Open access • Posted Content • DOI:10.1101/2020.09.04.279224

Mosaic evolution of molecular pathways for sex pheromone communication in a butterfly — Source link

Caroline M. Nieberding, Patrícia Beldade, Patrícia Beldade, Véronique Baumle ...+9 more authors

Institutions: Instituto Gulbenkian de Ciência, University of Lisbon, University of Paris, Institut national de la recherche agronomique

Published on: 05 Sep 2020 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Sex pheromone, Pheromone biosynthesis activating neuropeptide, Odorant binding and Bicyclus anynana

Mosaic evolution of molecular pathways for sex pheromone communication in a butterfly

Caroline M. Nieberding[1*], Patrícia Beldade[2,3], Véronique Baumlé[1], Gilles San Martin[1], Alok Arun[1], Georges Lognay[1], Nicolas Montagné[4], Lucie Bastin-Héline[4], Emmanuelle Jacquin-Joly[4], Céline Noirot[5], Christophe Klopp[5] & Bertanne Visser[6]

[1] Evolutionary Ecology and Genetics group, Biodiversity Research Centre, Earth and Life Institute, UCLouvain, Belgium

[2] Instituto Gulbenkian de Ciência, Oeiras, Portugal

[3] CE3C: Centre for Ecology, Evolution, and Environmental Changes, Faculty of Sciences, University of Lisbon, Portugal

[4] Sorbonne Université, INRAE, CNRS, IRD, UPEC, Université de Paris, Institute of Ecology and Environmental Sciences of Paris, Paris and Versailles, France

[5] INRA, Genotoul Bioinfo, Castanet-Tolosan, France

[6] Evolution and Ecophysiology group, Biodiversity Research Centre, Earth and Life Institute, UCLouvain, Belgium

*Corresponding author: caroline.nieberding@uclouvain.be

Author contributions: CMN conceived and designed the research, collected data, analyzed data, and prepared the manuscript; PB designed the transcriptomic analysis and discussed the manuscript; AA, VB, GSM, NM, and LBH collected data; CN, GSM, CK and GL analyzed data; BV collected data, analyzed data, and prepared the manuscript. All authors edited the manuscript.

Acknowledgments: CMN's team was supported by the Fonds National de la Recherche Scientifique (FNRS), FRFC grants 2.4600.10 and 2.4560.11 and CR grant 29109376, as well as UCLouvain ARC grant n◦ 10/15-031 and FSR grant n◦372 605031. BV was supported by FNRS CR grant 24905063. EJJ, NM and LBH were funded by the French National Research Agency (ANR-16-CE02-0003-01 and ANR-16-CE21-0002-01 grants).

Acknowledgments: We would like to thank Prof. Ken-Ichi Moto and Prof. Tetsu Ando for discussions about sex pheromone biosynthesis in Lepidoptera; Dr Jeroen Pijpe for mentoring us about, and providing access to, a Bioanalyzer Systems (Agilent) at the LUMC hospital in Leiden (the Netherlands); Marleen van Eijk for assistance with preparing the butterfly tissues. This is publication BRC321 of the Biodiversity Research Center.

Competing interests: The authors declare no competing interests.

Abbreviations: MSP, male sex pheromone; OR, odorant receptor; OBP, odorant binding protein; PBAN, pheromone biosynthesis activating neuropeptide.

Keywords: Olfactory communication; Desaturase; Reductase; Odorant receptor; Odorant binding protein; Chemosensory protein; PBAN; Phylogeny

Abstract

Unraveling the origin of molecular pathways underlying the evolution of adaptive traits is essential for understanding how new lineages emerge, including the relative contribution of conserved, ancestral traits, and newly evolved, derived traits. Here, we investigated the evolutionary divergence of sex pheromone communication from moths (mostly nocturnal) to butterflies (mostly diurnal) that occurred ~98 million years ago. In moths, females typically emit pheromones to attract male mates, but in butterflies pheromones and used by females for mate choice. The molecular bases of sex pheromone communication are well understood in moths, but have remained virtually unexplored in butterflies. We used a combination of transcriptomics, real time qPCR, and phylogenetics, to identify genes involved in different steps of sex pheromone communication in the butterfly *Bicyclus anynana*. Our results show that the biosynthesis and reception of sex pheromones relies both on moth-specific gene families (reductases) and on more ancestral insect gene families (desaturases, olfactory receptors, odorant binding proteins). Interestingly, *B. anynana* further appears to use what was believed to be the moth-specific neuropeptide Pheromone Biosynthesis Activating Neuropeptide (PBAN) for regulation of sex pheromone production. Altogether, our results suggest that a mosaic pattern best explains how sex pheromone communication evolved in butterflies, with some molecular components derived from moths, and others conserved from more ancient insect ancestors. This is the first large-scale analysis of the genetic pathways underlying sex pheromone communication in a butterfly.

Introduction

Evolution of new life forms occurs through the transition from an ancestral to a descendant clade, where the newly formed clade generally expresses a mosaic of traits conserved from the ancestor, as well as newly evolved, derived, traits. Mosaic evolution is indeed a recurring pattern in paleontology (Cracraft 1970; Gómez-Robles *et al.* 2014; Xu *et al.* 2017). For example, *Tiktaalik roseae,* believed to represent the transition from fishes to amphibians (the "fishapod"; ~375 Mya), shares some traits with more primitive sarcopterygian fishes (e.g. body scales, fin rays, lower jaw and palate) and other traits more typical of tetrapods (e.g. shortened skull roof, a modified ear region, a mobile neck, a functional wrist joint, and other features) (Daeschler *et al.* 2006). Investigating the genetic bases of ancestral and derived phenotypic traits is essential to obtain a mechanistic explanation of how mosaic evolution takes place. Examples of studies investigating the mechanistic basis of mosaic evolution have accumulated in the last decade, including recent genomic evolution analyses identifying patterns of gene loss, retention, and *de novo* evolution (e.g. van Gestel *et al.* 2019; Fernández & Gabaldón 2020; Guijarro-Clarke *et al.* 2020). Other patterns in the genetic bases of trait evolution have suggested a role for hybridization between species (e.g. Berner & Salzburger 2015; Stryjewski & Sorenson 2017; Marques *et al.* 2019) or co-option of molecular pathways that acquired new functions (e.g. Shubin et al. 2009, including a butterfly, Shirai et al. 2012). Derived phenotypic traits can thus be generated through different molecular mechanisms that need to be identified case-by-case.

Here, we focused on the genetic bases of divergence in sex pheromone communication during the evolutionary transition from moths to butterflies, which occurred ~98 Mya. Sex pheromone communication is used by most insects, including butterfly species, for finding, identifying, and assessing the quality of potential mating partners (Myers 1972; Birch *et al.* 1990;

Vane-Wright & Boppre 1993; Nieberding *et al.* 2008; Sarto Monteys *et al.* 2016). Sex pheromone communication is under strong selection, because it determines mating success and consequently an individual's contribution to the next generation (Svensson 1996; Moore *et al.* 1997; Smadja & Butlin 2009; Hansson & Stensmyr 2011; Wyatt 2014). Molecular pathways for sex pheromone biosynthesis, its regulation, and pheromone reception have been identified in several moth species (Tillman *et al.* 1999; Jurenka 2004; Blomquist *et al.* 2012; Leal 2013; Yew & Chung 2015; Zhang *et al.* 2015; Rafaeli 2016). Compared to other insects, moths appear to have evolved moth-specific genes or gene lineages involved in sex pheromone communication (Jurenka 2004; Pelosi *et al.* 2006; Vieira *et al.* 2007; Leal 2013; de Fouchier *et al.* 2014; Helmkampf *et al.* 2015; Yew & Chung 2015; Bastin-Héline *et al.* 2019). For biosynthesis, most moths use a limited number of enzymes for desaturation, chain-shortening, reduction, acetylation or oxidation of *de novo* synthesized saturated fatty acids (Jurenka 2004; Liénard *et al.* 2014; Helmkampf *et al.* 2015; Yew & Chung 2015), which generate tremendous chemical diversity of pheromone components and high species-specificity of pheromone blends (Tillman *et al.* 1999; Jurenka 2004). Some of these enzyme families comprise Lepidoptera-specific subclades, such as the Δ9- and Δ11-desaturases (Jurenka 2004; Yew & Chung 2015; Helmkampf et al 2014). Similarly, novel and derived odorant receptor and odorant binding protein subclades have evolved in moths, which bind specifically to sex pheromone chemicals (Pelosi *et al.* 2006; Vieira *et al.* 2007; Leal 2013; de Fouchier *et al.* 2014; Bastin-Héline et al 2019). We aimed to investigate whether butterflies use moth-specific molecular pathways, more ancestral insect pathways, or have evolved butterfly-specific molecular pathways for pheromone communication.

Butterfly sex pheromones have some ecological specificities that contrast with those of moths, for which it is used by female moths to signal their location to mating partners in the dark or at dusk. In contrast, butterfly sex pheromones are generally produced by males and the

113 importance of sex pheromone communication in butterflies was less clear due to their diurnal
114 lifestyle. Studies on some butterfly species have, however, revealed that sex pheromone
115 communication is important for determining mating success (e.g. Costanzo and Monteiro 2007;
116 Nieberding et al. 2008) and fuels speciation events (Bacquet *et al.* 2015). Butterfly sex pheromones
117 can convey refined information with regard to the identity or quality of potential mating partners
118 and can be critical for mate choice and species recognition (Andersson 1994; Johansson & Jones
119 2007). We focused on the butterfly *Bicyclus anynana,* whose sex pheromone composition has been
120 previously identified and functionally validated (Nieberding *et al.* 2008, 2012). Moreover,
121 experimental manipulation, including the addition of synthetic sex pheromone perfumes
122 (Nieberding *et al.* 2008, 2012), and artificial induction of "anosmia" (Costanzo & Monteiro 2007;
123 van Bergen *et al.* 2013), confirmed the importance of *B. anynana* sex pheromone for mating
124 success. More than a hundred chemical components have been identified on *B. anynana* adult male
125 and female bodies (Heuskin *et al.* 2014), but the composition of the male sex pheromone ("MSP"
126 hereafter) consists of three main volatile components: (Z)-9-tetradecenol (MSP1), hexadecanal
127 (MSP2) and 6,10,14-trimethylpentadecan-2-ol (MSP3) (Nieberding *et al.* 2008). The identification
128 and functional validation of these three MSPs and their role in *B. anynana* chemical communication
129 set this species apart from other lepidopterans, where the role of specific chemical components
130 often remains elusive in relation to fitness (but see Nieberding PLoS ONE 2008 where specific
131 components were identified). MSP1 and MSP2 are derived from fatty acids and are typically found
132 in the pheromone blends of many moth species (Jurenka 2004), which led us to hypothesize that
133 *B. anynana* uses the same genes as moths for sex pheromone communication.
134 To investigate if butterflies use what were believed to be moth-specific gene families,
135 despite 98 million years since the divergence of butterflies from moths (Wahlberg *et al.* 2013), we
136 used RNA-seq to identify genes that could be involved in *B. anynana* pheromone production,

biosynthesis regulation, and reception. We compared transcript abundance of sex pheromone-related adult tissues (male pheromone producing structures, heads and antennae) with control tissues (female wings and heads), and validated our findings with qPCR. We identified specific candidate genes involved in the different olfactory communication functions and used phylogenetic analyses to identify the molecular origin of those genes. Our results reveal that sex pheromone communication in *B. anynana* evolved through a mosaic of ancestral insect genes, and more derived lepidoptera-specific genes.

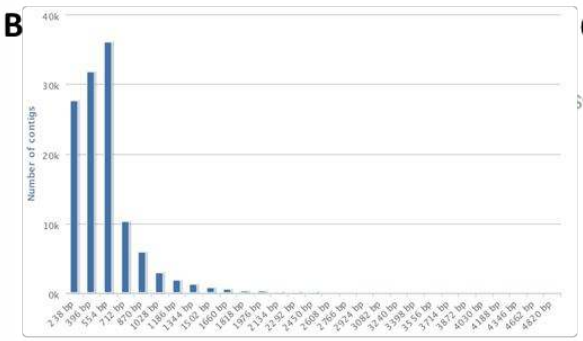
Results and Discussion

Candidate gene identification by transcriptome sequencing

Numerous publications document gene expression studies focusing on chemical communication in Lepidoptera, but only two studied butterflies (Briscoe *et al.* 2013; van Schooten et al 2016), and none of those investigated butterfly sex pheromone communication. Here, we produced six RNA libraries from different adult tissues that were specifically chosen to cover the various steps of male pheromone communication (Figure 1): pheromone biosynthesis (which occurs in dedicated structures on male wings, called androconia; Nieberding *et al.* 2008), its neuro-regulation (in the brain), and pheromone reception (in antennae). Approximately 500 male and female *B. anynana* adults were dissected and relevant tissues assigned to different libraries (Figure 1A). For pheromone synthesis, we compared transcripts in male androconia (Library "androconia") with those in remaining adult male wing parts (Library "male wings") and adult female wings (Library "female wings") as controls. For regulation of pheromone communication, we compared transcript abundance between adult male heads (where the regulation of pheromone synthesis takes place; Library "adult male heads") and adult female heads (Library "adult female heads", control). For

160 pheromone reception, we compared transcripts between adult male and female antennae (the tissue
161 where pheromone reception takes place; Nieberding et al (2008); Library "antennae") and adult
162 heads (Libraries "male heads" and "female heads") as controls. Two other libraries were also
163 analyzed, corresponding to pupal wings in males (Library "pupal male wings"), and females
164 (Library "pupal female wings"), but these data will not be discussed to focus solely on adults, the
165 stage during which pheromone communication takes place. The total of 737,206 Roche 454 reads
166 obtained from the different tissues sampled in *B. anynana* were assembled into 43,149 contigs,
167 with 76,818 remaining non-assembled singletons (Figure 1B and C, Supplementary Table 1). The
168 transcriptome contained 48.6% near-complete sequence information compared to the insect
169 database (S: 43%; D: 5.6%; F:23%; M: 28.4%; BUSCO v4.0.6; Simão *et al.* 2015). The transcripts
170 were annotated based on reference genomes for several butterfly species (including *B. anynana*;
171 Nowell *et al.* 2017), as well as other relevant insect species (see materials and methods section
172 'Transcriptome assembly, quantification, and annotation').

173 Using the digital differential display (DDD) tool (of NCBI's UniGene database; $p < 0.05$),
174 a total of 422 contigs were found to be differentially expressed when tissue-specific libraries were
175 compared (Supplementary Table 1). Expression differences were validated by real time
176 quantitative PCR analyses on 10 selected candidate chemical communication genes, showing that
177 relative differences in expression levels in our transcriptome dataset matched those quantified by
178 RT-qPCR (Supplementary Figure 1).
179



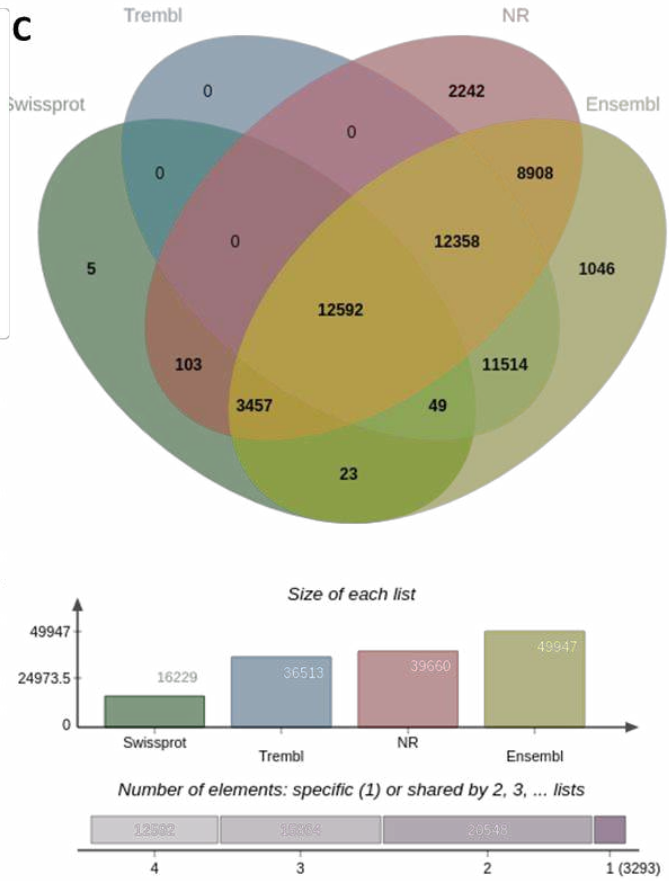


Figure 1: A) Experimental design for the transcriptomics experiment showing the 3 steps involved in male sex pheromone communication and the corresponding tissues sampled to produce the RNA libraries (also including developmental libraries, not shown here). The number of sequenced and cleaned reads per library is provided. B) Information about the number of contigs in the transcriptome. C) Venn diagram of annotated contig with regard to databases: Swissprot, Trembl, NR and 10 species of Ensembl Lepbase.

B. anynana *sex pheromone biosynthesis*

B. anynana was the first butterfly for which sex pheromone biosynthesis molecular pathways were investigated and compared to those of moths (Liénard *et al.* 2014). One gene related to pheromone communication was overexpressed in *B. anynana* male androconial wing tissues compared to male and female control wing samples (i.e., 'type 1' contigs in Supplementary Table 1), an aldose reductase-like gene. This gene was also overexpressed in the male androconial wing tissue alone. Moreover, a Δ9-desaturase gene was also found to be overexpressed in this library. In contrast to earlier findings in *B. anynana* (Liénard et al. 2014), no fatty-acyl reductase (FAR), nor Δ11-desaturase were found overexpressed in male androconial wing tissue (Supplementary Table 1).

Desaturases (that add a double bond to fatty acid substrates) were previously found to be involved in *B. anynana* MSP biosynthesis (Liénard *et al.* 2014). We, therefore, extended our search for desaturase genes for each of the libraries separately. We focused specifically on Δ9 and Δ11-desaturases, because these enzymes play an important role in moth pheromone biosynthesis (Liénard *et al.* 2014; Yew & Chung 2015). Previous work with *B. anynana* suggested that a Δ11-desaturase is involved in the production of MSP1 (Liénard *et al.* 2014). Both Δ9- and Δ11-desaturase were present in the transcriptome, mainly in antennae. A phylogenetic tree containing our Δ9- and Δ11-desaturase contigs further revealed a similar position within the larger desaturase phylogenetic tree, compared to earlier work (Liénard *et al* 2014; Supplementary Figure 2).

To get more insight into the role played by the Δ9 and Δ11 desaturase gene, we used RT-qPCR (cf. protocol in Arun *et al* 2015) to compare transcript abundance between different adult wing tissues, the main tissue producing MSP1 (and using RNA extracted from new samples). Δ9-desaturase transcript abundance was approximately four-fold higher than that of Δ11-desaturase (Welch Two Sample t-test on normalized Ct values; $t = -8.56$, $df = 9.11$, p-value <0.0001; Supplementary Figure 1). When comparing the spatial distribution of MSP1 on *B. anynana* body

parts with our RT-qPCR data for the two Δ-desaturase genes (Supplementary Figure 3A), the expression profile of the Δ9-desaturase gene, but not the Δ11-desaturase gene, matched MSP1 distribution (Supplementary Figure 3B and 3C, respectively). Indeed, the Δ9-desaturase gene showed overall significant variation in transcript abundance across tissues that correlated with the distribution pattern of MSP1 (nested ANOVA, n = 24 samples with 3 biological replicates and 2 technical replicates; $F_{3,6} = 883.5$; p-value < 0.01). Specifically, Δ9-desaturase was found to be significantly overexpressed in male wing parts containing the androconia that produce MSP1, compared to remaining male wing tissues (nested ANOVA; $F_{1,4} = 1{,}814.0$; p-value < 0.01) and female wings (nested ANOVA; $F_{1,4} = 50.5$; p-value < 0.01). Moreover, Δ9-desaturase gene expression was also found to be significantly overexpressed in male head tissue containing MSP1. No such match between gene expression and MSP1 abundance was found for the Δ11-desaturase gene, which showed no significant variation in transcript abundance across tissues known to contain MSP1 (nested ANOVA, $F_{3,6} = 0.07$; p-value < 0.01). Altogether, these findings suggest that a Δ9 desaturase plays a role in *B. anynana* pheromone biosynthesis.

We then searched for genes from a second gene family known to be involved in sex pheromone production in *B. anynana*: fatty acyl reductases (*far*), which convert fatty-acyl pheromone precursors to alcohol (Liénard *et al.* 2014). While more than 20 FARs have been experimentally characterized from 23 moth and butterfly species, all FARs implicated in moth and butterfly sex pheromone biosynthesis are restricted to a single clade, suggesting that one FAR group was exclusively recruited for pheromone biosynthesis (Tupec et al. 2017; 2019 and refs therein). In our transcriptome, two reductase contigs were annotated and identified in male and female antennae: enoyl-CoA reductase and fatty-acyl reductase 1, *far 1*. As *far1* and another fatty-acyl reductase, *far2*, were previously found to be involved in MSP2 and MSP1 biosynthesis, respectively (Liénard *et al.* 2014), we manually mined our transcriptome for *far1* and *far2* contigs

by n-blasting *far1* and *far2* specific gene sequences. Contigs matching *far1* were largely overexpressed in androconia (171 copies), compared to wing controls (0 copies; Supplementary Table 2). While contigs matching *far2* showed an overall low expression level in wing tissues (Supplementary Table 3), a previous qRT-PCR study revealed that *far2* gene expression matched MSP1 biosynthesis (Arun *et al.* 2015), highlighting the potential importance of *far2* for *B. anynana* pheromone production.

The low expression level of *far2* is surprising given the amount of MSP1 present on male wings (2 µg/individual on average); hence we suggest that alternative candidates for MSP1 biosynthesis could be aldo-keto reductases, two of which are among the most expressed genes in androconial male wing tissues (Supplementary Table 1). Indeed, fatty-acyl reductases are usually associated with the reduction of aldehyde into alcohols producing various sex pheromone components in moths (Moto *et al.* 2003; Ando *et al.* 2004), but aldo-keto reductases are regularly found highly expressed in sex pheromone transcriptomes of moth species (Gu *et al.* 2013; Zhang *et al.* 2014). Guo *et al.* (2014) and Yamamoto *et al.* (2016) have proposed that aldo-keto reductases are involved in sex pheromone biosynthesis in the moths *Helicoverpa armigera* and *Bombyx mori* by reducing 9-hexadecenal, 11-hexadecenal and 10E,12Z-hexadecadienal into alcohol. Our expression data suggest that an aldo-keto reductase, with or without *far2*, may be involved in MSP1 biosynthesis.

B. anynana *sex pheromone reception*

The genomes of the butterflies *Danaus plexippus* and *Heliconius melpomene* (i.e., species for which phylogenies of odorant receptor genes were available) have revealed a large number of genes belonging to families known to be involved in olfactory reception in moths, including odorant receptors, and odorant binding proteins (Zhan *et al.* 2011; Dasmahapatra *et al.* 2012; Briscoe *et al.*

2013). Specifically, the odorant receptor (OR) and odorant binding protein (OBP) gene families contain lineages specialized in the detection of sex pheromones in moths, the so-called pheromone receptors (PRs) and pheromone-binding proteins (PBPs) (Sakurai *et al.* 2004; Zhang & Löfstedt 2015; Bastin-Héline et al 2019; Vogt et al 2015). ORs are transmembrane receptors that bind volatile chemicals and are responsible for signal transduction in insect olfactory sensory neurons. They exhibit various response tuning breadths, and moth ORs involved in pheromone detection are often (but not always) highly specific to one or a few pheromone components (Zhang & Löfstedt 2015). We, therefore, expected to identify ORs binding to each of the three known chemical components of the *B. anynana* male sex pheromone: MSP1, 2, and 3 (Nieberding *et al.* 2008). We identified the obligatory co-receptor "Orco" and 16 ORs in the transcriptome, some of which were overexpressed in antennae compared to other adult tissues (Supplementary Table 4). Phylogenetic analysis revealed that ORs expressed in *B. anynana* were distributed among various lepidopteran OR lineages (Figure 2; Dasmahapatra *et al.* 2012), but none were located in the classically defined sex pheromone receptor clade (Bastin-Héline *et al.* 2019; Shen *et al.* 2020; Figure 2). Our results suggest that *B. anynana* sex pheromone reception may have evolved from lepidopteran OR lineages other than the sex pheromone lineage.

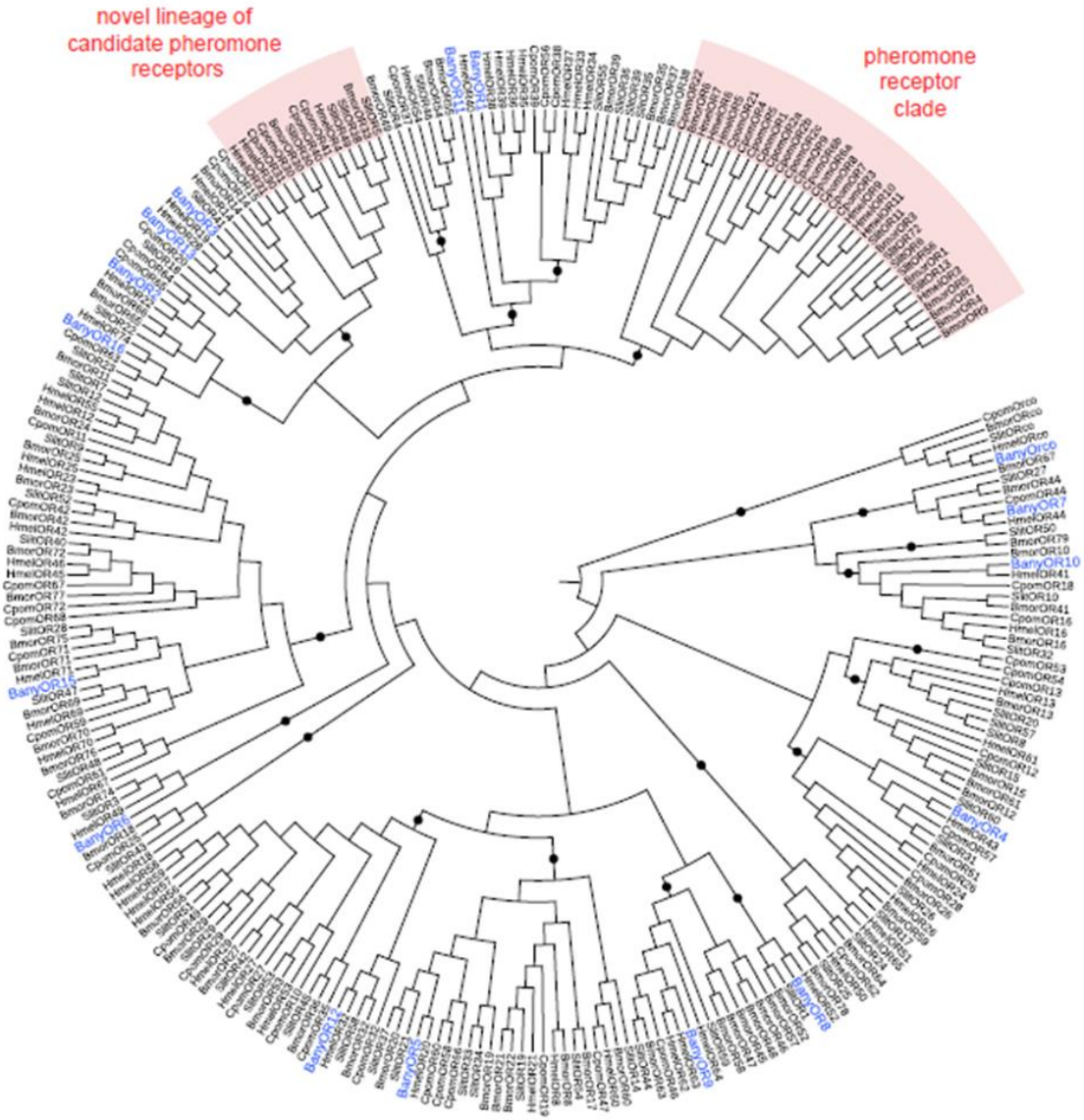


Figure 2: Maximum-likelihood phylogeny of Lepidopteran odorant receptors (OR), including the 16 ORs found in the *B. anynana* transcriptome (BanyOR, in blue). Bmor, *Bombyx mori*; Cpom, *Cydia pomonella*; Hmel, *Heliconius melpomene*; Slit, *Spodoptera littoralis*. Black circles indicate branchings highly supported by the approximate likelihood-ratio test (aLRT > 0.95).

Recent studies have revealed that moth PRs do not constitute a monophyletic clade and, instead, evolved several times during OR evolution (Yuvaraj et al 2017, Bastin-Héline et al 2019). Functional PRs that have been found outside of the PR clade in some moth species were identified based on their sex-biased expression. We, therefore, searched for potential *B. anynana* PRs by quantifying the expression levels between sexes using qPCR, expecting that PR in *B. anynana*

should be more expressed in male than female antennae. We further expected that gene expression levels would correlate with temporally varying physiological and biological needs. In moth species, PRs are critical for detecting the female sex pheromone and the male's behavioral and physiological responses to female sex pheromones were shown to be affected by moth age and mating status (Soques *et al.* 2010, Saveer et al 2012). We, therefore, tried to identify *B. anynana* candidate PRs by comparing RNA expression levels in females with different mating status (using qPCR). We expected that virgin females that had developed either in isolation (naive "virgin") or in the presence of male scent ("virgin sensitized") would exhibit higher expression levels for OR genes responsible for detecting the male sex pheromone, compared to mated females ("mated")(Zeng et al 2013 Biochem Biophys Res Commun; Immonen et al 2012 PRSB). This difference would be due to virgin females taking information about the composition of the male sex pheromone for choosing mates regarding their inbreeding level (van Bergen *et al.* 2013) or their age (Nieberding *et al.* 2012), and because recently mated females are much less receptive to courtship attempts in *B. anynana*. The candidate genes Ban_OR1, Ban_OR2 and Ban_Orco were selected for qPCR experiments because these genes displayed the highest expression among the 16 identified candidate ORs and were significantly overexpressed in antennae compared to control libraries (Supplementary Table 4). Orco expression was significantly decreased in mated compared to virgin (naïve or sensitized) females, but Bany_OR1 or Bany_OR2 were not (Figure 3). This finding suggests that the regulation of the expression of Orco could be a mediator of sex pheromone receptivity. Orco, and not specific parts of the odorant receptor dimer, such as OR1, OR2 or other odorant receptors that we did not test here, could be regulated by sex pheromone communication, similar to what was previously found in cockroaches (Soques *et al.* 2010; Latorre-Estivalis *et al.* 2015).

In addition to the work described above, we tried to functionally investigate if some specific OR candidate genes were responsible for the detection of male pheromone components using heterologous expression in *D. melanogaster* olfactory sensory neurons coupled to electrophysiological recordings. These experiments did not lead to functional validation, but the procedures followed and results obtained are described in Supplementary File 1.

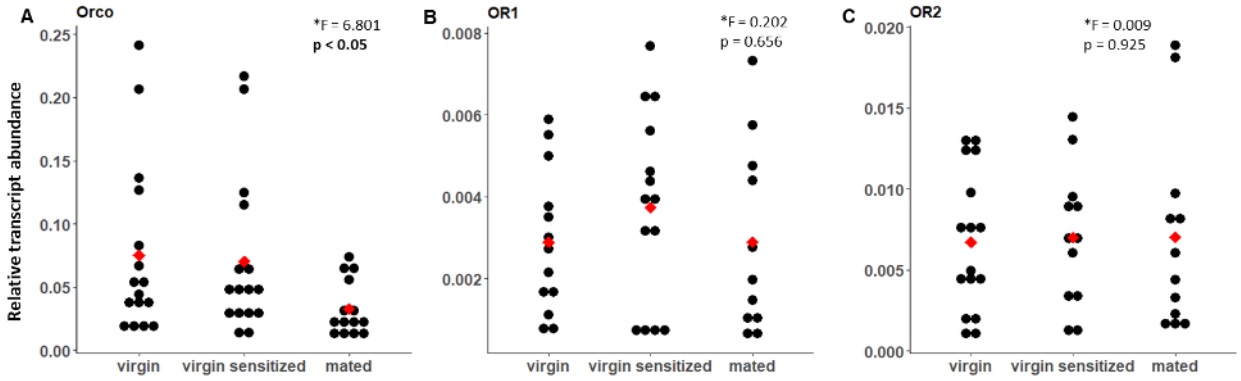


Figure 3: qPCR expression level of odorant receptors (OR) in antennae of female *B. anynana* with different mating status. Orco (A), but not OR1 (B) or OR2 (C), RNA expression level differed significantly in virgin naïve (left) and virgin sensitized (middle) compared to mated (right) females. Each treatment is the mean of 3 to 7 biological replicates. A nested ANOVA was used to test for differences between groups. F and p values are included for each graph. *log transformed data.

A second gene family specific to insects, the OBP family, is involved in olfaction by solubilizing semiochemicals once they have entered the aqueous lymph within olfactory sensilla (Leal et al 2013). OBPs have been proposed to play an important role in response sensitivity. In Lepidoptera, a dedicated lineage of OBPs (the so-called "pheromone-binding proteins" or PBPs) has evolved high affinity towards pheromone components (Vogt *et al.* 2015). We identified 46 contigs assembled into 13 OBP unigenes expressed in our *B. anynana* transcriptome (Supplementary Table 5), a number lower than what has been described in various transcriptomes from moth species (49 predicted OBPs in *Spodoptera littoralis* and *Manduca sexta;* Walker et al 2019) and in the genomes of two butterfly species (32 in *Danaus plexippus*, 51 in *Heliconius Melpomene*; Pelosi *et al.* 2014, 2018; Venthur & Zhou 2018). *B. anynana* expressed OBPs were

found in most subclades of the phylogenetic tree of lepidopteran OBPs, including general odorant binding protein 1 and 2 lineages, as well as classic, minus-C, plus-C and duplex OBP lineages (with categories based on the level of sequence homology and conserved amino acid signatures; Supplementary Figure 4). In Lepidoptera, the OBP gene family also includes a lineage of the PBPs, thought to transport pheromone molecules (Vogt *et al.* 2015). In moths, such as *Manduca sexta* and *Bombyx mori*, trichoid sensilla are associated with pheromone perception and express specifically PBP-A. No *B. anynana* expressed OBP clustered in the pheromone-binding protein-A or -B lineages (Supplementary Figure 4). This is similar to what has been found in other butterfly species: the PBP-A lineage is lacking in the genome of *Danaus plexippus* and the PBP-A and PBP-B lineages are absent from the genomes of *Heliconius melpomene* and *Melitaea cinxia* (Vogt *et al.* 2015). In contrast, we did find two candidate PBPs (Supplementary Table 5), expressed in *B. anynana* antennae, which belong to the pheromone-binding protein-C and -D lineages present in all butterfly genomes investigated to date (Vogt *et al.* 2015). These candidate PBPs indeed correspond to the two sole candidate PBP genes identified in the *B. anynana* genome, and are both most similar to two PBPs found in the antennae of *H. melpomene* (Dasmahapatra et al. 2012; Supplementary Figure 4). In most moths, PBP-C and PBP-D OBPs are expressed in basiconic sensilla and are associated with foraging (Vogt *et al.* 2015). Although we cannot exclude that we missed BanOBPs in our transcriptome, the lack of a PBP-A subgene family in *B. anynana,* as in the four other butterflies studied to date (*H. melpomene, D. plexippus, M. cinxia, P. polytes*), suggests that butterflies lost this gene lineage (at least in the Nymphalidae, the family of butterflies to which the sampled species belong), and possibly use other PBP lineages to functionally aid the OR-pheromone connection. The transcriptome was further mined for Chemosensory Proteins (CSPs), a third gene family potentially implicated in olfaction in insects (Sánchez-Gracia *et al.* 2009; Vieira & Rozas 2011), for which the results are available in Supplementary Table 6.

341

342 **B. anynana *sex pheromone regulation***

343 Eleven contigs were found overexpressed in male compared to female brains, but their role in the
344 regulation of sex pheromone processing remains open (Supplementary Table 1). Given its role as
345 a key regulator of female sex pheromone biosynthesis in many moth species (Bloch *et al.* 2013),
346 we focused our attention on Pheromone Biosynthesis Activating Neuropeptide (PBAN). We
347 hypothesized that PBAN could be involved in male sex pheromone regulation in *B. anynana*, and
348 looked for it in our transcriptome database. We identified one unigene annotated as PBAN
349 (BA_PBAN.1.1), which was expressed in adult heads. We used this sequence to obtain the
350 complete cDNA sequence of PBAN in *B. anynana* (RACE), Ban_PBAN (Figure 4A). The
351 phylogenetic reconstruction of PBAN across Lepidoptera shows monophyly of butterfly PBANs
352 (data not shown). We next investigated the PBAN cDNA tissue distribution using semi-quantitative
353 and quantitative PCR. PBAN was found to be expressed in adult heads, but not in other tissues,
354 and more strongly in males than in females (Figure 4B). PBAN in male moths is suspected to be
355 involved in male pheromone biosynthesis: the PBAN receptor of the moth *Helicoverpa armigera*
356 was found expressed in male hairpencils, and PBAN stimulation of the hairpencils was found to be
357 responsible for the production and release of male pheromonal components (Bober & Rafaeli
358 2010). Next, we used RT-qPCR and found that PBAN expression level in male brains correlated
359 with the amount of male sex pheromone found on male wings during the adult male's lifetime,
360 with maximum content around 15 days (Figure 4C; Nieberding *et al.* 2012).

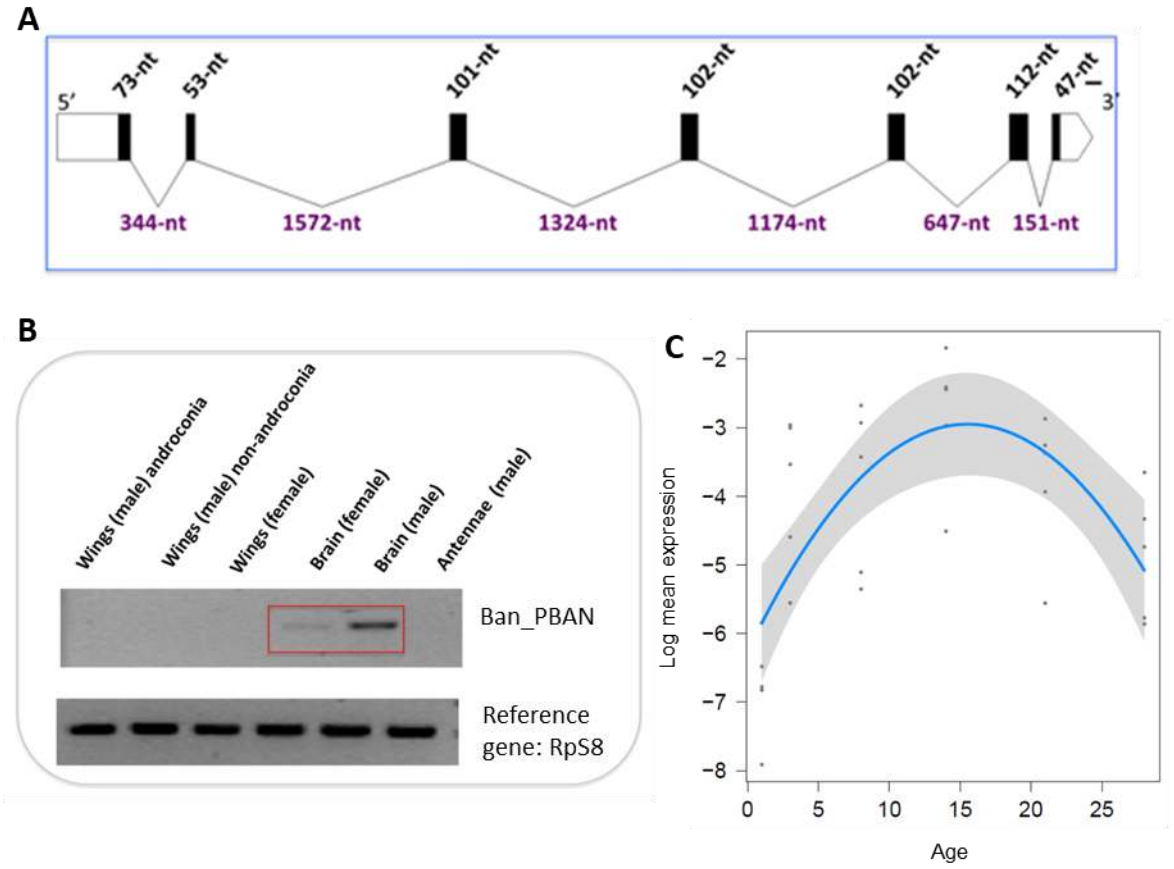


Figure 4: Pheromone binding activating neuropeptide (PBAN) expression in *B. anynana*. A) Structure of Ban_PBAN full length gene sequence (594 bp). The 7 exons (encoding 198 amino acids) are represented by black boxes and the 6 introns by lines. B) PBAN expression level quantified by semi-quantitative PCR in adult tissues (brains, wings, antennae) of males and females ranging from 3 to 14 days of age. Higher levels of PBAN are observed in male brains compared to the other adult tissues. C) PBAN expression level quantified by real time qPCR in adult male brains from 1 to 28 days of age.

In moths, production of volatile sex pheromones usually shows a circadian pattern that is regulated by PBAN and correlates with the female "calling" behavior (extrusion of the sex pheromone gland) during specific hours of the scotophase (Rosén 2002; Bloch, Hazan, and Rafaeli 2013). A circadian rhythm of male sex pheromone production was also found in the moth *Aphomia sabella* (Levi-Zada *et al.* 2014). We tested whether *B. anynana* displayed daily variation in courtship activity, MSP production, and PBAN expression in 8-day old adult males. We found that courtship activity peaked 7 to 12 hours after sunrise, and courtship activity was significantly higher in the afternoon compared to the rest of the day (Figure 5A; likelihood ratio test on the "Time^2"

effect, p= 0.005; Supplementary Table 7 and 8). Similarly, MSP production significantly varied during the course of the day, and peaked around maximum courtship activity, with MSP1/MSP2 and MSP2/MSP3 ratios displaying significant reversed changes during the day (Figure 5D; likelihood ratio test on the "Time² "p= 0.004 and p = 0.002, respectively; Supplementary Table 8). MSP amounts also displayed a slight, but non-significant, variation with time of the day (Figure 5C; likelihood ratio test on the "Time²" effect, $0.007 < p < 0.09$; Supplementary Table 8). MSP titers were estimated to be minimum around 11 hours after sunrise for MSP1 and MSP3, while the MSP2 titer was estimated to be at a maximum 12.4 hours after sunrise. We further found that PBAN expression significantly varied throughout the day as well (Figure 5B), with the highest expression 11 to 14 hours after sunrise. Daily variation in PBAN expression thus correlates both to male courtship activity and to male sex pheromone quantities found on male wings: all three traits peak during the afternoon and PBAN expression is maximal just before the peak in MSP2/MSP1 and MSP2/MSP3 ratios and MSP2 amount. This suggests that the daily regulation of male sex pheromone may be associated to circadian variation in PBAN expression, a neuropeptide that is specific to sex pheromone regulation in moths (Cheng *et al.* 2010).

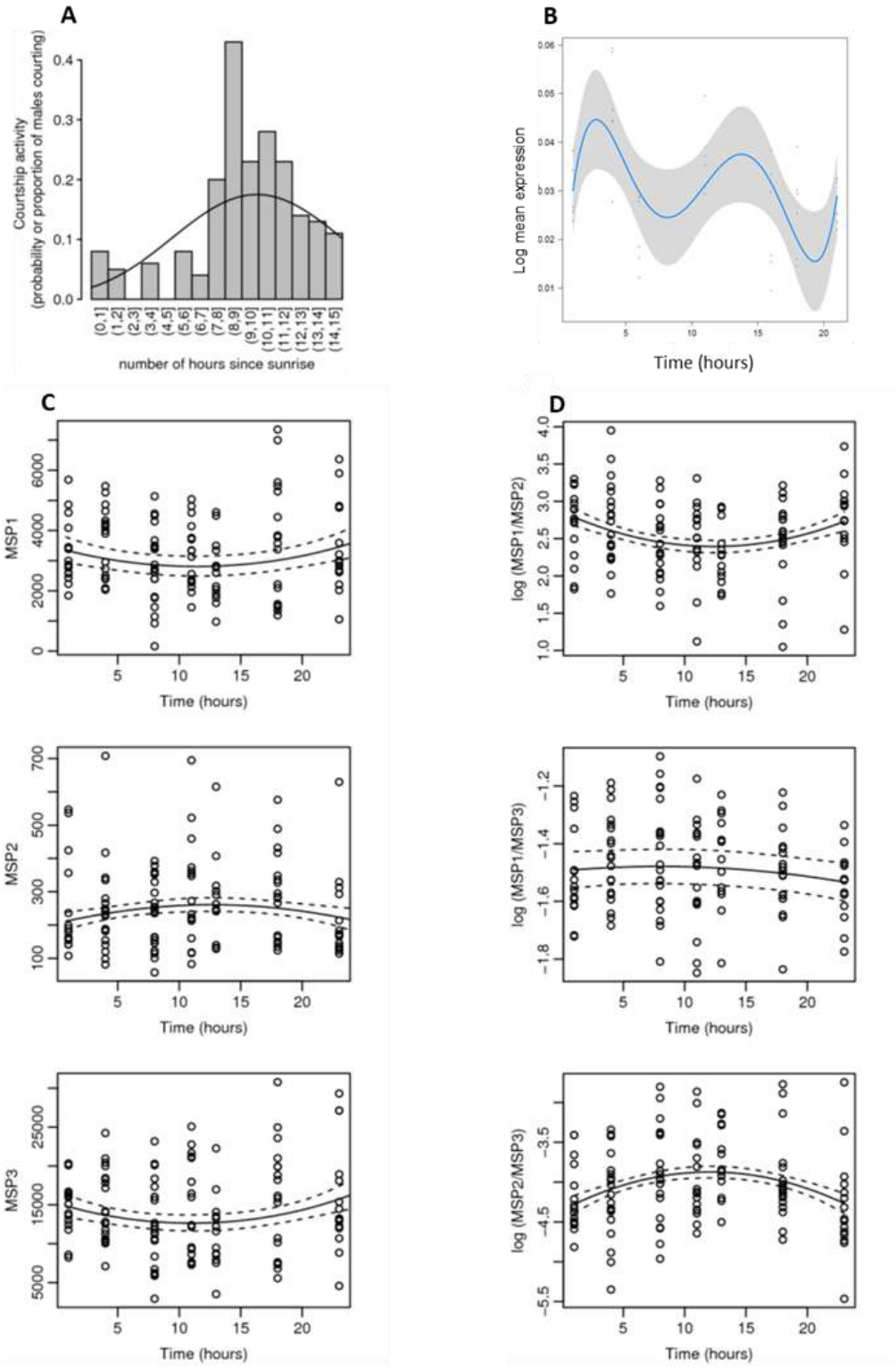


Figure 5: Daily variation in *B. anynana* male courtship activity (A), PBAN expression by qPCR (B), MSP production (C) and log MSP ratio production (D) .

386

In addition to the work described above, we aimed to functionally demonstrate the role of PBAN expression in regulating male sex pheromone biosynthesis in *B. anynana*. These experiments did not lead to functional validation of the role of PBAN, but all procedures followed and results obtained are described in Supplementary File 2.

Conclusions and perspectives

This is the first time a transcriptomics approach has been used to gain insight into the different aspects of sex pheromone communication in a butterfly. Our model, *B. anynana,* is the only butterfly species for which the sex pheromone composition and its role in male mating success in relation to different fitness-associated traits (e.g., age and inbreeding level) have been demonstrated (Costanzo & Monteiro 2007; Nieberding *et al.* 2008, 2012; van Bergen *et al.* 2013). The male sex pheromone composition and function in communication between sexes has been partially analyzed in only a few additional butterfly species, including *Pieris napi*, *Colias eurytheme*, *Danaus gilippus*, *Idea leuconoe*, and more recently *Heliconius melpomene* (Pliski & Eisner 1969; Taylor 1973; Grula *et al.* 1980; Sappington & Taylor 1990a, b; Nishida *et al.* 1996; Andersson *et al.* 2007; Darragh *et al.* 2017).

Mosaic evolution appears to have taken place at the molecular level based on our investigation of the pathways involved in the production, reception, and regulation of the sex pheromone in *B. anynana*: our data suggest that the biosynthesis of the three chemical components forming the male sex pheromone could be partly due to moth-specific genes (*far1* and *far2* for the MSP2 and MSP1 components, respectively) and partly due to genes present in insects other than moths (Δ9-desaturase, aldo-keto reductase for the MSP1 component). This is also likely the case for the MSP3 component whose synthesis is not expected to rely on moth-specific gene families,

as this pheromone component is not derived from fatty acids. None of the expressed odorant receptors or odorant binding proteins in *B. anynana* belonged to Lepidoptera-specific gene lineages responsible for sex pheromone reception in moths, suggesting that sex pheromone reception in this butterfly may have evolved independently from their moth ancestors. In contrast, we found that sex pheromone biosynthesis could be regulated by the neuropeptide PBAN in both moths and butterflies, an evolutionarily shared derived trait for Lepidoptera.

Recently, the genomes of 250 species of skippers (Hesperiidae; Li et al. 2019) and 845 North American butterfly species (Zhang *et al.* 2019) have been sequenced. A systematic comparative analysis of major gene families involved in moth sex pheromone communication in these ~1000 butterfly genomes would provide important information on the level of conservation of molecular pathways when butterflies diverged from moths about 98 million years ago.

Materials and methods

Insects

Bicyclus anynana (Butler, 1879) (Lepidoptera: Nymphalidae) originated from an outbred wild type population that was collected in Malawi, Africa, in 1988 using 80 gravid females. Since 1988, several hundred individuals have been used each generation to maintain high levels of heterozygosity (Brakefield *et al.* 2001) in a climate-controlled room at a temperature of 27ºC, a relative humidity of 70%, and a photoperiod of L:D 12:12. Larvae were kept under these conditions on maize plants *Zea mays*, and adults fed mashed banana *Musa acuminate* for all experiments, except when stated otherwise.

Transcriptome

Tissue collection

For the transcriptome dataset, several hundreds of virgin males and females were separated at the pupal stage in different cages, and tissues collected in March 2010. Pupal tissues were collected from male and female pupae 1 to 7 days after pupation (1 or 2 pupae per day after pupation and per sex); the wing imaginal discs were dissected as described in (Brakefield *et al.* 2009) . Tissues for adult libraries (wings, heads and antennae) were collected from adult virgin males and females aged 1, 3, 5, 8, 10 and 14 days after emergence (Supplementary Figure 5): ~50 adults were used per age category and per library for wing libraries, ~10 adults were used per age category and per library for head tissues, and ~5 adult females and 5 males were used per age category for the antennae library. Brain tissue was obtained by cutting the head and cutting off the eyes, the proboscis and the antennae; antennal tissue was collected for a similar number of adult males and females (Figure 1). Dissected tissues were conserved immediately at -20°C in RNAlater (Sigma-Aldrich).

RNA extraction

RNAs of all dissections were extracted in April 2010, within a month after collection of tissues, in an RNA-free environment, on ice, and using the RNeasy Mini kit and the RNAase free DNAase kit (Qiagen). After RNA extraction, 1 µl of each RNA extract was used for testing RNA quality and quantity using a Bioanalyzer Systems (Agilent) at the LUMC hospital in Leiden (the Netherlands, courtesy of Dr Jeroen Pijpe), and using the RiboGreen RNA quantification kit, respectively. The remaining RNA extract was stored at -80°C for cDNA synthesis. For cDNA synthesis, we first pooled all RNA extracts dedicated to the same library in one tube per library, in such a way that: *i*) the same amount of RNA was present for each sex (male and female), *ii*) each

life stage was represented by similar RNA amounts (days 1 to 7 after pupation for pupal tissue libraries, days 1 to 14 after emergence for adult tissue libraries; Supplementary Table 9). Total RNA yield was 27 to 40 µg per library as requested for sequencing.

mRNA isolation, cDNA synthesis and sequencing

mRNA capture, cDNA synthesis and tagging for Titanium 454-sequencing was performed by Biogenomics, a KU Leuven Research & Development Division of the Laboratory of Animal Diversity and Systematics (Leuven, Belgium). Between 370 and 1,340 ng (0.3 and 1.6%) mRNA yield was obtained for each library, providing enough mRNA (minimum 200 ng per library) for cDNA construction and tagging. Yet, cDNA synthesis failed when started from mRNA, which is why a SMART cDNA synthesis was performed from total RNA. A custom normalization step (based on the EVROGEN Trimmer kit) was optimized in collaboration with the Roche R&D department and applied to the cDNA libraries, as no validated normalization protocol was available from Roche in 2010 for Titanium cDNA sequencing. Each normalized library was quality checked for fragment length and integrity before sequencing. Each library was subjected to GS FLX Titanium Emulsion PCR and Sequencing, and each library was sequenced 5 times. After sequencing, data were processed through certified Roche software (GS Transcriptome Assembler/Mapper) and custom scripts for advanced analysis. Basic data analysis included read quality trimming and assembly into contigs, including potential alternative splicing products. The sequences were trimmed by removing low quality sequences, ambiguous nucleotides, adapter sequences, and sequences with lengths less than 20 nucleotides. The 454-sequencing generated 824,439 reads, with an average length of 293 base pairs and a total of 242,005,027 nucleotides (Supplementary Figure 6).

Transcriptome assembly, quantification, and annotation

Adaptors were removed with smartkitCleaner and adaptorCleaner. Raw sequences (reads) were cleaned with the software Pyrocleaner (Mariette et al. 2011[2]), using the following criteria: (1) complexity/length ratio less than 40 (using a sliding window approach based on 100 bp sequence length, and a step of 5 bp); (2) duplicate read removal (see bias associated with pyrosequencing, due to the random generation of duplicate reads); (3) removal of too long/too short reads (maximum and minimum read length = mean read length ± 2 SD); (4) removal of reads with too many undetermined bases (more than 4 %). Contaminations were discarded by searching hits against *E. coli*, phage and yeasts.

The reads were assembled *de novo* in 43,149 contigs of 488 base pairs on average with a total of 21,087,824 nucleotides (Supplementary Figure 6). The average GC content was 36.44%. The assembly was perform with tgic l (https://academic.oup.com/bioinformatics/article/19/5/651/239299) version 2.1 using standard parameters. The reads where realigned to the contigs and singletons with bwa aln version 0.7.2 using standard parameters and transformed in bam format, sorted and indexed with samtools version 0.1.19 with default parameters The bam files were then processed with samtools idxstats to extract expression measures in the form of numbers of reads aligned on each contig for every condition. These measures were than merged to produce the quantification file using unix cut and paste commands. Diamond was used to search for sequence homology between contig and the following generalist databases: UniProtKB/Swiss-Prot, UniProtKB/TrEMBLRelease of April, NR release of end of March.

The following species from the ensemble database were queried: *Bicyclus anynana* (nBA.0.1), *Calycopis cecrops* (v1.1), *Danaus plexippus* (v3), *Heliconius melpomene melpomene*

(Hmel2.5), *Junonia coenia* (JC v1.0), *Lerema accius* (v1.1), *Melitaea cinxia*, *Papilio machaon* (Pap_ma_1.0), *Phoebis sennae* (v1.1) and *Pieris napi* (v1.1).

Identification of specific gene families

We also mined the transcriptome for specific families of genes supposedly involved in sex pheromone communication based on the available evidence in moths (Lepidoptera): desaturases, reductases, odorant receptors (OR), odorant binding proteins (OBP), and chemosensory proteins (CSP). For this:

i) we downloaded the DNA sequence of every *B. anynana* contig named as a desaturase, reductase, OR, OBP or CSP in our transcriptome;

ii) we checked the homology of the sequence of each candidate contig with gene members of the same family in other Lepidoptera by performing a blastx in Genbank;

iii) every *B. anynana* contig that showed significant homology in step *ii* was blasted in the transcriptome, allowing us to find more *B. anynana* ESTs of the same gene family, even if some had not been annotated as such. All these contigs and ESTs were then "candidate members of each respective gene family". If no significant homology was found using blastx in step *ii*, the sequence was removed from the list of candidate members of the gene family;

iv) every *B. anynana* contig and EST candidate was then translated into an amino acid sequence using Expasy (https://web.expasy.org/translate/). When necessary, cdd analyses of domains were done. Using this procedure, 27 OR, 44 OBP and 70 CSP candidate members were found in the *B. anynana* transcriptome (Supplementary Tables 4, 7 and 8 for OR, OBP and CSP, respectively; for reductases and desaturases see Results). For example, 40 contigs were initially annotated as "odorant-binding protein" in our transcriptome, based on the characteristic hallmarks of the OBP protein families, including six highly conserved cysteines, i.e., "C" (in Lepidoptera C1-X25-30-

C2-X3-C3-X36-42-C4-X8-14-C5-X8-C6, with "X" being any amino acid; Xu *et al.* 2009). As sequence conservation between OBPs is low, i.e., between 25 and 50% identity for amino acid sequences, manually mining the transcriptome allowed us to find another seven OBP candidate members (Supplementary Tables 4, 7, and 8 for OR, OBP and CSP, respectively).

v) Candidate members were then manually aligned in Bioedit to group them into distinct expressed gene units, or unigenes: 17 Bany_OR unigenes, 9 Bany_OBP unigenes, including in some cases more "gene subunits" when contigs were similar enough to suggest that they represented different allelic variants of the same gene, such as Bany_OBP3, Bany_OBP4, Bany_OBP6 (Supplementary Table 5) and 8 Bany_CSP unigenes with some more gene subunits as well (Supplementary Table 6).

vi) The expression level of each candidate unigene across libraries was then obtained by pooling the number of copies in the *B. anynana* transcriptome of each EST and contig forming the unigene.

Phylogenies

For the OR phylogeny, amino acid sequences found in the *B. anynana* transcriptome were aligned with OR sequences previously identified in the genomes of *B. mori* and *H. melpomene* (Dasmahapatra *et al.* 2012) and in antennal transcriptomes of *C. pomonella* (Walker *et al.* 2016) and *S. littoralis* (Walker *et al.* 2019). Alignment was performed with MAFFT v7 (https://mafft.cbrc.jp/alignment/server/), and the maximum-likelihood phylogeny was built using PhyML 3.0 (Guindon *et al.* 2010). Branch support was assessed using a likelihood-ratio test (Anisimova & Gascuel 2006). Published datasets of Lepidoptera protein sequences from previous phylogenetic studies were used for testing the phylogenetic position of *B. anynana* ORs (Poivet *et al.* 2013; Supplementary File 3), OBPs (Vogt *et al.* 2015; Supplementary File 4), desaturases and reductases (Liénard *et al.* 2014; Supplementary Files 5 and 6).

551

552 ***Quantitative real time PCR***

553 Biological replicates were the mRNA extracted from either a single individual or formed by
554 pooling 3 to 5 individuals of various ages in experiments for the "reception" and the "production"
555 communication steps, respectively. Each treatment is represented by 3 to 7 biological replicates.
556 The protocol used for quantitative real time PCR is described in Arun *et al.* (2015). Briefly, total
557 RNA was extracted using the RNeasy Mini kit following manufacturer's instructions. Residual
558 DNA was removed after treating extracted RNA using a DNase enzyme. A nanodrop ND-1000
559 spectrophotometer was then used to assess the integrity of the RNA before conversion into cDNA.
560 qRT-PCR was carried out using the SYBR green dye in a 96-well thermocycler with parameters
561 described in Arun *et al.* (2015). Primer sequences for all genes are available in Supplementary
562 Table 10. Relative transcript abundance was calculated using the 2(-ΔCt) method. Statistical
563 significance of differences in expression levels expressed in Rq values after log-transformation was
564 tested using nested ANOVA with technical replicates nested with biological replicates; the model
565 was log (Rq) ~ treatment / biological replicate / technical replicate + Error (tissue / biological
566 replicate / technical replicate). R version 3.6.1 (R Core Team 2020) was used for statistical
567 analyses.

568

569 ***Behavioral experiments***

570 **Mating experiments for quantifying odorant receptor expression levels**

571 Naïve virgin females were reared in isolated conditions (devoid of the male secondary sexual traits
572 putatively involved in sexual communication, i.e., olfaction, vision and audition) directly after egg
573 collection. The virgin sensitized females were reared in a MSP-containing environment near cages

574 containing males. The sensitized mated females were reared in a MSP-containing environment and
575 mated at an age of 3 days. All females were sacrificed at day 5 and the antennal tissues were used
576 for RNA extraction and RT-qPCR analysis (as described above).
577
578 **Daily variation in courtship activity**
579 We tested whether courtship activity in *B. anynana* males varies throughout the day. A large
580 number of individuals were reared and age after emergence was recorded. The day before the
581 experiment, five males and four females between 2 and 12 days old were randomly chosen and
582 grouped in a cage (40cm x 30cm, cylindrical). The cages were placed in a room with a temperature
583 of ~27°C with natural light, and a 14:10 day-night regime. The butterflies were fed with banana
584 slices and had access to water during the course of the experiment. We used five cages per trial and
585 produced three trials with different individuals. A generalized mixed model with binomial error
586 distribution was used to characterize the courting activity of males during the day. The
587 presence/absence of courtship behavior for each male during 15 minutes of observations per hour
588 was used as the dependent variable. As we expected courtship activity to peak at some time point
589 during the day, we used "time of the day" (in the number of hours after natural sunrise) and its
590 second order polynomial as a fixed explanatory variable. The age of males (in days) was also
591 included as a fixed cofactor to control for the effect of age. The identity of each individual, cage
592 and trial were used as random effects with individual nested within cage and cage nested within
593 trial. We tested the model parameters with type III likelihood ratio tests, in which a model without
594 the explanatory variable of interest is compared to the full model, both models being estimated by
595 Maximum Likelihood.
596

Daily variation in male sex pheromone production

A number of butterfly couples were set up using adult virgin stock males and females. Three families were started from three couples that produced over 200 offspring. The three families were each partly reared into two different climate rooms that differed in the onset of artificial daylight (one at 9:00 am, the other at 12:00 pm). This allowed us to control for the potential effect climate cell-specific conditions on variation in MSP production. Forty to 80 males that emerged the same day were selected per family. MSP production of 8-day old males was sampled, an age at which each MSP component is produced in measurable quantities (Nieberding *et al.* 2008). Four to 7 males of each family were killed and conserved at -80° for subsequent pheromone analysis at seven sampling points during the day: 1, 4, 8, 11, 13, 18 and 23 hours after the onset of daylight. MSP production was measured as described below in the section "male sex pheromone quantification". We used mixed models with normal error distribution to characterize the variation of MSP production during the day. The titre of each MSP component and the ratios between pairs of MSP components were used as dependent variables. MSP titres were square root transformed and MSP ratios were log-transformed to improve the normality and homoscedasticity of the residuals. As we suspected MSP production to peak at some time point during the day, we used a second order polynomial equation with time and time^2 as a fixed explanatory variables and family as a random effect. We tested model parameters with type III likelihood ratio tests, in which a model without the explanatory variable of interest is compared to the full model, both models being estimated by Maximum Likelihood. We estimated the percentage of variation explained by the models and each of their components with pseudo R^2 based on ratios of sums of squared residuals. We followed Legendre & Legendre (1998) for the variance decomposition procedure.

Quantification of male sex pheromone production

MSP concentrations were determined as previously described (Nieberding *et al.* 2008, 2018). In short, one forewing and hindwing of each male was soaked in 600ul of hexane during 5 minutes. One ng/ul of the internal standard (palmitic acid) was then added. Extracts were analyzed on a Hewlett-Packard 6890 series II gas chromatograph (GC) equipped with a flame-ionization detector and interfaced with a HP-6890 series integrator, using nitrogen as carrier gas. The injector temperature was set at 240 °C and the detector temperature at 250 °C. A HP-1 column was used and temperature increased from the initial temperature of 50 °C by 15 °C/min up to a final temperature of 295 °C, which was maintained for 6 min.

References

Andersson, J., Borg-Karlson, A.-K., Vongvanich, N. & Wiklund, C. (2007). Male sex pheromone release and female mate choice in a butterfly. *J. Exp. Biol.*, 210, 964–970.

Andersson, M. (1994). *Sexual selection.* Princeton University Press.

Ando, T., Inomata, S. & Yamamoto, M. (2004). Lepidopteran Sex Pheromones. In: *The Chemistry of Pheromones and Other Semiochemicals I. Topics in Current Chemistry* (ed. Schulz, S.). Springer, Berlin, Heidelberg.

Anisimova, M. & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.*, 55, 539–552.

Arun, A., Baumlé, V., Amelot, G. & Nieberding, C.M. (2015). Selection and validation of reference genes for qRT-PCR expression analysis of candidate genes involved in olfactory communication in the butterfly *Bicyclus anynana*. *PLoS One*, 10, 1–17.

Bacquet, P.M.B., Brattström, O., Wang, H.L., Allen, C.E., Löfstedt, C., Brakefield, P.M., *et al.* (2015). Selection on male sex pheromone composition contributes to butterfly reproductive isolation. *Proc.*

644 *R. Soc. B Biol. Sci.*, 282, 20142734.

645 Bastin-Héline, L., de Fouchier, A., Cao, S., Koutroumpa, F., Caballero-Vidal, G., Robakiewicz, S., *et al.*

646 (2019). A novel lineage of candidate pheromone receptors for sex communication in moths. *Elife*, 8,

647 1–17.

648 Bengtsson, J.M., Gonzalez, F., Cattaneo, A.M., Montagné, N., Walker, W.B., Bengtsson, M., *et al.* (2014).

649 A predicted sex pheromone receptor of codling moth *Cydia pomonella* detects the plant volatile

650 pear ester. *Front. Ecol. Evol.*, 2, 1–11.

651 van Bergen, E., Brakefield, P.M., Heuskin, S., Zwaan, B.J. & Nieberding, C.M. (2013). The scent of

652 inbreeding: A male sex pheromone betrays inbred males. *Proc. R. Soc. B Biol. Sci.*

653 Berner, D. & Salzburger, W. (2015). The genomics of organismal diversification illuminated by adaptive

654 radiations. *Trends Genet.*, 31, 491–499.

655 Birch, M.C., Poppy, G.M. & Baker, T.C. (1990). Scents and eversible scent structures of male moths. *Annu.*

656 *Rev. Entomol.*, 35, 25–58.

657 Bloch, G., Hazan, E. & Rafaeli, A. (2013). Circadian rhythms and endocrine functions in adult insects. *J.*

658 *Insect Physiol.*, 59, 56–69.

659 Blomquist, G., Jurenka, R., Schal, C. & Tittiger, C. (2012). 12 - Pheromone Production: Biochemistry and

660 Molecular Biology. In: *Insect Endocrinology* (ed. Gilbert, L.). Academic Press, pp. 523–567.

661 Bober, R. & Rafaeli, A. (2010). Gene-silencing reveals the functional significance of pheromone

662 biosynthesis activating neuropeptide receptor (PBAN-R) in a male moth. *Proc. Natl. Acad. Sci. U. S.*

663 *A.*, 107, 16858–16862.

664 Brakefield, P., Beldade, P. & Zwaan, B. (2009). Dissection of larval and pupal wings from the African

665 butterfly *Bicyclus anynana*. *Cold Spring Harb. Protoc.*

666 Brakefield, P.M., Filali, E. El, Van Der Laan, R., Brueker, C.J., Saccheri, I.J. & Zwaan, B. (2001). Effective

667 population size, reproductive success and sperm precedence in the butterfly, *Bicyclus anynana*, in

668 captivity. *J. Evol. Biol.*, 14, 148–156.

669 Briscoe, A.D., Macias-Muñoz, A., Kozak, K.M., Walters, J.R., Yuan, F., Jamie, G.A., *et al.* (2013). Female

670 behaviour drives expression and evolution of gustatory receptors in butterflies. *PLoS Genet.*, 9,

671 e1003620.

672 Cheng, Y., Luo, L., Jiang, X., Zhang, L. & Niu, C. (2010). Expression of pheromone biosynthesis activating

673 neuropeptide and its receptor (PBANR) mRNA in adult female *Spodoptera exigua* (Lepidoptera:

674 Noctuidae). *Arch. Insect Biochem. Physiol.*, 75, 13–27.

675 Costanzo, K. & Monteiro, A. (2007). The use of chemical and visual cues in female choice in the butterfly

676 Bicyclus anynana. *Proc. R. Soc. B Biol. Sci.*, 274, 845–851.

677 Cracraft, J. (1970). Mandible of archaeopteryx provides an example of mosaic evolution. *Nature*, 226,

678 1268.

679 Daeschler, E.B., Shubin, N.H. & Jenkins, F.A. (2006). A Devonian tetrapod-like fish and the evolution of

680 the tetrapod body plan. *Nature*, 440, 757–763.

681 Dani, F.R., Michelucci, E., Francese, S., Mastrobuoni, G., Cappellozza, S., La Marca, G., *et al.* (2011).

682 Odorant-binding proteins and chemosensory proteins in pheromone detection and release in the

683 silkmoth Bombyx mori. *Chem. Senses*, 36, 335–344.

684 Darragh, K., Vanjari, S., Mann, F., Gonzalez-Rojas, M.F., Morrison, C.R., Salazar, C., *et al.* (2017). Male sex

685 pheromone components in *Heliconius* butterflies released by the androconia affect female choice.

686 *PeerJ*, 5, e3953.

687 Dasmahapatra, K.K., Walters, J.R., Briscoe, A.D., Davey, J.W., Whibley, A., Nadeau, N.J., *et al.* (2012).

688 Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*,

689 487, 94–98.

690 Fernández, R. & Gabaldón, T. (2020). Gene gain and loss across the metazoan tree of life. *Nat. Ecol. Evol.*,

691 4, 524–533.

692 de Fouchier, A., Montagné, N., Mirabeau, O. & Jacquin-Joly, E. (2014). Current views on the function and
693 evolution of olfactory receptors in Lepidoptera. *Short Views Insect Biochem. Mol. Biol. S*, 2, 385–
694 408.
695 van Gestel, J., Ackermann, M. & Wagner, A. (2019). Microbial life cycles link global modularity in
696 regulation to mosaic evolution. *Nat. Ecol. Evol.*, 3, 1184–1196.
697 Gómez-Robles, A., Hopkins, W.D. & Sherwood, C.C. (2014). Modular structure facilitates mosaic
698 evolution of the brain in chimpanzees and humans. *Nat. Commun.*, 5, 4469.
699 Gong, D.P., Zhang, H. jie, Zhao, P., Lin, Y., Xia, Q.Y. & Xiang, Z.H. (2007). Identification and expression
700 pattern of the chemosensory protein gene family in the silkworm, *Bombyx mori*. *Insect Biochem.*
701 *Mol. Biol.*, 37, 266–277.
702 Gong, D.P., Zhang, H.J., Zhao, P., Xia, Q.Y. & Xiang, Z.H. (2009). The odorant binding protein gene family
703 from the genome of silkworm, *Bombyx mori*. *BMC Genomics*, 10, 332.
704 Grosse-Wilde, E., Kuebler, L.S., Bucks, S., Vogel, H., Wicher, D. & Hansson, B.S. (2011). Antennal
705 transcriptome of Manduca sexta. *Proc. Natl. Acad. Sci. U. S. A.*, 108, 7449–7454.
706 Grula, J.W., McChesney, J.D. & Taylor, O.R. (1980). Aphrodisiac pheromones of the sulfur butterflies
707 Colias eurytheme and C. Philodice (Lepidoptera, Pieridae). *J. Chem. Ecol.*, 6, 241–256.
708 Gu, S.H., Wu, K.M., Guo, Y.Y., Pickett, J.A., Field, L.M., Zhou, J.J., *et al.* (2013). Identification of genes
709 expressed in the sex pheromone gland of the black cutworm *Agrotis ipsilon* with putative roles in
710 sex pheromone biosynthesis and transport. *BMC Genomics*, 14, 636.
711 Guijarro-Clarke, C., Holland, P.W.H. & Paps, J. (2020). Widespread patterns of gene loss in the evolution
712 of the animal kingdom. *Nat. Ecol. Evol.*, 4, 519–523.
713 Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms
714 and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML
715 3.0. *Syst. Biol.*, 59, 307–321.

716 Guo, H., Del Corso, A., Huang, L.Q., Mura, U., Pelosi, P. & Wang, C.Z. (2014). Aldehyde reductase activity
717 in the antennae of Helicoverpa armigera. *Insect Mol. Biol.*, 23, 330–340.
718 Hansson, B.S. & Stensmyr, M.C. (2011). Evolution of insect olfaction. *Neuron*, 72, 698–711.
719 Helmkampf, M., Cash, E. & Gadau, J. (2015). Evolution of the insect desaturase gene family with an
720 emphasis on social hymenoptera. *Mol. Biol. Evol.*, 32, 456–471.
721 Heuskin, S., Vanderplanck, M., Bacquet, P., Holveck, M.J., Kaltenpoth, M., Engl, T., *et al.* (2014). The
722 composition of cuticular compounds indicates body parts, sex and age in the model butterfly
723 *Bicyclus anynana* (Lepidoptera). *Front. Ecol. Evol.*, 2, 1–16.
724 Johansson, B.G. & Jones, T.M. (2007). The role of chemical communication in mate choice. *Biol. Rev.*, 82,
725 265–289.
726 Jurenka, R. (2004). Insect Pheromone Biosynthesis. In: *The Chemistry of Pheromones and Other*
727 *Semiochemicals I. Topics in Current Chemistry, vol 239* (ed. Schulz, S.). Springer, pp. 97–132.
728 Keys, D.N., Lewis, D.L., Selegue, J.E., Pearson, B.J., Goodrich, L. V., Johnson, R.L., *et al.* (1999).
729 Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science (80-.).*, 283,
730 532–534.
731 Latorre-Estivalis, J.M., Omondi, B.A., DeSouza, O., Oliveira, I.H.R., Ignell, R. & Lorenzo, M.G. (2015).
732 Molecular basis of peripheral olfactory plasticity in *Rhodnius prolixus*, a Chagas disease vector.
733 *Front. Ecol. Evol.*, 3, 74.
734 Leal, W.S. (2013). Odorant reception in insects: Roles of receptors, binding proteins, and degrading
735 enzymes. *Annu. Rev. Entomol.*, 58, 373–391.
736 Legeai, F., Malpel, S., Montagné, N., Monsempes, C., Cousserans, F., Merlin, C., *et al.* (2011). An
737 expressed sequence tag collection from the male antennae of the noctuid moth *Spodoptera*
738 *littoralis*: A resource for olfactory and pheromone detection research. *BMC Genomics*, 12, 86.
739 Legendre, P. & Legendre, L. (1998). *Numerical Ecology*. 2nd edn. Elsevier, Amsterdam.

740 Levi-Zada, A., David, M., Fefer, D., Seplyarsky, V., Sadowsky, A., Dobrinin, S., *et al.* (2014). Circadian
741 release of male-specific components of the greater date moth, *Aphomia (Arenipses) Sabella*, using
742 sequential SPME/GC/MS analysis. *J. Chem. Ecol.*, 40, 236–243.
743 Li, W., Cong, Q., Shen, J., Zhang, J., Hallwachs, W., Janzen, D.H., *et al.* (2019). Genomes of skipper
744 butterflies reveal extensive convergence of wing patterns. *Proc. Natl. Acad. Sci. U. S. A.*, 116, 6232–
745 6237.
746 Liénard, M.A., Wang, H.L., Lassance, J.M. & Löfstedt, C. (2014). Sex pheromone biosynthetic pathways
747 are conserved between moths and the butterfly *Bicyclus anynana*. *Nat. Commun.*, 5, 3957.
748 Marques, D., Meier, J. & Seehausen, O. (2019). A combinatorial view on speciation and adaptive
749 radiation. *Trends Ecol. Evol.*, 34, 531–544.
750 Moore, P.J., Reagan-Wallin, N.L., Haynes, K.F. & Moore, A.J. (1997). Odour conveys status on
751 cockroaches. *Nature*, 389, 25.
752 Moto, K., Yoshiga, T., Yamamoto, M., Takahashi, S., Okano, K., Ando, T., *et al.* (2003). Pheromone gland-
753 specific fatty-acyl reductase of the silkmoth, *Bombyx mori*. *Proc. Natl. Acad. Sci. U. S. A.*, 100, 9156–
754 9161.
755 Myers, J. (1972). Pheromones and courtship behavior in butterflies. *Am. Zool.*, 12, 545–551.
756 Nieberding, C.M., Fischer, K., Saastamoinen, M., Allen, C.E., Wallin, E.A., Hedenström, E., *et al.* (2012).
757 Cracking the olfactory code of a butterfly: The scent of ageing. *Ecol. Lett.*, 15, 415–424.
758 Nieberding, C.M., San Martin, G., Saenko, S., Allen, C.E., Brakefield, P.M. & Visser, B. (2018). Sexual
759 selection contributes to partial restoration of phenotypic robustness in a butterfly. *Sci. Rep.*, 8, 1–
760 13.
761 Nieberding, C.M., de Vos, H., Schneider, M. V., Lassance, J.M., Estramil, N., Andersson, J., *et al.* (2008).
762 The male sex pheromone of the butterfly *Bicyclus anynana*: Towards an evolutionary analysis. *PLoS*
763 *One*, 3, 1–12.

764 Nishida, R., Schulz, S., Kim, C.S., Fukami, H., Kuwahara, Y., Honda, K., *et al.* (1996). Male sex pheromone
765 of a giant danaine butterfly, Idea leuconoe. *J. Chem. Ecol.*, 22, 949–972.
766 Nowell, R.W., Elsworth, B., Oostra, V., Zwaan, B.J., Wheat, C.W., Saastamoinen, M., *et al.* (2017). A high-
767 coverage draft genome of the mycalesine butterfly *Bicyclus anynana*. *Gigascience*, 6, 1–7.
768 Pelosi, P., Iovinella, I., Felicioli, A. & Dani, F.R. (2014). Soluble proteins of chemical communication: An
769 overview across arthropods. *Front. Physiol.*, 5, 1–13.
770 Pelosi, P., Zhou, J.J., Ban, L.P. & Calvello, M. (2006). Soluble proteins in insect chemical communication.
771 *Cell. Mol. Life Sci.*, 63, 1658–1676.
772 Pelosi, P., Zhu, J. & Knoll, W. (2018). Odorant-binding proteins as sensing elements for odour monitoring.
773 *Sensors*, 18, 3248.
774 Pliski, T. & Eisner, T. (1969). Sex pheromone of the queen butterfly: Biology. *Science (80-.).*, 164, 1170–
775 1172.
776 R Core Team. (2020). R: A language and environment for statistical computing.
777 Rafaeli, A. (2016). Revelations on the regulatory mechanisms in moth sex-pheromone signals. In:
778 *Management of Insect Pests to Agriculture: Lessons Learned from Deciphering their Genome,*
779 *Transcriptome and Proteome* (eds. Czosnek, H. & Ghanim, M.). Springer, pp. 115–129.
780 Rosén, W.Q. (2002). Endogenous control of circadian rhythms of pheromone production in the turnip
781 moth, Agrotis segetum. *Arch. Insect Biochem. Physiol.*, 50, 21–30.
782 Sakurai, T., Nakagawa, T., Mitsuno, H., Mori, H., Endo, Y., Tanoue, S., *et al.* (2004). Identification and
783 functional characterization of a sex pheromone receptor in the silkmoth *Bombyx mori*. *Proc. Natl.*
784 *Acad. Sci. U. S. A.*, 101, 16653–16658.
785 Sánchez-Gracia, A., Vieira, F.G. & Rozas, J. (2009). Molecular evolution of the major chemosensory gene
786 families in insects. *Heredity (Edinb).*, 103, 208–216.
787 Sappington, T.W. & Taylor, O.R. (1990a). Developmental and environmental sources of pheromone

788 variation in Colias eurytheme butterflies. *J. Chem. Ecol.*, 16, 2771–2786.

789 Sappington, T.W. & Taylor, O.R. (1990b). Disruptive sexual selection in Colias eurytheme butterflies. *Proc.*

790 *Natl. Acad. Sci. U. S. A.*, 87, 6132–6135.

791 Sarto Monteys, V.I., Quero, C., Santa-Cruz, M.C., Rosell, G. & Guerrero, A. (2016). Sexual communication

792 in day-flying Lepidoptera with special reference to castniids or "butterfly-moths." *Bull. Entomol.*

793 *Res.*, 106, 421–431.

794 Shen, S., Cao, S., Zhang, Z., Kong, X., Liu, F., Wang, G., *et al.* (2020). Evolution of sex pheromone

795 receptors in Dendrolimus punctatus Walker (lepidoptera: Lasiocampidae) is divergent from other

796 moth species. *Insect Biochem. Mol. Biol.*, 122, 103375.

797 Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E.M. (2015). BUSCO:

798 Assessing genome assembly and annotation completeness with single-copy orthologs.

799 *Bioinformatics*, 31, 3210–3212.

800 Smadja, C. & Butlin, R.K. (2009). On the scent of speciation: The chemosensory system and its role in

801 premating isolation. *Heredity (Edinb).*, 102, 77–97.

802 Soques, S., Vásquez, G.M., Grozinger, C.M. & Gould, F. (2010). Age and mating status do not affect

803 transcript levels of odorant receptor genes in male antennae of *Heliothis virescens* and *Heliothis*

804 *subflexa*. *J. Chem. Ecol.*, 36, 1226–1233.

805 Stryjewski, K.F. & Sorenson, M.D. (2017). Mosaic genome evolution in a recent and rapid avian radiation.

806 *Nat. Ecol. Evol.*, 1, 1912–1922.

807 Svensson, M. (1996). Sexual selection in moths: The role of chemical communication. *Biol. Rev.*, 71, 113–

808 135.

809 Taylor, O. (1973). Reproductive isolation in *Colias eurytheme* and *C. philodice* (Lepidoptera: Pieridae):

810 Use of olfaction in mate selection. *Ann. Entomol. Soc. Am.*, 66, 621–626.

811 Tillman, J.A., Seybold, S.J., Jurenka, R.A. & Blomquist, G.J. (1999). Insect pheromones - An overview of

812 biosynthesis and endocrine regulation. *Insect Biochem. Mol. Biol.*, 29, 481–514.

813 Tupec, M., Buček, A., Janoušek, V., Vogel, H., Prchalová, D., Kindl, J., *et al.* (2019). Expansion of the fatty

814 acyl reductase gene family shaped pheromone communication in Hymenoptera. *Elife*, 8, 1–33.

815 Tupec, M., Buček, A., Valterová, I. & Pichová, I. (2017). Biotechnological potential of insect fatty acid-

816 modifying enzymes. *Zeitschrift fur Naturforsch. - Sect. C J. Biosci.*, 72, 387–403.

817 Vane-Wright, R.I. & Boppre, M. (1993). Visual and chemical signalling in butterflies: functional and

818 phylogenetic perspectives. *Philos. Trans. - R. Soc. London, B*, 340, 197–205.

819 Venthur, H. & Zhou, J.J. (2018). Odorant receptors and odorant-binding proteins as insect pest control

820 targets: A comparative analysis. *Front. Physiol.*, 9, 1–16.

821 Vieira, F.G. & Rozas, J. (2011). Comparative genomics of the odorant-binding and chemosensory protein

822 gene families across the arthropoda: Origin and evolutionary history of the chemosensory system.

823 *Genome Biol. Evol.*, 3, 476–490.

824 Vieira, F.G., Sánchez-Gracia, A. & Rozas, J. (2007). Comparative genomic analysis of the odorant-binding

825 protein family in 12 *Drosophila* genomes: Purifying selection and birth-and-death evolution.

826 *Genome Biol.*

827 Vogt, R.G., Große-Wilde, E. & Zhou, J.J. (2015). The Lepidoptera odorant binding protein gene family:

828 Gene gain and loss within the GOBP/PBP complex of moths and butterflies. *Insect Biochem. Mol.*

829 *Biol.*, 62, 142–153.

830 Wahlberg, N., Wheat, C.W. & Peña, C. (2013). Timing and patterns in the taxonomic diversification of

831 Lepidoptera (butterflies and moths). *PLoS One*, 8, 1–8.

832 Walker, W.B., Gonzalez, F., Garczynski, S.F. & Witzgall, P. (2016). The chemosensory receptors of codling

833 moth *Cydia pomonella*-expression in larvae and adults. *Sci. Rep.*, 6, 1–15.

834 Walker, W.B., Roy, A., Anderson, P., Schlyter, F., Hansson, B.S. & Larsson, M.C. (2019). Transcriptome

835 Analysis of Gene Families Involved in Chemosensory Function in Spodoptera littoralis (Lepidoptera:

836 Noctuidae). *BMC Genomics*, 20, 1–20.

837 Wyatt, T.D. (2014). *Pheromones and animal behaviour: chemical signals and signatures*. 2nd edn.

838 Cambridge Univ Press, Cambridge (UK).

839 Xu, X., Currie, P., Pittman, M., Xing, L., Meng, Q., Lü, J., *et al.* (2017). Mosaic evolution in an

840 asymmetrically feathered troodontid dinosaur with transitional features. *Nat. Commun.*, 8.

841 Xu, Y.L., He, P., Zhang, L., Fang, S.Q., Dong, S.L., Zhang, Y.J., *et al.* (2009). Large-scale identification of

842 odorant-binding proteins and chemosensory proteins from expressed sequence tags in insects.

843 *BMC Genomics*, 10, 1–13.

844 Yamamoto, K., Higashiura, A., Suzuki, M., Shiotsuki, T., Sugahara, R., Fujii, T., *et al.* (2016). Structural

845 characterization of an aldo-keto reductase (AKR2E5) from the silkworm *Bombyx mori*. *Biochem.*

846 *Biophys. Res. Commun.*, 474, 104–110.

847 Yew, J.Y. & Chung, H. (2015). Insect pheromones: An overview of function, form, and discovery. *Prog.*

848 *Lipid Res.*, 59, 88–105.

849 Zhan, S., Merlin, C., Boore, J.L. & Reppert, S.M. (2011). The monarch butterfly genome yields insights into

850 long-distance migration. *Cell*, 147, 1171–1185.

851 Zhang, D.D. & Löfstedt, C. (2015). Moth pheromone receptors: Gene sequences, function, and evolution.

852 *Front. Ecol. Evol.*, 3, 1–10.

853 Zhang, J., Cong, Q., Shen, J., Opler, P. & Grishin, N. (2019). Genomics of a complete butterfly continent.

854 *bioRxiv*, 845, 829887.

855 Zhang, J., Walker, W.B. & Wang, G. (2015). Pheromone reception in moths: From molecules to behaviors.

856 In: *Progress in Molecular Biology and Translational Science* (ed. Glatz, R.). Academic Press, pp. 109–

857 128.

858 Zhang, Y.N., Xia, Y.H., Zhu, J.Y., Li, S.Y. & Dong, S.L. (2014). Putative pathway of sex pheromone

859 biosynthesis and degradation by expression patterns of genes identified from female pheromone

860 gland and adult antenna of </i>Sesamia inferens</i> (Walker). *J. Chem. Ecol.*, 40, 439–451.

861