome during fall migration, but not in spring when they leave the overwintering areas and start migrating northwards (Ruiz Vargas et al., 2018). In the painted lady on the other hand, empirical data point towards lower mating propensity in females during the migratory phase indicating reduced investment in reproduction (Stefanescu et al., 2021). Besides that the oogenesis-flight syndrome potentially can be applied to the migrating female painted ladies, the species exhibits life-history characteristics that reflect long-distance migratory behaviour within different circuits that cover almost all continents – the best-studied circuit spans the range between sub-Saharan Africa and northern Europe (Menchetti et al., 2019; Stefanescu et al., 2013; Talavera et al., 2018; Talavera & Vila, 2016). In contrast to the monarch, painted ladies never enter diapause and the annual migratory circuit involves multiple generations (6-8, or even more) (Menchetti et al., 2019; Talavera & Vila, 2016). The species is highly polyphagous and can therefore utilize the wide range of host plants encountered during alternating seasons in different parts of the almost cosmopolitan distribution range (Ackery, 1988). The painted lady butterfly therefore constitutes an attractive model for investigating both the mechanistic basis of different aspects of migratory behaviour in general and the genetic underpinnings of the oogenesis-flight syndrome specifically.

The behavioural and physiological changes induced by environmental cues are mediated by the activation of different regulatory elements involved in transcriptional regulation. One way to characterise the genes and pathways involved is therefore to assess the chromatin state in regulatory elements. Chromatin accessibility and transcriptional activity is largely determined by histone tail modifications. For example, acetylation of the histone tail residue lysine 27 (H3K27ac) reduces the affinity of DNA to the histones. This leads to accessible chromatin around promoters and enhancers of actively transcribed genes (Rada-Iglesias et al., 2011). H3K4me3 on the other hand is predominantly positioned within promoters and proximal regulatory elements and has been associated with transcriptional activity via recruitment of transcription factors and histone modifiers (Santos-Rosa et al., 2002), although the precise molecular mechanisms here are not yet fully understood (Beacon et al., 2021; Howe et al., 2017). Chromatin immunoprecipitation followed by DNA-sequencing (ChIP-seq) is a method to identify DNA-sequences associated with attached proteins. By quantifying the enrichment of specific histone tail modifications, both the position and activity of regulatory elements and the adjacent genes can be characterized.

Here we used an experimental set-up where recently eclosed female painted lady butterflies were exposed to environments that differed in the availability of host plants. By quantifying differences in activity of regulatory elements between females from the treatment groups, we provide a first glimpse into how resource availability affects activation of different genetic pathways in a migratory butterfly.

Materials and methods

Experimental set-up

Behavioral changes including induction of oviposition have been observed to start approximately 4 - 5 days after eclosure, both in *V. cardui* (Stefanescu et al., 2021) and other nymphalid butterfly species (Wiklund & Friberg, 2022). This suggests that the 'decision' to invest in migration or reproduction, which can be influenced by, among other things, host plant availability for oviposition, should have been made at this time. To test how the response to host plant access affects the behaviour and investment into reproduction we exposed the females to presence or absence of host plants and sampled the females in the morning five days after eclosure to analyse differential histone mark enrichment at this time point.

Offspring to one painted lady female (3^{rd} generation offspring of wild caught individuals captured in Catalonia in spring 2020) were reared under controlled conditions, including constant temperature at 23°C, constant day length at 16 hours (h) light, 4 h ambient light and 4 h darkness, and *ad libitum* access to host plants (*Malva sylvestris* , grown under controlled conditions in a common greenhouse). Upon pupation, the sex of each pupa was scored and when the imagines eclosed, individuals were immediately divided into two treatment cohorts (separate cages) using two replicate cages for each treatment. A first group of females was placed in one of two cages (40*40*40 cm) with five free-flying males and access to the host plant *M. sylvestris*

. The female individuals in the other treatment group were placed in one of two cages with five free-flying males and no host plants. All butterflies had *ad libitum* access to energy sources (20% sugar water) and the climatic settings were again a stable temperature at 23°C, constant day length at 16 h light, 4 h ambient light and 4 h darkness. The females used for the experiment were sampled in the morning on the fifth day after eclosure and immediately snap frozen in liquid nitrogen for subsequent DNA-extraction. The host plants were exchanged every second day and the number of eggs were recorded in each specific cage. A total of 47 females were sampled, 12 from each of the four treatments/replicates, except for one treatment (with host plants) where only 11 females could be obtained.

ChIP-Seq preparation and sequencing

To obtain an overview of the chromatin state and regulatory element activity in *V. cardui* females during a critical time point of the oogenesis-flight syndrome progression, we focused on head and antennae. The rationale behind this selection was that nervous tissue is highly concentrated to the head of butterflies and environmental cues are perceived and processed by the sensory organs. Analysing head tissue was therefore a logical first step to understand potential differential regulatory element activities related to the propensity for migration.

Heads, including antennae, of the snap frozen individuals were ground to fine powder in liquid nitrogen. Ice cold PBS with a Protease Inhibitor Cocktail (PIC) was added before cross-linking with 1% formaldehyde for seven minutes at room temperature. The reaction was halted with 125 mM glycine and incubated at room temperature for five minutes. Samples were then centrifuged for five minutes at 500 rpm in 4°C and subsequently washed twice with ice-cold PBS + PIC. The samples were continuously flash frozen at -80°C until all samples had been processed.

For the chromatin immunoprecipitation we used the SimpleChIP® Enzymatic Chromatin IP Kit (Cell Signaling Technology Inc.) with some modifications (see below). The chromatin was prepared by homogenizing the tissue with a pestle in eppendorf tubes and then pooling four samples (three pools per treatment/replicate). The cells were lysed with DTT lysis buffer according to the manufacturer's instructions. The chromatin was fragmented by adding 0.5 µl Micrococcal Nuclease to each pool and incubation at 37°C for 10 minutes. The nuclear membrane was lysed with a Bioruptor Pico sonicator using six cycles of 30 second (s) pulses and 30 s pauses. The samples from each treatment were then pooled to get a single sample per replicate. The fragment size distribution of the chromatin was quantified with a 2100 Bioanalyzer (Agilent Inc.) after extracting DNA from a subset of the sample product. The immunoprecipitation was performed according to the manufacturer's protocol with some minor modifications (see below). A small aliquot (10µl) of the digested chromatin was set aside before adding antibodies as an input control sample. We used specific antibodies against the histone tail modifications H3K27ac and H3K4me3, which are associated to active enhancers and promoters. 2.5 µg of Rabbit monoclonal H3K27ac, H3K4me3 or negative control IgG was added to one aliquot from each replicate, respectively. Each aliquot contained approximately 3 µg of digested chromatin. The reactions were incubated overnight and antibody-chromatin complexes were extracted with magnetic beads according to the manufacturer's protocol (except the extended 2 hours at 65°C with rotation and vortexing five times to eluate the chromatin from the beads). Reverse cross-linking was performed with Proteinase K lysis for 2 h at 56°C, and finally, DNA purification was performed with columns provided in the kit. The input control sample was extracted together with the immunoprecipitated samples. The purity and fragment distribution of the DNA was assessed with Nanodrop (Thermo Scientific Inc.) and a 2100 Bioanalyzer (Agilent Inc.). A total number of 12 libraries were prepared (2 treatments x 2 replicates x 3 extractions including two precipitations and the input sample for each pool) with SMARTer ThruPLEX DNA-seq and sequencing of 2x150 bp reads was performed on a single Illumina NovaSeq6000 S4-300 lane at the National Genomics Infrastructure (NGI, see acknowledgements) in Stockholm.

Data processing and analysis

We used the *nf-core-chipseq* pipeline version 1.2.2 (Ewels et al., 2020; Patel et al., 2021) to identify differentially enriched peaks. The pipeline includes *Trimgalo-*

re(https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) for trimming and adapter removal. We also used *BWA-mem* (Li, 2013) to map the reads to a high-contiguity genome assembly of the painted lady (Lohse et al., 2021). The Model based Analysis of ChIP-Seq (*MACS2)* package (Zhang et al., 2008) was applied to identify read coverage significantly higher than the random genome-wide variation and to construct consensus peaks. For determining number and location of enrichment of histone marks we used only consensus peaks called in all four groups (treatment*replicates). The analysis of differential activation with hostplant availability as contrast was performed with this requirement relaxed to account for differential enrichment of peaks absent in one of the treatments. The count data for the peaks were transformed with *Voom* (Law et al., 2014), for linear models in the R-package *Limma* (Ritchie et al., 2015) to detect differentially enriched peaks between the two treatment groups. To correct for multiple testing, the p-values were adjusted with the Benjamini-Hochberg false discovery rate as implemented in *Limma* . We used previously available annotation information (Shipilina et al., 2022) to identify the gene located closest to each differentially activated region. Potential functions of candidate genes were obtained from the annotation in combination with homology searches with*BLAST* to the NCBI database using the nucleotide sequence of each candidate gene (Altschul et al., 1990). Additional functional information was extracted from *Flybase* (https://flybase.org).

Results

Reproduction

As a proxy for oviposition behaviour we counted the number of eggs laid in each treatment group. The females in the treatment group with access to *M. sylvestris* had started laying eggs in large numbers at day five after eclosure. In total, we counted > 850 and > 500 eggs in the two replicate cages with host plants after harvesting all females. In the two cages without host plants, we found no (0) or a smaller number of eggs (90) in each replicate, respectively. To investigate if host plant abundance had an effect on enhancer and promoter activity of specific genes, we proceeded with analysis of the ChIP-Seq data from the two treatment groups.

Results from the ChIP-Seq analysis

For the precipitation with H3K27ac, a total number of 14,100 activity peaks were identified across all groups (two treatment groups * two replicates). Of these, 3,368 were located in proximal regions (< 2.0 kb away from either the transcriptional start [TSS] or end sites [TES]) of annotated coding genes in the painted lady butterfly genome assembly (Lohse et al., 2021; Shipilina et al., 2022). Hence, a majority of the H3K27ac activity peaks were detected in intergenic regions. For the precipitation with H3K4me3, 9,846 activity peaks were identified in total, of which 4,744 were found within proximal regions from either the TSS or the TES of genes. Of the total number of H3K4me3-peaks, 5,383 overlapped with H3K27ac-peaks. The distribution pattern was consistent with the established association between H3K4me3 and proximal regulatory elements, with over half of the peaks annotated in promoters, introns or exons and a decline in the density of H3K4me3 from first introns (and exons) to genic regions further downstream (Figure 1). As expected, given the higher density of regulatory elements in the vicinity of gene regions, there was a pronounced enrichment of activity peaks close to the TSS and a considerable drop in activity signal just downstream of the TES for both the acetylation (H3K27ac) and the methylation (H3K4me3) mark (Figure 1). The reduced activity directly upstream the TSS is expected since histones are dislodged from the TSS during active transcription.

Hosted file

`image1.emf` available at https://authorea.com/users/593676/articles/628552-resource-availability-affects-activity-profiles-of-regulatory-elements-in-a-long-distance-butterfly-migrant

Figure 1. a) Regulatory element activity profiles summarized jointly over all annotated genes in the painted lady butterfly genome assembly. The Y-axis shows the average genome-wide enrichment of H3K27ac (both promoter and enhancer regions) and H3K4me3 (predominantly promoter regions) for all genes in the anno-

tation. The X-axis shows the absolute positions (3 kb) of the upstream and downstream stretches around the transcribed regions and the relative position summarised over all genes within the transcribed regions. TSS and TES represent the transcription start and end sites, respectively. b) Number of shared H3K27ac and H3K4me3 peaks among all groups, genome wide and within 2 kb of a TSS. The overlap represents peaks with both H3K27ac and H3K4me3 enrichment. The bottom panels show the number of observed H3K27ac (c) and H3K4me3 (d) activity peaks in different genomic regions. The pie charts illustrate the overall number of activity peaks across genomic regions and the barplot shows the percentage of exons/introns with observed activity peaks in consecutive exons (brown) and introns (blue) along genes (x-axis).

Gene categories with differential regulatory element activity

To obtain information about potential pathways affected by the treatment (host plant availability), we identified differentially acetylated or methylated regions between the treatment groups. For the H3K27ac mark, we identified 9 proximal gene regions with significant differential activity peaks between treatment groups (Figure 2). Seven of the genes in the vicinity of H3K27ac peaks with differential activity had annotation information, suggesting that they are involved in regulation of female receptivity to mating (1 gene), chorion (insect egg shell) gene amplification (1 gene), pheromone binding activity / sensory perception of smell (1 gene), actin binding in muscle tissue and motor neuron to muscle communication (2 genes), regulation of juvenile hormone (1 gene) and mini-chromosome maintenance (1 gene) (Supplementary Table 1). For the H3K4me3 mark, there were no significantly differentially activated regions after adjusting for multiple testing.

Hosted file

`image2.emf` available at https://authorea.com/users/593676/articles/628552-resource-availability-affects-activity-profiles-of-regulatory-elements-in-a-long-distance-butterfly-migrant

Figure 2. A volcano plot showing the significance level (Y-axis, - logarithmic p-value) and logarithmic fold change in activity (X-axis) in the comparison between treatment groups for the analysis of H3K27ac. Non-significant loci are indicated with grey dots and loci with significantly (adjusted for multiple testing) different acetylation activity with red dots.

We proceeded and intersected the signals from the two different histone marks. This analysis unveiled that six of the 9 regions with significantly different activity signals in H3K27ac also had considerable activity peaks for H3K4me3, and an additional proximal differentially activated peak close to the gene *SLC* which has been inferred to be involved in carbohydrate metabolism.

Discussion

Here we used an experimental set-up where female painted lady butterflies were exposed to environments that varied in host plant abundance. We combined this with analysing regulatory element activity to get insights into the pathways that were differentially regulated between painted lady individuals exposed to different environmental conditions. Besides giving information about differences in transcriptional activity between female butterflies exposed to different environmental settings, this is one of the first attempts to characterise the genome wide distribution of H3K27ac and H3K4me3 in butterflies in detail.

We found that females that had access to host plants had laid significantly more eggs than females in cages without host plants. This is in line with the observations that female butterflies can sense both relative abundance of host plants and presence of conspecifics (Mugrabi-Oliveira & Moreira, 1996), and the amount of secondary compounds for oviposition selection (Reudler Talsma et al., 2008). The higher egg laying propensity in the treatments where females had access to host plants can obviously just be a consequence of availability of host plant substrate, and not indicate a delayed investment in reproduction. However, the observation supports previous results that indicate a higher frequency of reproductively active females in areas where host plants are abundant (Stefanescu et al., 2021).

In order to investigate how the differences in environment affected the activity of regulatory elements, we harvested the females from the different experimental cohorts at day five after eclosure and analysed regulatory element activity using ChIP-Seq. Information about the genome-wide regulatory landscape in Lepidoptera is limited to a few species. As a first step, we therefore identified genome-wide activity peaks for both H3K27ac and H3K4me3. Previous studies in Lepidoptera have shown that histone tail modifications (and regulatory activity) can vary across tissues and between developmental and metabolic stages (Cheng et al., 2018; Lewis et al., 2016). In the painted lady butterfly, 4,744 of the H3K4me3 peaks were located in proximal regions of genes potentially representing promoters and proximal regulatory elements. This is approximately in the same range as what has previously been observed in both *B. mori* (n = 5,599 'proximal elements') and *Heliconius erato* (n = 5,399) (Cheng et al., 2018; Lewis et al., 2016). The observed agreement in both absolute numbers and the spatial distribution of activity peaks suggests that the genome wide distribution of these active histone marks has been accurately characterized in the painted lady butterfly.

The core question in this study was to investigate potential differences in regulatory element activity between treatment groups that could inform on how environmental differences affect gene regulation and, ultimately, the behaviour of individual butterflies. Only H3K27ac showed loci with significant differential activity between the treatment groups and 9 of those were in the vicinity of coding genes. When overlaid with information from the analysis of H3K4me3, we found that seven regions with differential activity coincided with H3K4me3 modifications. Hence, it is likely that the chromatin in the regions of these particular genes was accessible, since H3K4me3 activity was present in both treatments. In addition, H3K27ac and H3K4me3 have been shown to interact to enhance the transcriptional activity (Zhao et al., 2021) and it is therefore plausible that differential acetylation corresponds to significant differences in activity between individuals exposed to different environmental conditions (i.e. different host plant densities).

The potential functions of the significantly differentially activated genes were assessed and we found gene ontology information for eight of the candidate genes. This set included two genes associated with the mini-chromosome maintenance (*MCM*) – the genes *MCM6* and *GEMININ* - which have been shown to be involved in the regulation of *MCM6* (Kushwaha et al., 2016). The *MCM* gene family consists of several gene copies that jointly affect chromatin unwinding and the complex has been shown to be involved in DNA-replication (Tsuruga et al., 2016). Interestingly, *GEMININ* , which showed a significant activity peak in individuals exposed to high host plant density, has been shown to be involved in DNA-replication control in *B. mori* (Tang et al., 2017) and expressed in insect ovaries during oogenesis (Quinn et al., 2001). It is hence tempting to speculate that the higher activity is associated with the more advanced reproductive mode in this treatment group. We also found that the genes *TTN* (titin-like) and *SLC* (solute carrier family 2) showed higher activity in the treatment where females had access to host plants. *SLC* is a member of a large family of transmembrane genes where the gene product forms a solute carrier that mediates transport of, for example, glucose and amino acids across cell membranes (Hediger et al., 2004). The titin-like gene product is a structural protein involved in muscle formation and function (https://flybase.org/reports/FBgn0086906.html; accessed 2022-12-09) and a different activity of regulatory regions of *TTN* in the treatment group with host plants is in line with the predictions of the oogenesis-flight syndrome, since flight muscle activity should be reduced when reproductive mode is more advanced compared to when individuals are in the migratory phase. In agreement with this, we found that an activity dependent transporter gene (*ARC1*) had significantly higher acetylation levels in the second intron in the group without host plant. In *Drosophila* , *ARC1* mediates transfer of mRNA from motor neurons to the muscle tissue (Ashley et al., 2018) and it has also been linked body fat accumulation (Mosher et al., 2015). Both functions can be directly linked to migratory behaviour since motor neuron activity and efficient fat storage are key components of long-distance flight. The gene *SPR* (sex peptide receptor) also showed differential regulatory element activity between treatment groups, with a higher activity in the individuals exposed to an environment without host plants. *SPR* is expressed in the central nervous system and the reproductive tract in *Drosophila melanogaster* females where it has been shown to regulate post-mating responses, for example sperm release, egg-laying capacity and reduced receptivity (Avila et al., 2015; Haussmann et al., 2013). It seems plausible that this pattern reflects delayed female receptivity in female painted ladies when no host plants are available for egg laying.

Another interesting gene that showed differential regulatory activity was the odorant binding protein, *OBP* . *OBP* expression is associated with pheromone binding activity and perception of smell (Hekmat-Scafe et al., 2002). Since pheromone signalling likely is a key component for courtship and mating acceptance in many lepidopterans, the difference in activity could be a result of differences in reproductive status between treatment groups. Finally, the butterflies with access to host plants showed a higher acetylation peak in the second intron of a juvenile hormone esterase (*JHE)* . *JHE* is involved in degradation of juvenile hormone (*JH)* , a key hormone in female reproductive maturation (Herman & Dallman, 1981) and associated with dispersal versus reproduction decisions in many insects (Ramaswamy et al., 1997). The increased activation of *JHE* in the butterflies with access to host plants is perhaps counterintuitive. However, since we harvested females five days after eclosure, it is possible, or even likely, that the differences in regulatory element activity we detected result from cascading effects that were initiated at an earlier time point. This does obviously not only concern the *JHE* / *JH*regulation, but all other differentially activated regions. Future efforts to characterize the genetic / regulatory underpinnings of the oogenesis-flight syndrome might benefit from sampling across multiple time points to cover the temporal dynamics of regulatory cascades.

Our results provide a starting point to investigate how regulation of the identified candidate genes affect individual behavioural strategies in painted lady butterflies and could also be relevant for future assessments of the generality of specific pathways underlying the oogenesis-flight syndrome in insects at a larger scale.

Acknowledgements

We would like to acknowledge members of the Backström lab, especially Orazioluca Paternò, for their work in the butterfly lab and Patrick Nylund at the Department of Immunology, Genetics and Pathology, Rudbeck Laboratory, Uppsala University, for providing the sonicator. This work was funded by the Swedish Research Council FORMAS (FORMAS research grant #2019-00670 to N.B.). The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. This work was also supported by NBIS/SciLifeLab long-term bioinformatics support (WABI). R.V. was supported by project PID2019-107078GB-I00/MCIN/AEI/10.13039/501100011033. G.T. was supported by the grant PID2020-117739GA-I00 MCIN / AEI / 10.13039/501100011033; all authors were also supported by the grant LINKA20399 from the CSIC iLink program.

Data availability

The raw sequence data is deposited in the European Nucleotide Archive (ENA) with accession number PRJEB59028. Scripts for the analyses are available in GitHub (https://github.com/EBC-butterfly-genomics-team).

References

Ackery, P. R. (1988). Hostplants and classification: A review of nymphalid butterflies. *Biological Journal of the Linnean Society* ,*33* , 95–203.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* , *215* , 403–410.

Ashley, J., Cordy, B., Lucia, D., Fradkin, L. G., Budnik, V., & Thomson, T. (2018). Retrovirus-like Gag protein Arc1 binds RNA and traffics across synaptic boutons. *Cell* , *11* , 262–274.

Avila, F. W., Mattei, A. L., & Wolfner, M. F. (2015). Sex peptide receptor is required for the release of stored sperm by mated*Drosophila melanogaster* females. *Journal of Insect Physiology* , *76* , 1–6.

Beacon, T. H., Delcuve, G. P., López, C., Nardocci, G., Kovalchuk, I., van Wijnen, A. J., & Davie, J. R. (2021). The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes. *Clinical Epigenetics* , *13* , 138.

Bhaumik, V., & Kunte, K. (2018). Female butterflies modulate investment in reproduction and flight response to monsoon-driven migrations. *Oikos* , *127* , 285–296.

Chapman, J. W., & Drake, V. A. (2019). Insect migration. In *In Encyclopedia of animal behavior.* (M.D. Breed, J. Moore). Academic Press.

Chapman, J. W., Reynolds, D. R., & Wilson, K. (2015). Long-range seasonal migration in insects: Mechanisms, evolutionary drivers and ecological consequences. *Ecology Letters* , *18* , 287–302.

Cheng, D., Cheng, T., Yang, X., Zhang, Q., Fu, J., Feng, T., Gong, J., & Xia, Q. (2018). The genome-wide transcriptional regulatory landscape of ecdysone in the silkworm. *Epigenetics and Chromatin* ,*11* , 48.

Dingle, H. (2014). *Migration. The biology of life on the move.*Oxford University Press.

Dingle, H., & Drake, V. A. (2007). What is migration? *Bioscience* , *57* , 113–121.

Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology* , *38* , 3.

Haussmann, I. U., Hemani, Y., Wijesekera, T., Dauwalder, B., & Soller, M. (2013). Multiple pathways mediate the sex-peptide-regulated switch in female *Drosophila* reproductive behaviours. *Proceedings of the Royal Society B: Biological Sciences* , *280* , 20131938.

Hediger, M. A., Romero, M. F., Peng, J. B., Rolfs, A., Takanaga, H., & Bruford, E. A. (2004). The ABCs of solute carriers: Physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflügers Archiv* , *447* , 465–468.

Hekmat-Scafe, D. S., Scafe, C. R., McKinney, A. J., & Tanouye, M. A. (2002). Genome-wide analysis of the odorant-binding protein gene family in *Drosophila melanogaster* . *Genome Research* , *12* , 1357–1369.

Herman, W. S., & Dallman, S. H. (1981). Endocrine biology of the painted lady butterfly *Vanessa cardui* . *Journal of Insect Physiology* , *27* , 163–168.

Howe, F. S., Fischl, H., Murray, S. C., & Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? Françoise S. Howe, Harry Fischl, Struan C. Murray, Jane Mellor. *BioEssays* , *39* , 1–12.

Jiang, X. F., Luo, L. Z., & Sappington, T. W. (2010). Relationship of flight and reproduction in beet armyworm, *Spodoptera exigua*(Lepidoptera: Noctuidae), a migrant lacking the oogenesis-flight syndrome. *Journal of Insect Physiology* , *56* , 1631–1637.

Johnson, C. G. (1969). *Migration and dispersal of insects by flight.* Methuen.

Kushwaha, P. P., Rapalli, K. C., & Kumar, S. (2016). Geminin a multi task protein involved in cancer pathophysiology and developmental process: A review. *Biochimie* , *131* , 115–127.

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* , *15* , R29.

Lewis, J. J., van der Burg, K. R. L., Mazo-Vargas, A., & Reed, R. D. (2016). ChIP-Seq-annotated *Heliconius erato* genome highlights patterns of cis-regulatory evolution in Lepidoptera. *Cell Reports* , *16* , 2855–2863.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv* , *00* , 1–3.

Lohse, K., Wright, C., Talavera, G., Garcia-Berro, A., & Darwin Tree of Life Consortium. (2021). The genome sequence of the painted lady, *Vanessa cardui* , Linnaeus 1758. *Wellcome Open Research* ,*6* , 324.

Menchetti, M., Guéguen, M., & Talavera, G. (2019). Spatio-temporal ecological niche modelling of multigenerational insect migrations. *Proceedings of the Royal Society Series B: Biological Sciences* ,*286* , 20191583.

Mosher, J., Zhang, W., Blumhagen, R. Z., D'Alessandro, A., Nemkov, T., Hansen, K. C., Hesselberth, J. R., & Reis, T. (2015). Coordination between *Drosophila* Arc1 and a specific population of brain neurons regulates organismal fat. *Developmental Biology* ,*405* , 280–290.

Mugrabi-Oliveira, E., & Moreira, G. R. P. (1996). Conspecific mimics and low host plant availability reduce egg laying by *Heliconius erato phyllis* (Fabricius) (Lepidoptera, Nymphalidae). *Revista Brasileira de Zoologia* , *13* , 929–937.

Patel, H., Wang, C., Ewels, P., Silva, T. C., Peltzer, A., Behrens, D., Garcia, M., mashehu, Rotholandus, Haglund, S., & Kretzschmar, W. (2021). *nf-core/chipseq: Nf-core/chipseq v1.2.2 - Rusty Mole* . Zenodo.

Quinn, L. M., Herr, A., McGarry, T. J., & Richardson, H. (2001). The*Drosophila* Geminin homolog: Roles for Geminin in limiting DNA replication, in anaphase and in neurogenesis. *Genes and Development* , *15* , 2741–2754.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* , *470* , 279–283.

Ramaswamy, S. B., Shu, S., Park, Y. I., & Zeng, F. (1997). Dynamics of juvenile hormone-mediated gonadotropism in the Lepidoptera. *Insect Biochemistry and Physiology* , *35* , 539–558.

Reudler Talsma, J. H., Biere, A., Harvey, J. A., & van Nouhuys, S. (2008). Oviposition cues for a specialist butterfly: Plant chemistry and size. *Journal of Chemical Ecology* , *34* (9), 1202–1212.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* ,*43* , e47.

Ruiz Vargas, N., Rowe, L., Stevens, J., Armagost, J. E., Johnson, A. C., & Malcolm, S. B. (2018). Sequential partial migration across monarch generations in Michigan. *Animal Migration* , *5* , 104–114.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C. T., Schreiber, S. L., Mellor, J., & Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* , *419* , 407-411.

Shipilina, D., Näsvall, K., Höök, L., Vila, R., Talavera, G., & Backström, N. (2022). Linkage mapping and genome annotation give novel insights into gene family expansions and regional recombination rate variation in the painted lady (*Vanessa cardui*) butterfly.*Genomics* , *114* , 110481.

Stefanescu, C., Páramo, F., Åkesson, S., Alarcón, M., Ávila, A., Brereton, T., Carnicer, J., Cassar, L. F., Fox, R., Heliölä, J., Hill, J. K., Hirneisen, N., Kjellén, N., Kühn, E., Kuussaari, M., Leskinen, M., Liechti, F., Musche, M., Regan, E. C., ... Chapman, J. W. (2013). Multi-generational long-distance migration of insects: Studying the painted lady butterfly in the Western Palaearctic. *Ecography* ,*36* (4), 474–486.

Stefanescu, C., Ubach, A., & Wiklund, C. (2021). Timing of mating, reproductive status and resource availability in relation to migration in the painted lady butterfly. *Animal Behaviour, 172* , 145–153.

Talavera, G., Bataille, C., Benyamini, D., Gascoigne-Pees, M., & Vila, R. (2018). Round-trip across the Sahara: Afrotropical painted lady butterflies recolonize the Mediterranean in early spring. *Biology Letters* , *14, 20180274* .

Talavera, G., & Vila, R. (2016). Discovery of mass migration and breeding of the painted lady butterfly *Vanessa cardui* in the Sub-Sahara: The Europe–Africa migration revisited. *Biological Journal of the Linnean Society* , *120* , 274–285.

Tang, X. F., Chen, X. Y., Zhang, C. D., Li, Y. F., Liu, T. H., Zhou, X. L., Wang, L., Zhang, Q., Chen, P., Lu, C., & Pan, M. H. (2017). Two Geminin homologs regulate DNA replication in silkworm, *Bombyx mori* . *Cell Cycle* , *16* , 830–840.

Tigreros, N., & Davidowitz, G. (2019). Flight-fecundity trade-offs in wing-monomorphic insects. *Advances in Insect Physiology* ,*56* , 1–41.

Tsuruga, H., Yabuta, N., Hosoya, S., Tamura, K., Endo, Y., & Nojima, H. (2016). HsMCM6: A new member of the human MCM/P1 family encodes a protein homologous to fission yeast Mis5. *Genes to Cells* ,*2* , 381–399.

Wegener, G. (1996). Flying insects: Model systems in exercise physiology. *Experientia* , *52* , 404–412.

Wiklund, C., & Friberg, M. (2022). Testing the migration syndrome: Comparative fecundity of migratory and non-migratory nymphaline butterflies. *Ecological Entomology* , *2022* , 1–7.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Li, W., & Liu, S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology* ,*9* , R137.

Zhao, W., Xu, Y., Wang, Y., Gao, D., King, J., Xu, Y., & Liang, F. S. (2021). Investigating crosstalk between H3K27 acetylation and H3K4 trimethylation in CRISPR/dCas-based epigenome editing and gene activation. *Scientific Reports* , *11* , 15912.

Supplementary information

Supplementary Table 1. Annotation information, genomic position and adjusted p-values for the significantly differentially activated genes identified in the comparison between treatment groups.

Gene name	**DToL gene name**	**Position**	**Adjusted p-value**	**Intersect**	**Distance TSS**
ODP	Vcard_DToL11162	Intergenic	5,07E-04	m3 + ac	-82,050
GEMININ	Vcard_DToL09546	Intergenic	5,25E-04	m3 + ac	-10,138
unknown	Vcard_DToL13032	Intergenic	5,25E-04	m3 + ac	-2,342
JHE	Vcard_DToL14308	Intron 2	7,14E-04		2,391
MCM6	Vcard_DToL16085	Intron 10	9,32E-03	m3 + ac	6,417
SPR	Vcard_DToL08760	Intergenic	1,65E-02	m3 + ac	-3,566
TTN	Vcard_DToL15083	Intergenic	1,95E-02	m3 + ac	-15,072
ARC1	Vcard_DToL15961	Intron 2	1,95E-02		5,281
unknown	Vcard_DToL15738	Intergenic	3,95E-02		62,802
SLC	Vcard_DToL16288	Intergenic	4,87E-02	m3 + ac	18,134

Gene name	**Inferred ortholog / function**
ODP	BLAST: Odorant-binding protein 19d.
GEMININ	DNA replication, eggshell (chorion) gene amplification.
Unknown	NA
JHE	BLAST: 80% match to juvenile hormone esterase-like.
MCM6	DNA replication licencing factor Mcm6 (mini-chromosome maintenance)
SPR	G-protein coupled receptor; sex peptide receptor-like isoform X2.
TTN	Titin-like, actin binding, muscle constituent.
ARC1	BLAST: activity-regulated cytoskeleton associated protein 1-like.
Unknown	NA
SLC	Solute carrier family 2, facilitated glucose transporter member 8.